Харківський національний
університет радіоелектроніки

Державне підприємство
"Південний державний
проектно-конструкторський
та науково-дослідний інститут
авіаційної промисловості"

Kharkiv National
University of Radio Electronics

State Enterprise
"Southern National Design
&
Research Institute
of Aerospace Industries"

СУЧАСНИЙ СТАН НАУКОВИХ ДОСЛІДЖЕНЬ ТА ТЕХНОЛОГІЙ В ПРОМИСЛОВОСТІ

INNOVATIVE TECHNOLOGIES AND SCIENTIFIC SOLUTIONS FOR INDUSTRIES

№ 2 (28), 2024

Щоквартальний науковий журнал

Харків
2024

No. 2 (28), 2024

Quarterly scientific journal

Kharkiv
2024

INNOVATIVE TECHNOLOGIES AND SCIENTIFIC SOLUTIONS FOR INDUSTRIES

No. 2 (28), 2024

ЗАСНОВНИКИ
Харківський національний університет радіоелектроніки,
Державне підприємство "Південний державний проектно-конструкторський та науково-дослідний інститут авіаційної промисловості"

АДРЕСА РЕДАКЦІЇ:
Україна, 61166, м. Харків, проспект Науки, 14
Інформаційний сайт: http://itssi-journal.com
https://journals.uran.ua/itssi
E-mail редколегії: journal.itssi@gmail.com

ESTABLISHERS
Kharkiv National University of Radio Electronics,
State Enterprise "National Design & Research Institute of Aerospace Industries"

EDITORIAL OFFICE ADDRESS:
Ukraine, 61166, Kharkiv, Nauka Ave, 14
Information site: http://itssi-journal.com
https://journals.uran.ua/itssi
E-mail of the editorial board: journal.itssi@gmail.com

Журнал включено до Переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата наук наказом Міністерства освіти і науки України від 16.07.2018 №775 (додаток 7).

Витяг з реєстру суб'єктів у сфері медіа – реєстрантів від 25.04.2024 № 1410. Ідентифікатор медіа R30-03878.

СУЧАСНИЙ СТАН НАУКОВИХ ДОСЛІДЖЕНЬ ТА ТЕХНОЛОГІЙ В ПРОМИСЛОВОСТІ

№ 2 (28), 2024

Затверджений до друку Науково-технічною Радою Харківського національного університету радіоелектроніки
(Протокол № 5 від 14 червня 2024 р.)

ЗМІСТ

INNOVATIVE TECHNOLOGIES AND SCIENTIFIC SOLUTIONS FOR INDUSTRIES

No. 2 (28), 2024

Approved for publication by the Scientific and Technical Council of the Kharkiv National University of Radio Electronics
(Protocol No. 5 d.d. 14 June 2024)

CONTENTS

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Головний редактор журналу, доктор технічних наук, професор
Євгеній БОДЯНСЬКИЙ
Editor-in-Chief, Doctor of Technical Sciences, Professor
Yevgeniy BODYANSKIY

Шановні друзі!

У новому номері наукового журналу "Сучасний стан наукових досліджень та технологій в промисловості" ми вкотре зібрали найцінніші напрацювання та здобутки, що їх наразі мають наші науковці. Варто наголосити на тому, що ці результати одержані в умовах воєнного стану й водночас вони є цінними для промисловості та обороноздатності нашої держави.

Зокрема, в номер увійшли статті присвячені сучасним системам захисту, застосункам штучного інтелекту, управлінню групою роботів, інклюзії, кібергігієні, гібридним нейронним мережам, сталому розвитку в умовах воєнного стану тощо.

Від імені наукової спільноти, висловлюю вдячність Збройним Силам України, завдяки мужності яких ми залишаємося в рідних містах, на робочих місцях в alma mater та продовжуємо торувати науковий шлях задля розвитку та майбутнього нашої держави.

Також вітаю усіх українців та українок з Днем Незалежності! Це свято сповнилось для кожного з нас особливим змістом, воно символізує нашу свободу, силу та єдність. Це також і день пам'яті усіх, хто протягом віків боровся за здобуття Незалежності. Сьогодні ми продовжуємо цей шлях і, безперечно, маємо бути гідними своєї держави та разом творити її інноваційний та справедливий простір.

Впевнений, що наука в цьому процесі матиме надважливе значення. Тому висловлюю вдячність всім науковцям, які не полишають дослідження та представляють свої здобутки на сторінках нашого видання.

Разом ми – сила! Тож, єднаймось, будьмо незалежними та розвиваймо науку задля майбутнього! Нехай наша країна процвітає, а кожен з нас відчуває гордість за те, що ми – українці.

Dear friends!

In the new issue of our journal "Innovative Technologies and Scientific Solutions for Industries" we have once again collected the most valuable developments and achievements of Ukrainian scientists. It is worth emphasizing that these results were obtained under martial law, and at the same time they are valuable for the industry and defense capabilities of our country.

The issue contains articles on a wide range of problems, including modern defense systems, artificial intelligence applications, robot group management, inclusion, cyber hygiene, hybrid neural networks, sustainable development under martial law, etc.

On behalf of the scientific community, I express my gratitude to the Armed Forces of Ukraine! Thanks to their courage, we stay in our hometowns, at our workplaces in our alma mater, and continue to pave the scientific path for the development and future of our country.

I also congratulate all Ukrainians on Independence Day! This holiday has a special meaning for each of us, symbolizing our freedom, strength and unity. It is also a day of remembrance for all those who have fought for centuries to gain independence. Today, we continue this path and, of course, we must be worthy of our country and together create its innovative and fair space.

I am confident that science will play a crucial role in this process. Therefore, I express my gratitude to all the scientists who do not give up research and share their achievements on the pages of our publication.

Together we are a Force! So, let's unite, be independent and develop science for the future! May our country prosper and each of us feel proud to be Ukrainians.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

UDC 004.91 DOI: https://doi.org/10.30837/2522-9818.2024.28.006

O. BARKOVSKA, V. SERDECHNYI

INTELLIGENT ASSISTANCE SYSTEM FOR PEOPLE WITH VISUAL IMPAIRMENTS

Subject of the Research: The creation of an intelligent assistance system for people with visual impairments. Nowadays, the task of developing effective intelligent assistance systems that allow people with vision problems to achieve maximum independence is important and relevant, as existing systems have a number of drawbacks, such as limited autonomy, limited integration with other devices and systems, limited analysis of dynamic obstacles, and limited user feedback capabilities, which in most cases are restricted to voice guidance. **Objective:** The objective of this work is to create a generalized functional model of an intelligent assistance system for people with visual impairments, which has enhanced autonomy, integration with other devices and systems, the ability to analyze dynamic objects and predict their movement trajectory, and provide diverse feedback to the user. **Tasks:** To achieve the set objective, the following tasks were accomplished: a generalized functional model of the proposed intelligent assistance system for people with visual impairments was created; the functional dependencies of the components of the developed model were substantiated; a review of the basic modules of the proposed system model was conducted. **Methods:** The methods used include functional modeling methods and system analysis methods. **Results:** The following results were obtained: a functional model of an intelligent assistance system for people with visual impairments was proposed. This system surpasses existing analogs in a number of functional capabilities: detection of static and dynamic obstacles with prediction of dynamic obstacles' movement trajectory, the ability to operate in various conditions (indoors, outdoors, in light or dark, in different weather conditions), support for integration with other systems and devices, and a high level of autonomy. **Conclusions:** The developed system model has enhanced autonomy, integration with other devices and systems, the ability to analyze dynamic objects and predict their movement trajectory, and provides diverse feedback to the user.

Keywords: system; assistant; vision; LiDAR; video; trajectory; prediction; intelligent system; recognition; classification.

Introduction

According to the IAPB [1], in 2020 there were at least 1.1 billion people with visual impairment worldwide, of whom 43 million were completely blind. In 2023, the number of people with visual impairments increased to 2.2 billion, according to WHO [2]. Statistics show that the number of people who need help due to complete or partial lack of vision is growing every year.

In everyday life, these people face a number of challenges. As noted in the study [3], some of the key problems of people with visual impairments are a sense of burden and dependence (due to the need for help from others, especially with transportation), social interaction (such people often cannot perceive non-verbal signals, which makes social interaction difficult), and the use of aids (the use of things like a cane or guide dog marks a person as blind, which can lead to discrimination and lower self-esteem). In his article [4], author and disability rights activist Samiak Lalit also points out a number of problems that complicate the lives of people with visual impairments, including difficulty navigating in space, sorting clothes (most visually impaired people identify things by their shape and texture, and this makes organizing laundry a difficult task), gaining independence through modern devices (the author points out that the necessary equipment that can allow a blind person to live independently is not easily found in local stores or markets. Because of this, a person needs to make a lot of effort to get every device that can make them one step closer to independence).

People with vision problems are also more susceptible to serious life events. According to a study [8], the prevalence of serious life events for such people is higher than in the general population (60%, $p < 0.001$), especially in the case of fire or explosions, serious accidents, sexual violence, life-threatening illnesses or injuries, and severe human suffering.

Thus, it can be concluded that the task of creating effective intelligent assistance systems that will allow people with vision problems to gain maximum independence from other people, solve everyday problems more easily, and gain a higher level of life safety is currently important and relevant.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Analysis of existing systems

To date, researchers have proposed a number of different intelligent systems to assist people with vision problems.

For example, in [5] (2018), a system called "A Smart Personal AI Assistant for Visually Impaired People" is proposed, which is based on Google's Cloud API platform and allows recognizing objects and textual information in photos (using the Cloud Vision API) and communicating with a chatbot to obtain the necessary information (using Dialogflow). The user can interact with the system through voice commands to his or her smartphone, and the system generates answers to questions or descriptions of objects and text recognized in the photos in audio format and plays them back for the user.

The article [6] (2019) describes a system called ANSVIP, whose main purpose is to assist in navigation for visually impaired people. The system uses the Google ARCore platform (specifically, the SLAM technology is used) running on a smartphone running on the Android OS. The use of SLAM technology allows the system to work indoors, unlike traditional GPS localization, which can work intermittently indoors. A field-based path planning method was created to keep a person away from various obstacles to prevent collisions during real-time path generation. In addition, the authors proposed a two-channel user feedback mechanism. The first channel is a classic voice guidance, and the second is tactile gloves. The left glove sets the direction of a person's movement, and the right glove warns of obstacles.

Article [9] (2020) describes an intelligent assistance system for the movement of visually impaired people that receives data from a video camera, performs object detection in video frames, and, if an obstacle is detected on the user's path, warns him or her with a corresponding voice message using the corresponding Text to Speech converter module.

The paper [10] (2023) shows a prototype of an assistance system for visually impaired people based on the Raspberry Pi module due to its low cost, small size, and ease of integration. The user's proximity to an obstacle is measured using a camera and ultrasonic sensors. Feedback to the user is realized through voice guidance. It is noted that the system can work both indoors and outdoors. In addition, the system proposed the introduction of a reading module and integration with a pulse oximeter, which allows the system to determine when a person is in danger and call the emergency number.

Paper [7] (2023) proposes a deep learning-based assistance system for visually impaired people called DeepNAVI. The main components of the system are a smartphone, a wireless bone headset, and six software modules: an obstacle detection module, a distance measurement module, a location measurement module, a motion detection module, and a scene recognition module. The system uses data from a smartphone's video camera and is able to detect 20 different types of obstacles and 20 different types of scenes. Interaction with the user takes place through voice guidance.

Analyzing these works, we can identify a number of common weaknesses of all the systems under consideration, which are listed in Table 1.

Table 2 shows the main areas for improving such systems.

Table 1. *The main disadvantages of the considered assistance systems*

Disadvantage	Description of the disadvantages
Limited autonomy	Most systems are dependent on external devices (smartphones, Raspberry Pi) or have limited functionality in offline mode. This makes them less useful in situations where there is no access to these devices.
Limited integration	None of the systems have full integration with other systems and devices. This makes it difficult to use them in conjunction with other assistive technologies, which could significantly expand their functionality.
Limited analysis of dynamic interference	Most systems can only detect static obstacles, which is not always enough for safe navigation in the real world. In addition, none of the systems can predict the trajectory of dynamic objects, which is also a significant limitation in the context of predicting a possible collision.
Limited feedback	Feedback is mostly limited to voice interaction, which may be insufficient for people with hearing impairments.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Table 2. *Key areas for improving assistance systems*

Areas for improvement	Description
Increasing autonomy	Developing your own computing modules or using more energy-efficient components can reduce dependence on external devices and increase battery life.
Expanding integration	Integration with other systems, such as GPS, maps, public transportation, smart home, etc., can greatly expand the capabilities of systems and make them more useful in everyday life.
Improving the analysis of dynamic interference	The use of more advanced computer vision and machine learning algorithms can allow systems to better detect and analyze moving objects, which will increase user safety.
Development of various feedback channels	The introduction of haptic, vibration, or visual feedback can make systems more accessible to people with various disabilities.
Expanding the functionality	Adding new features, such as text, emotion, and gesture recognition, can make systems more adaptive to user needs.

These improvements will make systems for helping visually impaired people more efficient, versatile, and easy to use.

The theoretical foundations for the creation of such systems [5–7, 9, 10] are being actively developed and published in the following articles: [12] focuses on the fact that smartphones and wearable devices with built-in cameras are the main means of supporting the most advanced computer vision solutions that allow both positioning and controlling the area around the user; [13] substantiates the interdisciplinary importance and great social impact of such assistive technologies for visually impaired and blind people; [14] – a comparative review of wearable and portable assistive devices for visually impaired people to show the progress in assistive technologies for this group of people, highlighting the advantages and disadvantages of systems to further improve the level of safety, independence and mobility for visually impaired people.

Results and discussion

Assistance systems for visually impaired people (ASP) are complex systems with a large number of components that interact with each other. The components of the systems include modules for detecting the human environment in the form of audio or video data, modules for analyzing the environment near the person, modules for making decisions about the presence of danger, modules for feedback to the user, etc. Representation of such systems in the form of generalized models allows us to clearly define the components of the system, their functions and interaction between them, their influence on each other, identify weaknesses or potential problem areas in the system, plan and manage research aimed at improving systems, etc.

The article proposes an ASP model in the form of a generalized functional flowchart, as this will provide an opportunity to show the components of the system in more detail and provide a basis for applying a systematic approach to future research, which involves taking into account all the relationships and influences between the components. The above system model (Figure 1) will be used in the future to develop a prototype, conduct simulations and test the system, establish criteria for evaluating the effectiveness of the system and its components, which will contribute to a more objective analysis of the research results.

One of the main features of this system is that its practical implementation is envisaged by a separate module independent of the user, which moves alongside the user and performs all the necessary functions to accompany him or her and ensure his or her safety. In this case, the use of a smartphone is necessary only to ensure communication between the user and the assistant module.

In addition, the system uses data of high heterogeneity. The system works with input data of several types: video frames (data coming from video cameras), coordinates of points in space (data coming from LiDAR sensors), audio data (data coming from a microphone), and temperature sensor data. This approach is expected to improve the quality of object and scene detection, as well as environmental conditions.

The system functionality is provided by a number of modules, namely: Environment Conditions Detection, Video & LiDAR data merging, Objects Detection & Tracking, Trajectory Prediction, Trajectory Prediction Model Continuous Training, Generating an Assistance Decision, Generating an AI Assistant Module Control Signal, Assistance Decision Signal Conversion, Assistant Module Controller, External System & Sensors Data

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Processing. The modules are interconnected and interdependent, which can be seen in the proposed model of an intelligent assistance system for visually impaired people.

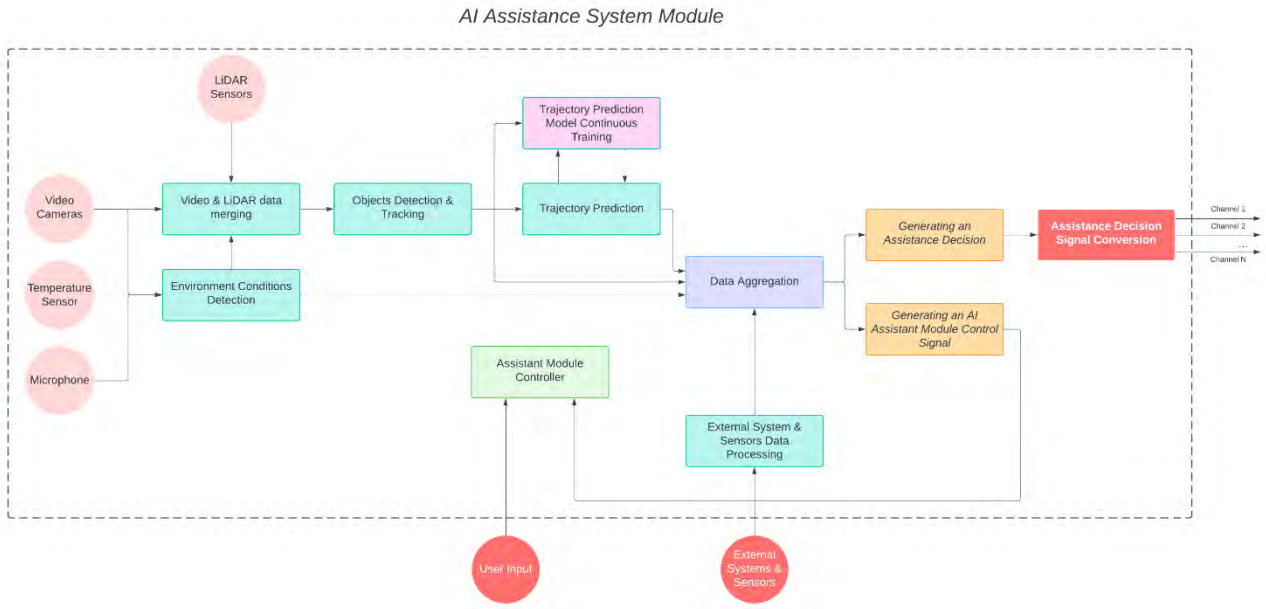


Fig. 1. Generalized functional model of an intelligent assistance system

Let us consider these modules and their functional interconnection in detail in the following subsections of the article.

Analysis of modules of the model of the system of assistance to visually impaired people

The first stage of the system's operation is the determination of environmental conditions, which is realized through a separate intelligent model called Environment Conditions Detection. This model uses different types of input data, namely video data, microphone data, and temperature sensor data. The task of this model is to solve the problem of classification and generate appropriate class labels that describe environmental conditions. Table 3 shows all the proposed class labels.

The stage of environmental conditions detection is necessary for two main purposes: taking these conditions into account for more accurate prediction of the trajectory of dynamic objects, notifying the user of weather conditions that require additional attention (for example, ice) and taking these conditions into account for applying the appropriate image preprocessing stack (for example, in the dark).

The labels of the defined classes are transferred to the Video & LiDAR data merging module, which contains an adaptive video frame pre-processing stack. The composition and parameters of this stack depend on the corresponding class labels, which are determined by the environmental detection module. The diagram of the Video & LiDAR data merging module is shown in Figure 2.

If, according to the environmental conditions, data pre-processing is not required, the Video Frames Preprocessing stage will be ignored and the video data will be directly transferred to the process of combining with LiDAR data.

The Objects Detection & Tracking module is designed to detect and track objects based on integrated video and LiDAR data. At this stage, the data on the identified objects are divided into static (stationary) and dynamic (moving). Data on dynamic objects are used separately to predict the trajectory of these objects and determine their possible coordinates after a certain period of time. Such prediction allows to assess the degree to which a moving object can be dangerous for a person. A diagram of this process is shown in Figure 3.

The first division is already the result of the work of an intelligent object detection model based on YOLO v9, the features and specifications of which are described in detail by the creators on the web resource [16].

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Table 3. *Classes of environmental conditions*

Group	Class label
Weather conditions	Sunny
	Cloudy
	Rain
	Snow
	Fog
	Wind
	Thunderstorm
	Ice
Lighting conditions	Bright sunlight
	Cloudy
	Twilight
	Darkness
	Artificial lighting
Acoustic conditions	Silence
	Noise of transport
	Rain noise
	Thunder
	Natural noise (birds, water, wind)
	Extraordinary car sounds (siren, car horn)
Temperature conditions	Low temperature (–30°C ÷ +5°C)
	Moderate temperature (+5°C ÷ +20°C)
	High temperature (+20°C ÷ +30°C)
	Heat (>+30°C)
Type of terrain	Urban environment
	Countryside
	Forest
	Beach
	Residential premises
	Office space
	Commercial premises
	Public premises
	Sports premises
	Industrial premises
Specific events	Repair work
	Fire
	Explosion
	Police operation
	Medical emergency
	Holiday (concert, parade, festival, wedding)

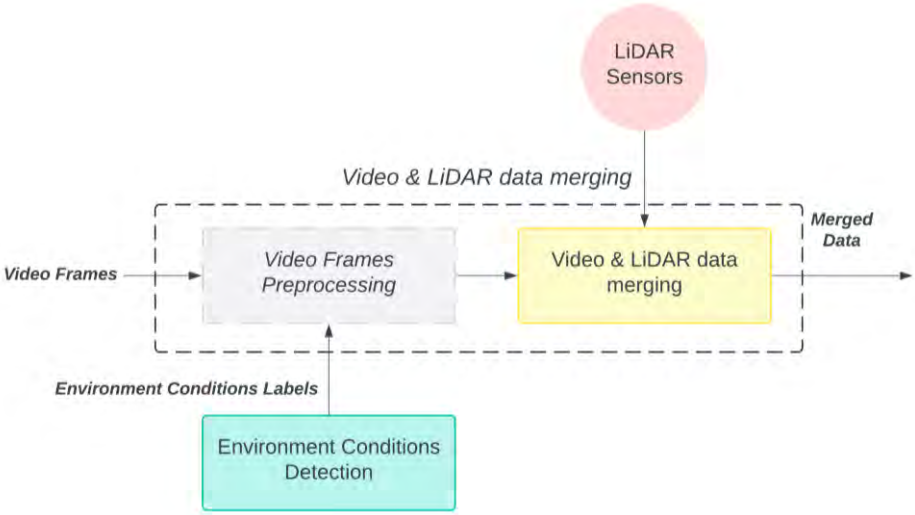


Fig. 2. Video & LiDAR data merging module

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

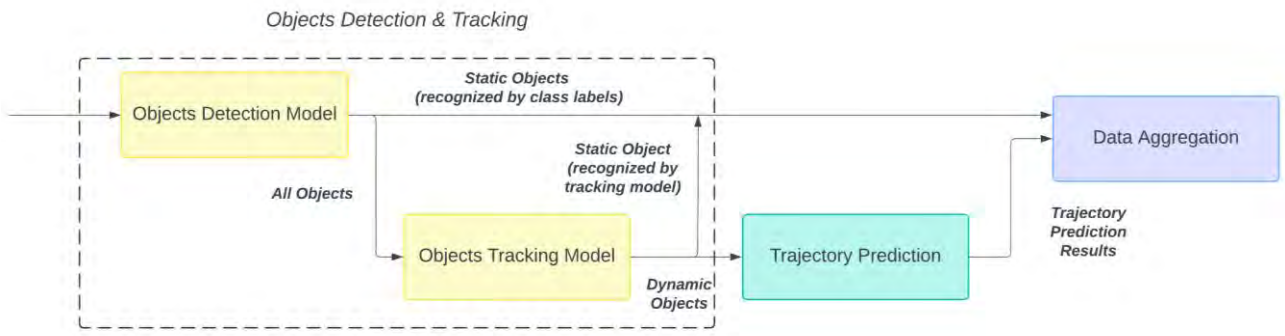


Fig. 3. The process of identifying static and dynamic objects with motion trajectory predictions

Objects are divided into static (e.g., house, bench, pole) and potentially dynamic (e.g., car, bus) according to the defined class labels. The location of potentially dynamic objects is recorded for a certain period of time to determine the presence of movement. Based on the results of such fixation, the objects are re-divided into static and dynamic. If no motion is detected (for example, a car is parked but not moving), such objects are transferred from the potentially dynamic category to the static category. If the motion is detected, such objects are defined as dynamic and their coordinates are passed on to the motion trajectory prediction model.

The task of the Trajectory Prediction module is to predict the trajectories of dynamic objects and their future coordinates in a certain period of time. To improve the accuracy of this module, it is proposed to create a mechanism for retraining this model in real time based on data on the real and predicted location of dynamic objects. The scheme of this mechanism is shown in Figure 4.

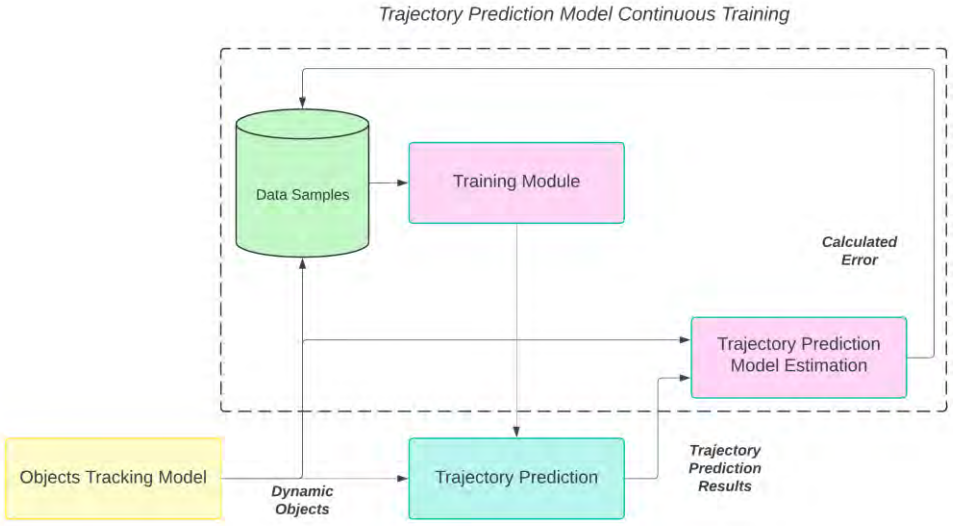


Fig. 4. A mechanism for retraining an intelligent model for predicting object trajectories

The system saves information about the movement of objects (their location at different time intervals) to a separate local database Data Samples to form training data. In addition, the accuracy of the system's predictions is calculated at a certain frequency (selected experimentally) by calculating the difference between the predicted location and the actual location of the object at the corresponding time. This approach allows you to evaluate the accuracy of the model during its operation and launch the retraining mechanism if necessary (if the prediction accuracy is satisfactory, then retraining is not required).

The following data is used to generate an assistant's decision in the Generating an Assistance Decision module:

– data of static and dynamic objects (types of objects, their current location);

– results of predicting the location of dynamic objects;

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

– current environmental conditions;
– data from external systems and devices.

Items 1–3 are formed as a result of the operation of the system modules discussed earlier. Data from external systems and devices is data that the system receives from additional devices that can improve the safety and comfort of a visually impaired person. To work with such data, the system contains a separate module External System & Sensors Data Processing, which is designed to establish communication, receive and process data from external devices. Such devices can be: a pulse oximeter, a smartwatch or bracelet, an air quality sensor, smart home devices, etc. As a result of the Generating an Assistance Decision module, a label of the corresponding class and additional parameters (usually numerical) are generated. In fact, this intelligent model solves both classification and regression tasks. Examples of classes and their parameters are shown in Table 4.

The generated result is transferred to the Assistant Decision Signal Conversion module of the system. The task of this module is to convert the class label and additional parameters to the format of the corresponding user notification channel. The system model includes several feedback channels. They are shown in Figure 5.

Table 4. *Class labels and parameters that result from the Generating an Assistance Decision module*

Group	Class	Parameters
Navigation	Adjustment of the route	Number of steps and direction to adjust
	The need to stop	–
	Start of movement	–
	The need to turn around	Number of degrees and direction of rotation
	Appearance of an obstacle on the way	Distance to an obstacle and its type
	Description of the environment	Text description of objects and scenes along the route
Security (this group has a higher priority than the others)	Approaching vehicles	Distance to the vehicle, type (car, motorcycle, bicycle), direction of movement, speed
	Change of traffic light signal	Traffic light signal color (green, yellow, red), time until the signal changes
	Edge of the sidewalk/roadway	Distance to edge, type of edge (curb, steps), direction
	Falling objects	Distance to object, type of object (branch, glass, other), direction of fall
	Suspicious activity	Type of activity (person following the user, loud conversation, fight), distance to the source of activity, direction
Health	Heart rate deviation	Current heart rate, deviation from the norm (if any)
	Abnormal blood oxygen levels	Current SpO2 value, deviation from the norm (if any)
	Air quality abnormalities	Current values of pollutants (PM2.5, PM10, CO2, etc.), deviation from the norm (if any)
	Medication reminders	Name of medication, time of administration
	Determine the time	Current time
	Weather information	Temperature, precipitation, wind, other parameters
	Notifications from your smartphone	Notification text

Fig. 5. User notification channels

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

There are two main channels of user notification: voice and tactile (via vibration). Voice commands can be transmitted by the assistant module to earbuds, a smartphone, or a portable Bluetooth speaker. Vibrations can be transmitted to a smartwatch, fitness bracelet, or specialized tactile gloves similar to those proposed in [6].

The system is positioned as a separate assistant module. This module can be made in the form of a ground, air, or combined (ground + air) drone. To control this module, it is proposed to introduce a separate Assistant Module Controller unit. The task of this unit is to autonomously control the physical module by transmitting appropriate signals to rotate the wheels, increase or decrease speed, stop the module, etc.

Table 5. *Class labels and parameters that result from the Assistant Module Controller module*

Group	Class	Parameters
Ground mode	Start of movement	–
	Stopping	–
	Increase speed	Speed increase
	Reduce speed	Reducing the speed
	Turn right	Turning angle
	Turn left	Turning angle
Air mode	Take off	Takeoff speed, Takeoff angle
	Landing	Landing speed, Landing angle
	Increase altitude	Climb rate
	Reduce altitude	Altitude decrease
	Turn right	Turning angle
	Turn left	Turning angle

A separate intelligent model is responsible for generating the corresponding signals, which is labeled Generating an AI Assistant Module Control Signal in Figure 1. This model, just like Generating an Assistance Decision, solves two tasks: classification and regression. As a result, this model generates a label of the corresponding class and additional parameters. Examples of classes and their parameters are shown in Table 5.

As a result of the work done, a model of an intelligent assistance system for visually impaired people has been developed. A modern system of intellectual assistance for visually impaired people should meet a number of functionalities, namely: real-time detection of static obstacles, real-time detection of dynamic obstacles and prediction of the trajectories of these obstacles (in order to timely notify the user of danger), recognition of objects (those that are not obstacles) and scenes (classification of the environment), localization and navigation, feedback (sound, tactile notification of the user, etc.), the ability to work in p

Table 6 shows the results of a comparative analysis of the developed model of an intelligent assistance system and its existing analogues.

Table 7 shows the results of the analysis of the compliance of the existing systems and the proposed system with the last five criteria.

Table 6. *Analysis of the compliance of the considered systems of intellectual assistance to visually impaired people with the first part of the specified functionalities*

Intelligent assistance system	Static interference detection	Detection and analysis of dynamic interference	Recognize objects and scenes	Localization and navigation	Feedback
A Smart Personal AI Assistant for Visually Impaired People [5]	No	No	Recognize objects in photos	No	Voice interaction
ANSVIP system [6]	Yes	No	Yes	Yes	Voice interaction, tactile interaction with gloves
The system proposed by Kushal Kumar [9]	Yes	No	Recognize objects (no scene analysis)	No	Voice interaction
System proposed by Surabhi Suresh [10]	Yes	No	Yes	No	Voice guidance and text reading
DeepNAVI system [7]	Yes	Partially (can distinguish moving objects from static ones)	Yes	Yes	Voice interaction
The proposed system	Yes	Yes	Yes	Yes (using external devices, for example, a smartphone with GPS)	Voice interaction, tactile interaction with gloves, smartwatch, fitness bracelet

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Table 7. *Analysis of the compliance of the considered systems of intellectual assistance to visually impaired people with the second part of the specified functionalities*

Intelligent assistance system	Ability to work in different conditions	Support for different work scenarios	Integration with other systems and devices	Autonomy	Reliability in critical situations
A Smart Personal AI Assistant for Visually Impaired People [5]	No	Yes (work in chatbot mode)	No	No	No
ANSVIP system [6]	No	No	No	Yes Yes (depends only on the availability of a smartphone)	No
The system proposed by Kushal Kumar [9]	No	No	No	No	No
System proposed by Surabhi Suresh [10]	Work indoors and outdoors (work in difficult weather conditions is not investigated)	Yes Yes (work in the mode of reading inscriptions)	No	Yes (depends only on the availability of the Raspberry Pi module)	Yes (includes a pulse oximeter and can notify emergency contacts if the user is in danger)
DeepNAVI system [7]	No	No	No	Yes Yes (depends only on the availability of a smartphone)	No
The proposed system	Yes	Yes	Yes	Yes	Yes (due to the integration of data from external devices)

Based on these data, we can see that the proposed system outperforms the analogs considered, offering a more versatile and effective solution in different conditions and scenarios.

Conclusion

In this article, we have proposed a generalized functional model of an intelligent assistance system for visually impaired people, which has increased autonomy, integration with other devices and systems, the ability to analyze dynamic objects and predict their trajectory, and provide various feedback to the user. This system outperforms existing analogs, providing a more versatile and effective solution in various conditions and scenarios. High autonomy is ensured by implementing the system as a separate independent module, has advanced integration with other devices, is able to analyze dynamic objects and provide user feedback through various channels.

The functional dependencies are substantiated and the basic components of the modules of the developed model presented in the article are reviewed.

One of the first large modules is the intelligent module "Environment Conditions Detection", whose task is to determine the environmental conditions. The module has two main goals: to take into account weather conditions to more accurately predict the trajectory of dynamic objects, to notify the user of weather conditions that require additional attention (e.g., ice), and to take these conditions into account when applying the appropriate image preprocessing stack (e.g., in the dark). The Objects Detection & Tracking module is designed to detect and track objects based on integrated video and LiDAR data. At this stage, the data on the identified objects is divided into static (stationary) and dynamic (moving). The task of the Dynamic Objects Trajectory Prediction module is to predict the trajectories of dynamic objects and their future coordinates over a certain period of time. The accuracy of the module is improved by the mechanism of retraining this model in real time based on data on the real and predicted location of dynamic objects. The result generated in the Generating an Assistance Decision module is transferred to the Assistant Decision Signal Conversion module of the system. The task of this module is to convert the class label and additional parameters to the format of the corresponding user notification channel. The system uses several feedback channels: voice and tactile (by means of vibration).

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

The proposed system is focused more on assisting visually impaired people, but its functionality can be expanded for use in other areas, namely:

– autonomous driving of vehicles: the system can be used to develop autonomous driving systems that can safely and efficiently move on roads without human intervention;

– autonomous control of reconnaissance and strike drones and other robotic developments in the military industry: the system can be used to develop autonomous combat systems that can perform combat and reconnaissance missions without risking human life;

– automation of loading and unloading operations and cargo movement: the system can be used to automate processes at enterprises that require cargo movement, such as ports, transport hubs, postal and courier services, agricultural enterprises, retail warehouses, etc.

References

1. IAPB (2020), "Global Estimates of Vision Loss. The International Agency for the Prevention of Blindness", available at: https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/global/ (last accessed 25 March 2024).
2. WHO (2023), "Blindness and vision impairment. World Health Organization", available at: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment (last accessed 25 March 2024).
3. Lynsey, K. Romo,1, Cimmiaron, Alvarez, Melissa, J. Taussig (2023), "An examination of visually impaired individuals communicative negotiation of face threats", *Journal of Social and Personal RelationshipsVolume 40, Issue 1, January 2023*, DOI: https://doi.org/10.1177/02654075221114048
4. Samyak, Lalit, Wecapable (2021), "Daily Life Problems, Struggle and Challenges Faced by Blind People. Wecapable", available at: https://wecapable.com/problems-faced-by-blind-people/ (last accessed 25 March 2024)
5. Shubham, Melvin Felix, Sumer, Kumar, A. Veeramuthu (2018), "A Smart Personal AI Assistant for Visually Impaired People", *Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018),* DOI: http://dx.doi.org/10.1109/ICOEI.2018.8553750
6. Xiaochen, Zhang, Xiaoyu, Yao, Yi Zhu, Fei Hu (2019), "An ARCore Based User Centric Assistive Navigation System for Visually Impaired People", *Appl. Sci.* 9(5), 989 p, DOI: https://doi.org/10.3390/app9050989
7. Bineeth, Kuriakose, Raju, Shrestha, Frode, Eika Sandnes (2023), "DeepNAVI: A deep learning based smartphone navigation assistant for people with visual impairments", *Expert Systems with Applications 212 (2023) 118720,* DOI: https://doi.org/10.1016/j.eswa.2022.118720
8. Audun, Brunes, Trond, Heir (2021), "Serious Life Events in People with Visual Impairment Versus the General Population", *Int. J. Environ. Res. Public Health, 18,* 11536 p, DOI: https://doi.org/10.3390/ijerph182111536
9. Kaushal, Kumar (2023), "An intelligent Assistant for the Visually Impaired & blind people using machine learning, International Journal of Imaging and Robotics 20(3)". available at: https://www.researchgate.net/publication/342561852_An_intelligent_Assistant_for_the_Visually_Impaired_blind_people_using_machine_learning (last accessed 25 March 2024)
10. Surabhi, Suresh, Sandra S, Aisha Thajudheen, Subina Hussain, Amitha R (2023), "An Intelligent Voice Assistance System For The Visually Impaired People", *International Journal of Engineering Research & Technology (IJERT) Volume 11, Issue 04 (2023),* DOI: https://doi.org/10.17577/IJERTCONV11IS04022
11. "YOLOv9: A Leap Forward in Object Detection Technology", available at: https://docs.ultralytics.com/models/yolov9/ (last accessed 22 May 2024).
12. Real, S., Araujo, A. (2029), "Navigation systems for the blind and visually impaired: Past work, challenges, and open problems" *Sensors.* Vol. 19. №. 15. 3404 p. DOI:10.3390/s19153404
13. Bhowmick, A., Hazarika, S. M. (2017), "An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends", *Journal on Multimodal User Interfaces*. T. 11. P. 149–172. DOI: 10.1007/s12193-016-0235-6
14. Elmannai, W., Elleithy, K. (2017), "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions", *Sensors*. Vol. 17. №. 3. P. 565. DOI: 10.3390/s17030565
15. Sayama, H. (2015), "Introduction to the modeling and analysis of complex systems. Open SUNY Textbooks". available at: https://open.umn.edu/opentextbooks/textbooks/233 (last accessed 25 March 2024)
16. Yildirim, U., Campean, F., Williams, H. (2017), "Function modeling using the system state flow diagram", *Artificial Intelligence for Engineering Design Analysis and Manufacturing.* Vol.31. №. 4. P. 413–435. DOI: 10.1017/S0890060417000294

Надійшла (Received) 29.05.2024

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

Відомості про авторів / About the Authors

Барковська Олеся Юріївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри електронних обчислювальних машин, Харків, Україна; e-mail: olesia.barkovska@nure.ua; ORCID ID: https://orcid.org/0000-0001-7496-4353

Сердечний Віталій Сергійович – Харківський національний університет радіоелектроніки, аспірант кафедри електронних обчислювальних машин, Харків, Україна; e-mail: anton.havrashenko@nure.ua; ORCID ID: https://orcid.org/0009-0007-8828-5803

Barkovska Olessia – PhD (Engineering Sciences), Docent, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Electronic Computers, Kharkiv, Ukraine.

Serdechnyi Vitalii – Kharkiv National University of Radio Electronics, PhD student at the Department of Electronic Computers, Kharkiv, Ukraine.

УЗАГАЛЬНЕНА ФУНКЦІОНАЛЬНА МОДЕЛЬ СИСТЕМИ АСИСТУВАННЯ ЛЮДЯМ ІЗ ВАДАМИ ЗОРУ

Предметом дослідження є створення системи інтелектуального асистування для людей з вадами зору. У наш час завдання створення ефективних систем інтелектуального асистування, що дають змогу людям, які мають проблеми із зором, отримати максимальну незалежність, є важливою та актуальною, оскільки наявні системи мають низку недоліків, таких як обмежена автономність, обмежена інтеграція з іншими пристроями та системами, обмежений аналіз динамічних перешкод та обмежені можливості зворотного зв'язку з користувачем, які здебільшого зводяться лише до голосового супроводу. **Метою роботи** є створення узагальненої функціональної моделі системи інтелектуального асистування людям з вадами зору, яка має підвищену автономність, інтеграцію з іншими пристроями та системами, здатність аналізувати динамічні об'єкти та передбачувати траєкторію їхнього руху, забезпечувати різноманітний зворотний зв'язок з користувачем. Для досягнення поставленої мети були виконані такі **завдання**: створено узагальнену функціональну модель запропонованої системи інтелектуального асистування людям із вадами зору; обґрунтовано функціональні залежності складових модулів розробленої моделі; проаналізовано базові модулі запропонованої моделі системи. Упроваджені такі **методи**: функціональне моделювання, системний аналіз. Досягнуті **результати**: запропоновано функціональну модель системи інтелектуального асистування для людей з вадами. Ця система має переваги, якщо порівнювати з наявними аналогами за низкою функціональних можливостей: детектування статичних та динамічних перешкод із передбачення траєкторії руху динамічних перешкод, здатність роботи в різних умовах (у приміщенні, на вулиці, у світлий або темний час, за різних погодних умов), підтримка інтеграції інших систем та пристроїв, високий рівень автономності. **Висновки**: розробленій моделі системи властиві підвищена автономність, інтеграція з іншими пристроями та системами, здатність аналізувати динамічні об'єкти та передбачувати траєкторію їх руху, а також забезпечувати різноманітний зворотний зв'язок з користувачем.

Ключові слова: система; асистент; зір; LiDAR; відео; траєкторія; прогнозування; інтелектуальна система; розпізнавання; класифікація.

Бібліографічні описи / Bibliographic descriptions

Барковська О. Ю., Сердечний В. С. Узагальнена функціональна модель системи асистування людям із вадами зору. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 6–16. DOI: https://doi.org/10.30837/2522-9818.2024.28.006

Barkovska, O., Serdechnyi, V. (2024), "Intelligent assistance system for people with visual impairments", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 6–16. DOI: https://doi.org/10.30837/2522-9818.2024.28.006

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 004. 896
DOI: https://doi.org/10.30837/2522-9818.2024.2.017

І. Бінько, В. Шевель, Д. Крицький

КОМПЛЕКСНИЙ ПІДХІД ДО УПРАВЛІННЯ ФОРМУВАННЯМ ГРУПИ РОБОТІВ

Предмет дослідження – розроблення методів управління роями безпілотних літальних апаратів (БпЛА) за моделлю "провідничий – ведений". Вивчення наявних класифікацій та взаємодій між безпілотними апаратами в різних формаціях, таких як групи, зграї, асоціації та рої, з метою створення ефективної системи управління. **Мета роботи** – покращення якості взаємодії між безпілотними літальними апаратами за моделлю "провідничий – ведений" під час виконання польотної місії унаслідок постійного контролю між об'єктами. Забезпечення надійного виконання польотних місій способом упровадження нових методів управління, що беруть до уваги різні режими взаємодії між апаратами. **Завдання:** проаналізувати класифікацію наявних БпЛА; дослідити параметри та модель взаємодії безпілотних літальних апаратів у групах, зграях, асоціаціях, роях; створити сценарій взаємодії двох БпЛА за моделлю "провідничий – ведений"; розробити програму для візуалізації польоту безпілотних літальних апаратів за моделлю "провідничий – ведений"; випробувати політ за запропонованою моделлю на етапах, де є різні геопросторові об'єкти. **Методи:** моделювання для розроблення підсистеми візуалізації польоту БпЛА; графічне моделювання для створення моделі безпілотного літального апарата типу літак; теорія алгоритмів для розроблення сценарію взаємодії двох БпЛА; використання спеціалізованих програмних засобів для візуалізації та симуляції поведінки безпілотних літальних апаратів в умовах реального часу. **Результати:** розроблено класифікацію безпілотних літальних апаратів; створено графічну модель літака *Mini-Flight-M*; запропоновано схему взаємодії двох БпЛА в режимах "учитель" або "наставник"; створено програму для візуалізації польоту БпЛА за моделлю "провідничий – ведений"; випробувано політ за запропонованою моделлю на етапах, де є різні геопросторові об'єкти. Результати підтвердили ефективність розробленої моделі та показали можливість її застосування в різних сферах, зокрема екологічний моніторинг, рятувальні операції та інші автономні місії. **Висновки.** Запропонований підхід до управління роєм БпЛА за моделлю "провідничий – ведений" дає змогу покращити якість взаємодії між апаратами та забезпечити надійне виконання польотних місій. Подальші дослідження мають зосередитися на оптимізації енергоспоживання та забезпеченні надійного зв'язку між агентами рою. Також важливо розробити методи захисту роїв БпЛА від кібератак та інших загроз, щоб підвищити їх стійкість і надійність під час виконання складних місій.

Ключові слова: безпілотні літальні апарати; рій; модель "провідничий – ведений"; взаємодія; моделювання; візуалізація.

Вступ

Зацікавлення в робототехнічних системах зросло завдяки їх успішному застосуванню в різних сферах діяльності. З розвитком технологій управління роями БпЛА виникає потреба у створенні ефективних систем, які б забезпечували надійне та безпечне виконання польотних місій. Управління роями літальних апаратів є складним завданням, що передбачає розв'язання проблем комунікації, адаптації до мінливих умов середовища, оптимізації енергоспоживання та забезпечення безпеки системи від кібератак. Це викликає необхідність у розробленні ефективної системи керування роботами, що дозволяла б також забезпечувати взаємодію в рою, у покращенні якості взаємодії між БпЛА за моделлю "провідничий – ведений" під час виконання польотної місії за допомогою постійного контролю між об'єктами. Це досягається завдяки впровадженню нових методів управління, що беруть до уваги різні режими взаємодії між апаратами та забезпечують надійне виконання польотних місій. Управління на основі розподіленої системи передбачає, що одна людина може керувати групою роботів та надсилати команди на виконання складних завдань. Кожен робот обладнаний спеціальним технічним засобом для забезпечення зв'язку з наземною станцією та іншими роботами у формації. На відміну від централізованого, розподілений метод передбачає можливість передавати інформацію про кожен безпілотний літальний апарат назад до терміналу призначення завдань, зокрема декомпозицію та розподіл завдань. Унаслідок цього термінал призначення завдань виконує функцію міжмережного з'єднувача, що дає змогу групі БпЛА формувати мережну структуру, сприяючи обміну інформацією,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

координації завдань та вирішенню конфліктів між БпЛА.

Різні сучасні безпілотні апарати використовуються як у військовій, так і в громадській сферах життя, і щороку їх функціональні можливості покращуються та доповнюються (рис. 1). Іноді необхідно застосовувати кілька груп БпЛА для покриття більшої території або отримання декількох точок зору. Сукупність БпЛА, які використовуються одночасно для виконання конкретного завдання, називається роєм. Такі безпілотні літальні апарати працюють разом і передають своє положення та іншу корисну інформацію з огляду на заздалегідь визначені часові інтервали. Просторове розташування БпЛА один щодо одного в просторі є ключовим елементом для їх взаємодії [1]. Використання груп і комплексів малогабаритних безпілотних літальних апаратів може значно розширити сферу їх застосування. З низкою проблем, що ускладнюють застосування малогабаритних БпЛА, можна впоратися за допомогою групового принципу. Зокрема колективне використання безпілотних літальних апаратів актуальне в геодезії, географії, дозиметрії, безпеці, розвідці, у пошуку зниклих людей тощо.

Рис. 1. Класифікація БпЛА

Режим призначення завдань для групи БпЛА має помітні переваги, зокрема високий рівень успішності та стійкість до несподіваних подій під час виконання завдань. Крім того, завдяки попередньо визначеним цільовим завданням для кожного БпЛА споживання енергії та інші витрати можуть залишатися відносно низькими в процесі польоту. Під час виконання завдань кожен БпЛА також має гнучкість відрегулювати свою цільову траєкторію залежно від фактичних умов польоту. Отже, цей метод виявляє високу стійкість і значно підвищує рівень виконання завдань. Проте певні завдання все ще обмежені деякими обставинами. Ці обмеження зазвичай виникають через такі проблеми, як тривалість польоту, обчислювальне навантаження та складність місії. Наприклад, у разі продовженої тривалості польоту батареї БпЛА можуть бути обмежені та не забезпечувати достатньої енергії для завершення завдання. Обчислювальне навантаження також може бути перевищеним, особливо в разі складних обчислювальних завдань, що вимагають значних ресурсів.

У цій роботі використано методи моделювання та симуляції, зокрема розроблено програму для візуалізації польоту БпЛА, написану мовою сценаріїв *Lua* та із застосуванням інформаційної системи *CoppeliaSimEDU*. Ця програма дає змогу відтворити сцену та модель поведінки двох БпЛА, де один літальний апарат підпорядковується іншому – лідеру рою.

Запропонований підхід до керування роєм БпЛА покращує якість взаємодії між апаратами та забезпечує надійне виконання польотних місій. Результати випробувань підтвердили ефективність розробленої моделі та довели можливість її застосування в різних сферах, таких як екологічний моніторинг, рятувальні операції та інші автономні місії.

Аналіз останніх досліджень і публікацій

Сучасні дослідження в царині безпілотних літальних апаратів та їх застосування в групових системах стрімко розвиваються, зосереджуючись на різних підходах до управління роями для підвищення їх ефективності, автономності та надійності.

Існують два типи роїв за методами керування. Перший тип містить пристрої одного виду, що призначені для виконання одного спільного завдання

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

та діють як розподілений об'єкт. Такий рій можна вважати самостійно організованою системою, властивою для природних утворень (подібно до поведінки комах, птахів, риб, об'єднаних у рої чи зграї, які обмінюються інформацією та виконують спільне завдання, використовуючи колективний інтелект). Для такого рою не існує централізованої системи контролю за поведінкою кожного індивіда. Локальні та досить випадкові взаємодії призводять до глобальної ройової поведінки, яка не може контролюватися окремими агентами. У цьому разі маємо справу з багатоагентною системою, якій властива самоорганізована поведінка і яка в сукупності має демонструвати "розумну" поведінку.

Другий тип рою містить апарати, що мають неоднакове навантаження та виконують різні функції в межах загального завдання. Багатоагентний метод керування роєм передбачає наявність керування, і команди можуть надходити як від системи управління, розташованої поза роєм, так і від "призначеного" ("локального") лідера всередині рою. Лідер передає та виконує команди, що надходять від керівного центру. Інші агенти діють, дотримуючись простих правил [2].

Три можливі стратегії управління.

1. *Централізована*: дистанційне керування з виділеною базовою станцією, лідер рою призначається із центрального вузла.

2. *Децентралізована*: лідер рою визначається на основі будь-якого алгоритму і не залежить від центральної станції управління.

3. *Змішана*: поєднує переваги централізованої та децентралізованої стратегій, призначаючи лідера рою в разі необхідності на основі одного з алгоритмів із передачею прав керування оператору [3].

Сучасні дослідження з використання ройового інтелекту БпЛА здебільшого спрямовані на виконання військових завдань. Напрями досліджень майбутніх інтелектуальних ройових систем присвячені адаптивним можливостям автономії БпЛА й роботі у великих неоднорідних командах інтелектуальних агентів.

Розглянемо кілька проєктів, пов'язаних із роєм БпЛА.

- *CODE* – програма, спрямована на подолання обмежень на масштаб і рентабельність операцій БпЛА способом побудови співпраці та спільної автономії. Розробники намагаються створити відкриту, модульну архітектуру, стійку до обмежень пропускної здатності та проблем підключення [4].
- *Perdix* – система рою БпЛА, розроблена Управлінням стратегічних можливостей Міністерства оборони США. БпЛА *Perdix* не запрограмований заздалегідь – це колективний організм, що має розподілений мозок для прийняття рішень і адаптації один до одного. Весь рій не має лідера – кожен БпЛА *Perdix* "спілкується" та "співпрацює" з усіма пристроями, тому може адаптуватися до будь-яких змін у польоті [5].
- *LOCUST* – технологія виробництва дешевих автономних ройових БпЛА, яка дає змогу швидко запустити до 30 безпілотних літальних апаратів у повітря, що об'єднуються в інтелектуальні мережі. Керування БпЛА, з'єднаними між собою за допомогою адаптивної бездротової мережі, здійснюється наземним оператором [6].
- *Gremlins* – технологія, призначена для проведення електронної розвідки та придушення засобів протиповітряної оборони (ППО). Після виконання своєї місії літальні апарати мають повернутися до носія, де вони встановлюються на борт літака за допомогою спеціального обладнання. Серед технологій можна виокремити високошвидкісне цифрове управління польотом [7].
- *OFFSET* – технологія, що забезпечує можливості створення рою для малих міських наземних підрозділів. Серед особливостей – автономія рою, покращений інтерфейс, який дає змогу користувачам відстежувати та керувати багатьма безпілотними платформами [8].

Вітчизняні дослідження спрямовані на планування траєкторії БпЛА в групі без можливості адаптації формації та забезпечення автономності окремого апарата [9, 10].

Більшість моделей, використаних у розроблених проєктах, містять пристрої одного типу, що виконують одне спільне завдання та не мають централізованої системи контролю поведінки. Тому вирішено звернути увагу на БпЛА за другим типом управління, коли команди надходять від "призначеного" лідера. За допомогою використання інформаційної системи *CoppeliaSimEDU* створено програму, що відтворює сцену й модель поведінки двох БпЛА, де один літальний апарат підпорядковується іншому – лідеру рою.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Визначення не розв'язаних раніше частин загальної проблеми. Мета роботи й завдання

Незважаючи на значний прогрес у дослідженні та розробленні систем управління роями безпілотних літальних апаратів, існує низка нерозв'язаних питань, що потребують подальшого вивчення. Забезпечення надійного зв'язку між агентами рою в умовах реального часу є складним завданням, адже відсутність централізованого контролю та можливість втрати зв'язку здатні призвести до субоптимальних рішень та збоїв у роботі системи. Сучасні системи часто не можуть швидко адаптуватися до мінливих умов середовища, що обмежує їх ефективність у складних і невизначених ситуаціях. Більшість досліджень не приділяє достатньої уваги оптимізації енергоспоживання в роях БпЛА, що є критично важливим для тривалих місій. Наявні моделі часто не зважають на всі можливі варіанти взаємодії та поведінки агентів у рою, що, імовірно, спричинить неточні прогнози та помилкові рішення. Питання безпеки та захисту роїв БпЛА від кібератак та інших загроз залишаються недостатньо вивченими.

Метою роботи є покращення якості взаємодії між безпілотними літальними апаратами за моделлю "провідничий – ведений" під час виконання польотної місії з допомогою постійного контролю між об'єктами. Це досягається дослідженням і розробленням нових методів керування роями БпЛА з огляду на розподілені системи управління, що забезпечують високу оперативність і надійність виконання завдань у динамічних умовах.

Для досягнення поставленої мети необхідно виконати такі завдання: проаналізувати класифікацію БпЛА; детально дослідити сучасні підходи до класифікації безпілотних літальних апаратів, зважаючи на їх призначення та технічні характеристики; вивчити параметри та моделі взаємодії безпілотних літальних апаратів у наявних групах, зграях, асоціаціях, роях; розробити сценарій взаємодії двох БпЛА за моделлю "провідничий – ведений" з можливістю вибору режимів керування "вчитель" або "наставник"; створити програму для візуалізації польоту безпілотних літальних апаратів за моделлю "провідничий – ведений"; виконати випробування польоту за запропонованою моделлю на етапах, де є різні геопросторові об'єкти.

Для реалізації цих завдань упроваджуються такі методи: моделювання для розроблення підсистеми візуалізації польоту БпЛА, графічне моделювання для створення моделі безпілотного літального апарата типу літак, а також методи теорії алгоритмів для розроблення сценарію взаємодії двох БпЛА.

Об'єктом дослідження є модель БпЛА за схемою "провідничий – ведений", а предметом – процес імітації польоту безпілотних літальних апаратів за схемою "провідничий – ведений" із можливістю вибору режимів керування БпЛА "вчитель" або "наставник". Досягнуті результати передбачають класифікацію безпілотних літальних апаратів, графічну модель літака *Mini-Flight-M*, схему взаємодії двох БпЛА в режимах "вчитель" або "наставник", програму для візуалізації польоту БпЛА за моделлю "провідничий – ведений", а також результати випробувань польоту за запропонованою моделлю на етапах, де є різні геопросторові об'єкти.

Розв'язання перелічених завдань дасть змогу підвищити ефективність та надійність роїв БпЛА, що відкриє нові можливості для їх використання в різних сферах, зокрема екологічний моніторинг, рятувальні операції та інші автономні місії.

Матеріали й методи

У цьому розділі досліджено різні архітектури систем управління, що застосовуються для координації дій роботів у формаціях. Розглянуто три основні типи архітектур: централізовані, децентралізовані та розподілені. Кожен із цих підходів аналізується з погляду їх переваг, обмежень та придатності для особливого застосування (табл. 1).

Централізована архітектура

У цій моделі управління всі керувальні функції концентруються в одному центральному контролері. Це дозволяє детально аналізувати та оперативно керувати поведінкою кожного агента в системі. Дослідження науковців з MIT показало ефективність цієї архітектури в управлінні дронами для екологічного моніторингу, що забезпечило високу точність координації та адаптацію до мінливих умов.

Розподілена архітектура

У цьому разі кожен агент оснащений власним контролером, що дозволяє виконувати обчислення на місці на основі інформації від сусідніх агентів. Цей метод забезпечує високий рівень автономності

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

та надійності, оскільки система не залежить від одного центру контролю. Застосування технологій машинного навчання та сучасних засобів комунікації, наприклад *URLLC* в мережах 5G, покращує цю архітектуру, забезпечуючи ефективність в динамічних середовищах.

Децентралізована архітектура

Контрольні функції поділяються між декількома незалежними підсистемами. Це підвищує гнучкість системи та спрощує масштабування, але також створює виклики з координацією між підсистемами. Активне дослідження та розроблення нових методів і технологій необхідні для покращення координації та згуртованості агентів у таких системах.

Архітектури управління формацією є фундаментальними для координації та ефективної взаємодії між роботами або агентами в багатьох сферах – від автоматизованого виробництва до автономних транспортних систем. Основні типи архітектур управління формацією передбачають централізовані, децентралізовані та розподілені підходи, кожен з яких має переваги та обмеження залежно від особливостей застосування та середовища використання.

Таблиця 1. *Порівняння характеристик централізованих, розподілених і децентралізованих архітектур управління*

Критерій	Централізована архітектура	Розподілена архітектура	Децентралізована архітектура
Затримка в передачі інформації	Низька, якщо вузли розташовані біля контролера	Залежить від місцевих з'єднань та швидкості оброблення	Залежить від з'єднань між підсистемами
Масштабованість	Обмежена через залежність від одного контрольного центру	Висока, агенти можуть додаватися без утручання центру	Висока, але може виникати складність у координації
Надійність	Залежить від єдиної точки збою (центральний контролер)	Висока, збій одного агента мало впливає на інших	Висока, але потребує ефективного взаємозв'язку між підсистемами
Стійкість до помилок	Низька, помилка в центрі впливає на всю систему	Висока, система може продовжувати роботу в разі збоїв окремих агентів	Висока, але вимагає додаткових механізмів синхронізації та відновлення

1. Централізована архітектура управління формацією

Централізована архітектура становить один з основних підходів у системах управління формаціями роботів і передбачає зосередження всіх контрольних функцій у єдиному центральному контролері. Такий підхід дає змогу комплексно аналізувати стан кожного агента та оперативно реагувати на динаміку змін у формації, використовуючи загальнодоступну інформацію для визначення оптимальних стратегій дій.

У 2021 р. науковці з MIT розробили централізовану систему управління для координації групи автономних дронів, задіяних у складних місіях зі збору екологічних показників. Ця система дала змогу досягти високої точності в плануванні маршрутів і синхронізації дій дронів, що є критично важливим для адаптації до динамічних атмосферних умов, з якими вони стикаються під час виконання місій [11, 12].

Централізована архітектура має кілька переваг, зокрема здатність ефективно керувати складними операціями та оптимізувати загальну активність завдяки глобальному огляду станів усіх агентів. Однак вона також має недоліки, особливо у сфері масштабування через високі вимоги до обчислювальних і комунікаційних ресурсів. Крім того, залежність від центрального комп'ютера призводить до єдиної точки збою, що може спричинити перебої в роботі всієї системи в разі виникнення проблем.

Централізована архітектура залишається важливою там, де потрібна висока точність координації та здатність швидко реагувати на зміни в складних технічних і природних умовах. Однак важливо зважати на потенційні ризики й виклики, що вимагають додаткових заходів безпеки та резервування систем.

2. Розподілена архітектура управління формацією

Розподілена архітектура управління передбачає, що кожен агент оснащений вбудованим контролером, який обчислює сигнал управління на основі його поточного стану та станів сусідніх агентів. Такий підхід дозволяє здійснювати обмін інформацією за допомогою камер, *LIDAR*-ів або прямої комунікації між агентами, що забезпечує автономність кожного з них.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Наведемо переваги розподіленої архітектури.

– Легкість масштабування: взаємодія між агентами обмежується їх найближчими сусідами, що спрощує залучення нових агентів до системи.

– Надійність: система є стійкою до збоїв, оскільки не залежить від централізованого контролю. Помилка одного агента не спричиняє колапсу всієї системи.

У дослідженнях, присвячених управлінню БпЛА, розподілена архітектура дає змогу багатороторним дронам ефективно формувати польотні формації. Наприклад, експерименти з управління на основі консенсусу показують, як літальні апарати можуть самостійно адаптуватися до динамічних умов та координувати свої дії без зовнішнього втручання [13].

Дослідження 2023 року продемонстрували використання машинного навчання для прогнозування та адаптації поведінки агентів у формації. Ці технології забезпечують вищу ефективність і оперативність системи, зокрема в автономних автомобілях, де також розв'язуються проблеми толерантності до збоїв, затримки реакцій та управління ресурсами [14].

Розвиток технологій зв'язку, зокрема впровадження *URLLC* у мережах 5G, покращив можливості розподілених систем, сприяючи швидшому обміну інформацією та більш синхронізованим діям між агентами.

Сучасні дослідження та розроблення в галузі розподіленої архітектури управління продовжують розширювати можливості систем, покращуючи їх ефективність у динамічних середовищах. Застосування новітніх технологій та інноваційних підходів дає змогу цим системам більш ефективно виконувати складні завдання, збалансувавши локальну автономію із загальносистемною згуртованістю.

Розподілені архітектури управління все ще стикаються з викликами в складних середовищах, де важливо керувати численними локальними взаємодіями та забезпечувати послідовну поведінку агентів. Відсутність централізованої системи моніторингу може призвести до субоптимальних формацій.

3. Децентралізована архітектура управління

Децентралізована архітектура управління поділяє контрольну систему на незалежні підсистеми, кожна з яких має свій контролер. Такий підхід передбачає, що комунікація між підсистемами здійснюється зазвичай за допомогою механічних з'єднань, як, наприклад, фізичні пружини. Контролер кожної підсистеми має доступ лише до інформації про стани агентів у своїй власній підсистемі.

Ця конфігурація не ефективна для керування формаціями, оскільки контролер кожної підсистеми не має інформації про агентів інших підсистем. Це може ускладнити синхронізацію та координацію між різними підсистемами, необхідними для створення єдиної та згуртованої формації агентів.

Наприклад, у дослідженнях застосування децентралізованого керування багатоагентними системами, такими як БпЛА, увага зосереджується насамперед на методах управління на основі консенсусу [15]. Такі методи дають змогу кожній підсистемі взаємодіяти з іншими для досягнення спільних цілей без необхідності центрального управління, покращуючи ефективність та адаптивність під час виконання місій у динамічних умовах.

Крім того, децентралізовані механізми керування також вивчаються у контексті автономних автомобілів. У цих дослідженнях розглядаються технологічні та регулятивні аспекти, що впливають на галузь. Обговорюються виклики, пов'язані з децентралізованими системами в динамічних умовах, та їх наслідки для політики й управління.

Однак децентралізована архітектура стикається з викликами, зокрема з обмеженою можливістю координації між підсистемами, що, імовірно, спричинить проблеми із згуртованістю та синхронізацією. Відсутність центрального контролю може ускладнити моніторинг і реагування на непередбачені зміни в середовищі, що особливо важливо в оперативних сценаріях із високим рівнем ризику та невизначеності.

Необхідність подальшого вдосконалення децентралізованих систем і розроблення нових методів для підвищення їх ефективності, згуртованості та адаптивності є критичною. Це передбачає інтеграцію передових алгоритмів машинного навчання та розвиток комунікаційних технологій, що підтримують більш надійний та ефективний обмін інформацією між підсистемами.

Сучасні дослідження присвячені розробленню ефективніших протоколів передачі даних, які мінімізують час затримки. Ці методи передбачають використання алгоритмів для динамічного вибору маршрутів у мережі, основаних на поточних умовах зв'язку та обсягах трафіку.

Застосування алгоритмів машинного навчання для аналізу та прогнозування навантаження на мережу допомагає адаптувати систему управління

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

для оптимізації розподілу ресурсів і зменшення загальних затримок у системі.

Розроблення та впровадження більш потужних і ефективних комунікаційних модулів, таких як передові WiFi-модулі, технології 5G та системи на основі Li-Fi, забезпечує високу швидкість передачі інформації та зниження втрат сигналу.

Оптимізація апаратного забезпечення: інтеграція оптимізованого апаратного забезпечення, яке підтримує швидкісні комунікації та має високу обчислювальну потужність, зменшує час оброблення інформації та покращує загальну ефективність системи.

Зазначені техніки та технології можуть бути широко застосовані – від автономних транспортних засобів і дронів до промислових роботизованих систем, де швидка й надійна комунікація є ключовим фактором для забезпечення ефективності та безпеки оперативних процесів.

Останнім часом значно зросла увага до методів управління формуванням роботів, які важливі для розширення можливостей мультиагентних систем. Особливо це стосується використання розподілених стратегій, що дають змогу роботам динамічно адаптуватися до змін у середовищі без централізованого керівництва. Цей напрям передбачає різні стратегії управління та архітектурні рішення для мережних взаємодій між роботами.

Існує три методи до керування формацією: "лідер – послідовник", поведінковий метод і віртуальна структура.

1. Метод "лідер – послідовник"

У цьому підході роль лідера призначається одному агенту. Лідер прямує шляхом, заздалегідь визначеним планувальником місії, а інші роботи намагаються утримувати бажану формацію, рухаючись за ним (рис. 2). Існують два типи стратегії "лідер – послідовник": режим лідера, де послідовники безпосередньо підтримують формацію з лідером, та режим фронту, де кожен агент прямує за наступним агентом до досягнення лідера. Основні переваги цього підходу – простота в освоєнні та реалізації, а також прямолінійність аналізу стабільності. Однак цей метод має недоліки, зокрема залежність від одного робота для підтримки формації та його централізована природа, що ускладнює масштабування.

Цей підхід є одним із найпоширеніших у керуванні формацією роботів. Лідер визначає траєкторію руху, а інші роботи налаштовують свої дії, щоб утримувати задану відстань та орієнтацію щодо лідера. Сучасні дослідження впроваджують адаптивні методи й глибоке навчання для підвищення результативності цього підходу в складних умовах, наприклад підводне середовище, де звичайні засоби навігації можуть бути неефективними. Також розроблено розподілені оцінювачі, які дають змогу кожному послідовнику визначати стан лідера, використовуючи лише локальну інформацію, що зменшує залежність від централізованої комунікації. Консенсусні алгоритми в системах "лідер – послідовник" для мультиагентних систем із нелінійними властивостями дозволяють агентам синхронізувати свої стани відповідно до моделі лідера, забезпечуючи точнішу координацію та вищу стабільність формації. Це особливо важливо в умовах, де стандартні методи керування можуть не впоратися з комплексністю динамічного середовища [16]. Адаптивна безпечна система управління для формацій "лідер – послідовник" неголономних мобільних роботів в умовах невизначеності та потенційних кібератак забезпечує високий рівень надійності та безпеки. Система адаптує параметри управління для компенсації можливих збоїв інформації від лідера. Алгоритми основані на методах Ляпунова для аналізу стабільності та забезпечення безпеки, що підвищує загальну робастність системи й дає змогу ефективно протистояти зовнішнім загрозам і внутрішнім помилкам у системі керування [17].

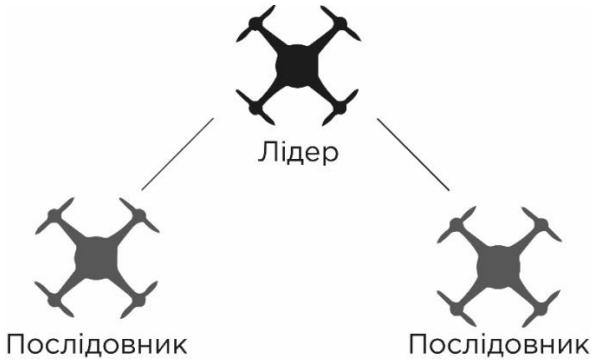


Рис. 2. Стратегія лідера в системах "лідер – послідовник"

2. Віртуальна структура

Підхід із використанням віртуальних структур для управління формацією роботів має вагоме значення в сучасній робототехніці. Цей метод допомагає створювати складні формації без необхідності фізичних з'єднань між роботами, використовуючи заздалегідь визначені правила

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

поведінки. У підході віртуальної структури всі роботи виконують роль віртуальних вузлів у заздалегідь заданій структурі. Це означає, що кожен робот має особливу траєкторію руху, синхронізовану з траєкторіями інших роботів для створення єдиної, цілісної формації. Цей метод забезпечує високу точність позиціонування та є дуже ефективним у стабільних умовах. Сучасні розробки використовують адаптивні та інтерактивні технології для підвищення гнучкості віртуальних структур. Наприклад, застосування слайдинг-мод контролю дозволяє роботам швидко адаптуватися до змін у динаміці оточення, мінімізуючи вплив затримок у відповідях та коливань у даних сенсорів [18]. Це покращує стійкість формації до зовнішніх перешкод і динамічних умов. Однак підхід із віртуальною структурою має деякі обмеження. Високі вимоги до обчислювальних ресурсів та необхідність точної синхронізації можуть створювати труднощі у швидко змінюваних умовах. Проблеми синхронізації особливо критичні, коли збільшується кількість роботів у формації. Дослідження покращення алгоритмів керування та розроблення нових моделей зворотного зв'язку можуть зменшити зазначені недоліки. Вивчення нових методів адаптації та оптимізації в умовах невизначеності також є важливим напрямом для подальшого розвитку технологій. Це забезпечить підвищену адаптивність і масштабованість віртуальних структур, даючи їм змогу бути ефективнішими в складних і динамічних середовищах.

Зосередження уваги на моделях машинного навчання, які можуть прогнозувати й компенсувати потенційні проблеми перед тим, як вони вплинуть на стабільність формації, також відкриває нові можливості для покращення ефективності робочих груп роботів.

3. Метод на основі поведінки

Поведінковий підхід у керуванні формацією роботів грунтується на принципах поведінки зграї тварин (рис. 3). Кожен робот налаштований діяти за простими локальними правилами, такими як уникнення зіткнень, узгодження швидкостей та підтримання когезії групи, завдяки чому формація динамічно адаптується до змін у середовищі без потреби в централізованому керуванні. Зазначений підхід може зіткнутися з обмеженнями у формуванні складних конфігурацій, тому сучасні дослідження спрямовані на розроблення алгоритмів, що реалізують складніші взаємодії та управління, внаслідок чого покращується адаптивність і масштабованість системи.

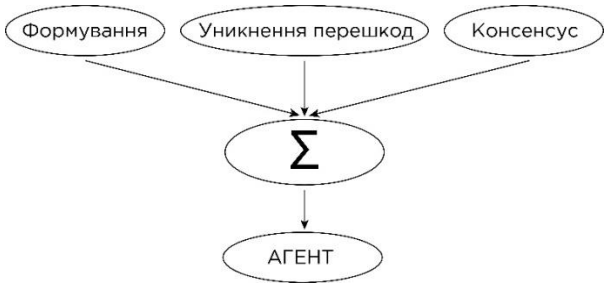


Рис. 3. Поведінковий підхід до формування групи

Найвідомішу роботу, яка вплинула на багатьох дослідників окресленої царини, запропонував Крейг Рейнольдс 1987 року. Автор висунув три евристичні правила, яких кожен агент має дотримуватися для збереження бажаної формації: центрування зграї, узгодження швидкостей та уникнення зіткнень. Зазначені правила допомагають агентам залишатися близько один до одного, вирівнювати свої швидкості та уникати зіткнень, використовуючи привабливі та відштовхувальні сили. Особливо важливі сучасні студії, які розширюють застосування поведінкових правил у більш складних і динамічних середовищах. Дослідження показує, як алгоритми глибокого навчання можуть оптимізувати рух і управління формацією в складних умовах підводного середовища, де стандартні навігаційні методи часто не ефективні [19]. Розподілене керування формацією мобільних роботів на основі біоінспірованої нейродинаміки та адаптивного ковзного інноваційного фільтра описано в роботі [20]. Автори пропонують біоінспіровані алгоритми, що використовують нейродинамічні моделі для забезпечення координації між роботами. Це сприяє покращенню реакції роботів на зміни умов довкілля, підвищує ефективність місії в складних сценаріях. Поведінкові підходи в керуванні формацією роботів продовжують розвиватися, пропонуючи гнучкі та адаптивні рішення для динамічних умов. Централізація управління відступає на другий план на користь розподілених і самоорганізованих систем, що можуть масштабуватися та адаптуватися до труднощів реального світу. Упровадження глибокого навчання та нейродинаміки є важливим кроком у підвищенні автономності роботів.

Для реалізації поставлених завдань було застосовано методи моделювання та симуляції.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Програма, розроблена для візуалізації польоту безпілотних літальних апаратів за моделлю "провідничий – ведений", написана мовою сценаріїв *Lua* та використовує інформаційну систему *CoppeliaSimEDU*. Ця програма відтворює сцену й модель поведінки двох БпЛА, де один літальний апарат підпорядковується іншому – лідеру рою.

Структура моделі візуалізації польоту містить такі елементи:

1) камери, що дають змогу бачити сцену;
2) світлові об'єкти для освітлення сцени;
3) середовище з такими властивостями, як навколишній світ, туман, колір фону тощо;
4) ґрунт, що містить об'єкти, згруповані в модель, зокрема декілька об'єктів дерев;
5) моделі безпілотних літальних апаратів *UAV-M* і *UAV-S*, розроблені в системі *SolidWorks* та імпортовані до наявної сцени.

Режими роботи програми

У програмі є декілька режимів роботи: *RP* та *RPYT*. У режимі викладача (*RP*) необхідні два пристрої введення: один для крену й висоти, другий – для інших функцій. У режимі наставника (*RPYT*) також необхідні два пристрої введення: один для крену, тангажу й тяги, другий – для інших функцій. Унаслідок використання функціональності перемикача можна змінювати параметри польоту.

Результати досліджень та їх обговорення

Одним із найбільш поширених підходів до формування управління є метод "лідер – послідовник", за умови якого "послідовник" визначає своє положення тільки щодо "лідера" (рис. 4). Ми припустили, що взаємодія між агентами має відбуватися на "глобальному" рівні замість того, щоб підлеглий реагував лише на інформацію локального рівня. Локальна інформація може бути отримана з параметрів контролера, тоді як глобальна інформація стосується положення об'єкта загалом щодо навколишнього середовища. Ці дві категорії джерел інформації не є взаємовиключними. У моделі використовувалася проста модель узгодження швидкості. Отже, це дослідження додатково вивчає гіпотезу про те, що зчеплення "прямування за лідером" може регулюватися лише глобальною інформацією без залучення локальних відомостей.

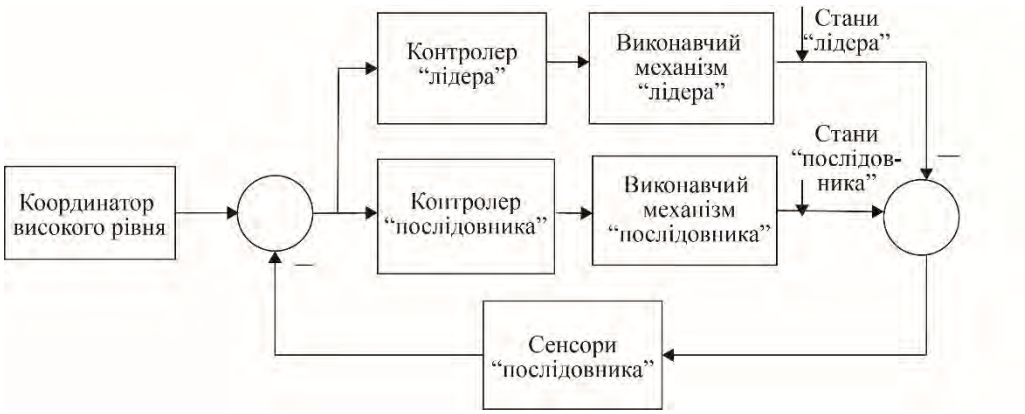


Рис. 4. Схема методу управління "лідер – послідовник"

У межах роботи запропоновано імітаційну модель, що координує рій відповідно до параметрів контролерів "лідера" та "послідовника" з огляду на збереження дистанції між БпЛА. За допомогою інформаційної системи *CoppeliaSimEDU* побудовано програму, що відтворює сцену й модель поведінки двох БпЛА, де один літальний апарат підпорядковується іншому – лідеру рою. Ініціатором візуалізації польоту є користувач. Програма написана мовою сценаріїв *Lua*, яка є розширеною мовою програмування та призначена для підтримки загального процедурного програмування.

Модель візуалізації польоту містить кілька елементів (рис. 5).

Сцена «1» містить такі елементи:

– кілька об'єктів камери, що дають змогу бачити сцену, якщо вони пов'язані з видом;

– кілька світлових об'єктів для освітлення сцени;

– кілька уявлень, які відтворюють те, що бачить камера;

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

– кілька сторінок, кожна з яких містить одне або кілька переглядів;

– середовище з такими властивостями, як навколишній світ, туман, колір фону тощо;

– ґрунт, що містить об'єкти, згруповані в модель, зокрема й декількох об'єктів дерев;

– моделі безпілотних літальних апаратів *UAV-M* і *UAV-S*, розроблені в системі *SolidWorks* та імпортовані до наявної сцени.

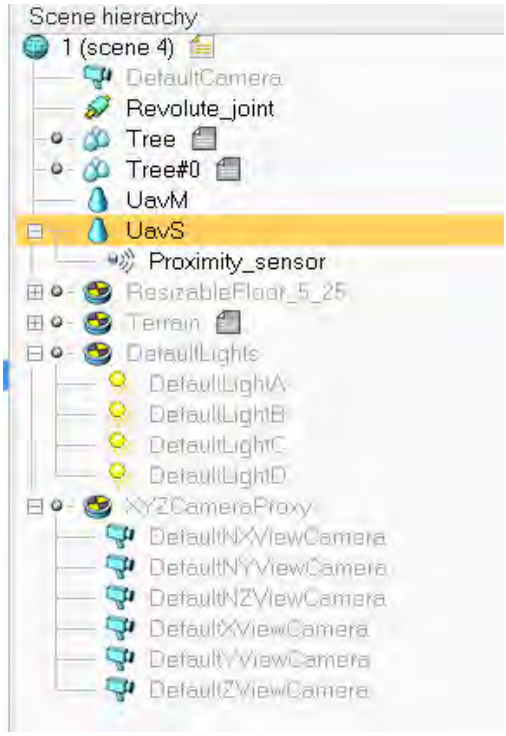


Рис. 5. Структура моделі візуалізації польоту

Одною з графічних моделей розроблених БпЛА для програми є *Mini-Flight-M* з такими тактико-технічними характеристиками (рис. 6):

– злітна маса до 90 кг;

– маса цільового та додаткового навантаження до 20 кг;

– тривалість польоту понад 30 хв;

– тяга двигуна 400–410 Н.

Є декілька режимів роботи в програмі: *RP* та *RPYT*.

У режимі викладача (*RP*) необхідні два пристрої введення: один для керування креном і висотою, а другий для виконання інших функцій. Використовуючи функціональність перемикача, відображеного в конфігурації, другий контролер також може змінювати крен і подачу.

Для роботи в режимі наставника (*RPYT*) також потрібні два пристрої введення: один для керування креном, тангажем і тягою, а другий для виконання інших функцій. Застосовуючи функціональність перемикача, відтворену в конфігурації, другий контролер також може змінювати крен, нахил, нишпорення й тягу.

Якщо під'єднано більше ніж один пристрій введення, можна переключитися на один з режимів. Спочатку необхідно обрати пристрій, що використовуватиметься для викладача, і зіставити його. Потім потрібно обрати пристрій для студента й також зіставити його. Після цього в нижній частині інтерфейсу користувача можна буде побачити відкриті пристрої та їх конфігурації.

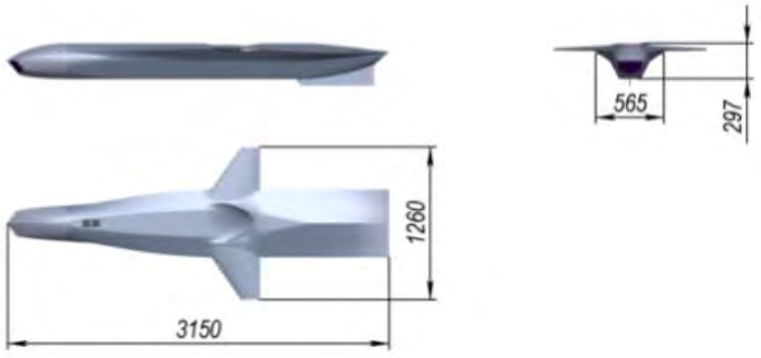


Рис. 6. Загальний вигляд ЛА *Mini-Flight-M*

Процес підготовки сцени передбачає три етапи:

1) створення або модифікація моделі полігону в *CoppeliaSimEDU*;

2) експорт моделі безпілотного літального апарата у формат *URDF*;

3) вказівка властивостей і параметрів усіх об'єктів у симуляторі.

Переходимо до налаштувань полігону. Спочатку все відключено для розрахунків – об'єкти не піддаватимуться дії сили тяжіння, не будуть

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

відскакувати один від одного та пролітати крізь інші об'єкти.

Сам полігон і підставки мають бути незміщеними, немов жорстко приклеєними до підлоги сцени, але в разі потрапляння на них ігрових об'єктів вони мають пружинити й на них потрібно звертати увагу. Тому необхідно вказати параметр *Body is respondable*, який означає, що на об'єкт потрібно зважати під час зіткнень. У моделюванні мають братися до уваги зіткнення з усіма об'єктами.

Для моделювання руху рою БпЛА на сцені було розміщено декілька моделей безпілотних літальних апаратів і відкориговані їх початкові позиції. Після додавання скрипта взаємодії з *Ros* та підскриптів моделювання керуючих програм було запущено моделювання поведінки команди БпЛА (рис. 7).

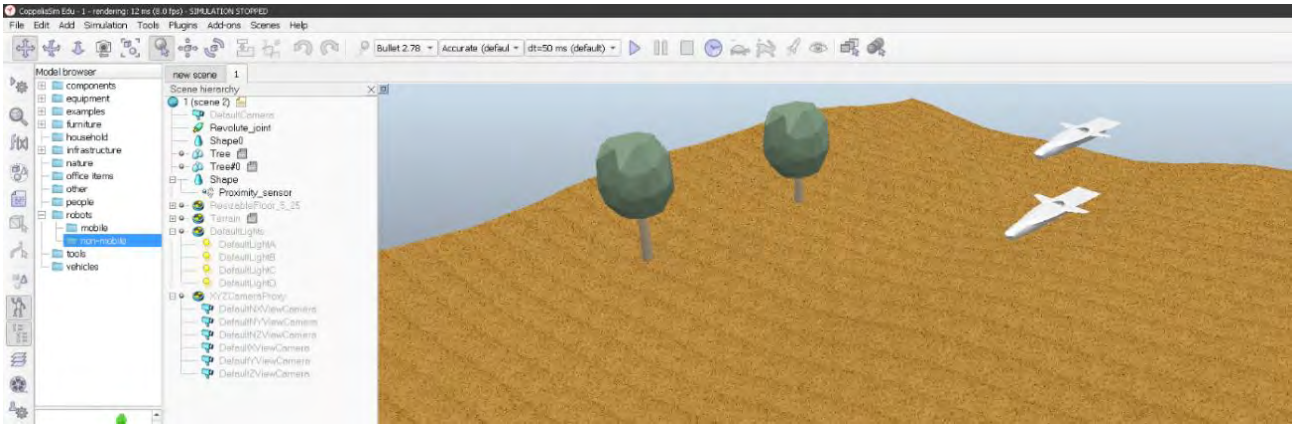


Рис. 7. Модель візуалізації польоту БпЛА за моделлю "провідничий – ведений"

Сукупність БпЛА, що використовуються одночасно для виконання конкретного завдання, називають роєм. Такі дрони працюють разом, повідомляючи про свою позицію та іншу корисну інформацію із заздалегідь визначеними інтервалами часу. Просторове розташування БпЛА один щодо одного є ключовим елементом для їх взаємодії [9]. Управління роєм – одна з найважливіших тем дослідження сукупної поведінки системи. Мета цього процесу полягає в тому, щоб регулярно й багаторазово розвертати та направляти літальні апарати на певній відстані один від одного та підтримувати досягнутий шаблон і надалі [21].

Як приклад руху рою дронів можна розглянути політ зграї птахів. Під час руху члени групи спілкуються на основі отриманої ними інформації, обмеженої зоною огляду. Згідно з дослідженнями поле зору становить 135° з бінокулярним перекриттям 20° (рис. 8) [10]. Такі особливості зумовлюють той факт, що члени зграї можуть рухатися один за одним у полі зору, відповідно створюючи лінії або *V*-подібну форму групи під час польоту.

На основі схеми польоту окремого об'єкта розглянемо алгоритм польоту рою за умови, що кожний БпЛА має обмежене поле зору, кут огляду становить 250°, а саме $-35° \le \theta \le 235°$.

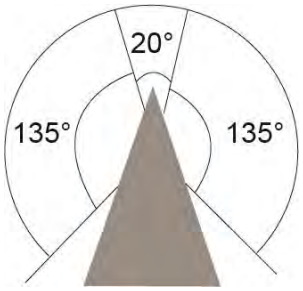


Рис. 8. Поле зору окремого БпЛА

Агент використовує неголономну модель руху, тому може рухатися за умови зміни руху основного об'єкта:

$$\begin{bmatrix} x_i(t+\Delta t) \\ y_i(t+\Delta t) \\ a_i(t+\Delta t) \end{bmatrix} = \begin{bmatrix} x_i(t) \\ y_i(t) \\ a_i(t) \end{bmatrix} + \begin{bmatrix} \cos a_i(t) & 0 \\ \sin a_i & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ \omega \end{bmatrix}, \qquad (1)$$

де (x_i, y_i, α_i) – декартові положення й провідна позиція агента i; v – лінійна швидкість у координатах кожного агента; ω – кутова швидкість. Вважатимемо, що всі члени рою можуть визначати відстань та кути огляду один одного щодо власних. l_{ij}, φ_{ij} – вимірювана відстань та кут нахилу агента j в діапазоні дії агента i, тоді

$$\begin{cases} x_{ij} = l_{ij} \cos \varphi_{ij} \\ y_{ij} = l_{ij} \sin \varphi_{ij} \end{cases}, \text{ де } -35^0 \le \varphi_{ij} \le 235^0, l_{ij} \in [0, R]. \qquad (2)$$

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Отже, для коректного формування рою БпЛА кожному агентові необхідно знаходити правильне місце для збереження відстані, кут руху щодо агента попереду, уникаючи перешкод та зіткнень один з одним. Комунікаційне оточення агента i – це його "колеги" в рої, розташовані в межах фіксованого радіуса R щодо нього:

$$\mathbb{N}(r_i) = \left\{ r_{j \in \mathrm{N} \neq i'} \left\| p_i - p_j \right\| \leq R \right\} , \quad (3)$$

де $\mathbb{N}$ – ділянка "комунікації"; N – кількість членів рою; r_i – агент i; p_i, p_j – положення в просторі агентів i та j; R – максимальний радіус взаємодії.

Під час руху рою довкілля може змінюватися, що зумовлює поділ єдиного рою на декілька суброїв. Рій зі змінними від часу характеристиками щодо агента i можна обчислити за формулою

$$S^t(r_i) = r_i \cup \left\{ r_{j \in N, j \neq i'} \left\| p_i^t - p_j^t \right\| \leq R \right\} . \quad (4)$$

Оскільки діапазон комунікації обмежений візуальним полем кожного об'єкта рою, маємо (рис. 9)

$$S_v^t(r_i) = r_i \cup \left\{ r_{j \in N, j \neq i'} \left\| p_i^t - p_j^t \right\| \leq R \wedge \varphi_{iy} \notin V \right\} . \quad (5)$$

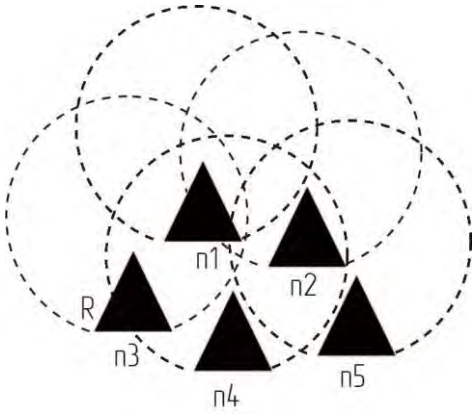


Рис. 9. Рій зі змінними від часу характеристиками та обмеженим візуальним полем

Щоб запобігати зіткненням із перешкодами та з метою огинання перешкод для нескладного рою можна використовувати спрощений алгоритм гістограми векторного поля (англ. *Vector Field Histogram, VFH*). Такий алгоритм визначає напрямок руху завдяки побудові гістограми векторного поля для уявлення полярної щільності перешкод (англ. *polar obstacle density, POD*). Поле сприйняття агента поділяється на n секторів із покриттям $360^\circ / n$. Тоді *POD* для кожного сектора

$$h^k(q_i) = \int_{\Omega k} P(p)^n \cdot \left(1 - \frac{d(qt_{i'} p)}{d_{\max}} \right)^m dp , \quad (4)$$

де $h^k(q_i)$ – полярна щільність перешкод у секторі k; $P(p)$ – імовірність того, що в точці p є перешкода; $d(q_{t_{i'}} p)$ – відстань від центра агента до точки p; d_{max} – максимальна дальність виявлення сенсорів; Ω_k – інтеграція.

$$\Omega_k = \left\{ p \in k \wedge d(q_{t_{i'}} p) < d_s \right\} . \quad (7)$$

Для утримання строю необхідно брати до уваги низьку обчислювальну складність та обмеження швидкості рою. З огляду на принцип *VFH* можна спроєктувати дії БпЛА з кутом огляду 250°, але в діапазоні $[-125°; 125°]$ з напрямком руху 0°. Унаслідок впливу сусідніх секторів згладжена полярна щільність перешкод за напрямком k становить:

$$p_k = \sum_{i=-1}^{l} w(i) f(k+i) , \quad (8)$$

$$f(k+1) = 1 - \frac{\min\left\{ d_s, d(k+i) \right\}}{d_s} , \quad (9)$$

де 1 – деяке додатне число для обчислення кожного напрямку $k \in [-a; a]$; $d(k+i)$ – відстань від центра агента до перешкоди за напрямком $k+i$; d_s – заздалегідь визначена безпечна відстань; $w(i)$ – вага сусідніх напрямків:

$$w(i) = \left\{ \begin{array}{c} \dfrac{l - |i| + 1}{\sum_{i=-1}^{l} (l - |i| + 1)'}, \quad -a \leq k + i \leq a \\ 0, \quad others \end{array} \right\} . \quad (10)$$

Такий вибір $w(i)$ забезпечує правильність тверджень: що далі сусідній напрямок від k, то менша його вага, і поточний напрямок руху $k(i=0)$ має найбільшу вагу.

Водночас необхідно брати до уваги, що для кожного окремого БпЛА рівняння руху матимуть такий вигляд:

$$\frac{dV}{dt} = \frac{1}{m(t)} \left[P \cos(\alpha + \phi_{ДВ}) - C_X \frac{\rho S}{2} V^2 \right] - g \sin\theta , \quad (11)$$

$$\frac{d\theta}{dt} = \frac{1}{m(t)} \left[\frac{1}{V} P \sin(\alpha + \phi_{ДВ}) - C_Y \frac{\rho S}{2} V^2 \right] - \frac{1}{V} g \cos\theta , \quad (12)$$

$$\frac{d\omega_Z}{dt} = \frac{1}{J_Z(t)} \left[M_Z + \Delta M_Z(\omega_Z) + \Delta M_Z(\varepsilon) \right] , \quad (13)$$

$$\frac{d\vartheta}{dt} = \omega_Z , \quad (14)$$

$$\frac{d\alpha}{dt} = \frac{d\vartheta}{dt} - \frac{d\theta}{dt} , \quad (15)$$

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

$$\frac{dH}{dt} = V \sin\theta\,, \tag{16}$$

$$\frac{dL}{dt} = V \cos\theta\,, \tag{17}$$

де

$$m(t) = m_0 - m_T сек \cdot t \cdot 0 < t < t_{ДВ.К}\,, \tag{18}$$

$$m(t) = m_0 - m_{сек} t_{ДВ} = const, \quad t_{ДВ.К} < t \le t_{сбр}\,, \tag{19}$$

$$m(t) = m_M = const, \quad t < t_{сбр}\,, \tag{20}$$

$$J_Z(t) = J_{Z0} - J_Z сек \cdot t\,, \tag{21}$$

$$0 < t \le t_{ДВ.К}\,, \tag{22}$$

$$J_Z(t) = J_{Z0} - J_Z сек \cdot t_{ДВ} = const, \quad t_{ДВ.К} < t \le t_{сбр}\,, \tag{23}$$

$$J_Z(t) = J_{ZM} = const, \quad t > t_{сбр}\,, \tag{24}$$

$$x_T = \frac{G_{M^x TM} + G_{0ДВ^x ТДВ - gm_{сек^x ТДВ} t}}{G_M + G_{0ДВ - gm_{сек} t}}, \quad 0 < t \le t_{ДВ.К}\,, \tag{25}$$

$$x_T = \frac{G_{M^x TM} + G_{0ДВ^x ТДВ - gm_{сек^x ТДВ} tДВ.К}}{G_M + G_{0ДВ - gm_{сек} tДВ.К}}, \quad t_{ДВ.К} < t \le t_{сбр}\,, \tag{26}$$

$$x_T = x_{TM} = const, \quad t > t_{сбр}\,. \tag{27}$$

Розглянута класифікація БпЛА дає змогу визначити основні завдання, що виконує рій. Запропонований підхід до управління роєм літальних апаратів з огляду на режим "провідничий – ведений" оснований на моделях польоту птахів, удосконалений завдяки впровадженню рівнянь руху кожного індивідуального агента. Такий підхід дозволяє прогнозувати траєкторію польоту кожного агента окремо та рою загалом.

Висновки

У процесі дослідження проаналізовано та розроблено методи управління роями безпілотних літальних апаратів за моделлю "провідничий – ведений". Розглянуті класифікації наявних БпЛА та параметри взаємодії в групах, зграях, асоціаціях та роях дали змогу визначити основні завдання, що виконує рій. Запропонований підхід до керування групою літальних апаратів зважає на режими "вчитель" або "наставник" для побудови рою. Розглянутий підхід оснований на моделях польоту птахів, удосконалений завдяки впровадженню рівнянь руху кожного індивідуального агента, що допомагає прогнозувати траєкторію польоту кожного агента окремо та рою загалом.

Створено програму для візуалізації польоту безпілотних літальних апаратів за моделлю "провідничий – ведений" та випробувано політ за запропонованою моделлю на етапах, де є різні геопросторові об'єкти. Результати підтвердили ефективність розробленої моделі та показали можливість її застосування в різних сферах, зокрема екологічний моніторинг, рятувальні операції та інші автономні місії.

Подальші дослідження мають бути зосереджені на оптимізації енергоспоживання та забезпеченні надійного зв'язку між агентами рою в умовах реального часу. Також важливо розробити методи захисту роїв БпЛА від кібератак та інших загроз, щоб підвищити їх стійкість і надійність під час виконання складних місій.

Список літератури

1. Jia G. W., Wang J. F. Research review of UAV swarm mission planning method. *Systems Engineering and Electronics.* 2021. Vol. 43. №. 1. P. 99–111. DOI: 10.3969/j.issn.1001-506X.2021.01.13
2. Do H. T. Formation control algorithms for multiple-uavs: a comprehensive survey / Do H. T. et al. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, *8*(27), Vol. 8(27):170230. 2021. DOI: 10.4108/eai.10-6-2021.170230
3. Na S., Niu H., Lennox B., Arvin F. Bio-Inspired Collision Avoidance in Swarm Systems via Deep Reinforcement Learning. *IEEE Transactions on Vehicular Technology,* 2022, 71(3), P. 2511–2526. DOI: 10.1109/TVT.2022.3145346
4. Pawełczyk M. Ł., Wojtyra M. Real World Object Detection Dataset for Quadcopter Unmanned Aerial Vehicle Detection. *IEEE Access,* 2020, Vol. 8, P. 174394–174409. DOI: 10.1109/ACCESS.2020.3026192
5. Zhang J., et al. Perdix: A Swarm of Swarming UAVs. *Journal of Field Robotics,* 2019, Vol. 36(6), P. 1240–1255.
6. Smith J., et al. LOCUST: Low-Cost UAV Swarm Technology for Tactical Operations. Defense Technology, 2020, Vol. 16(3), P. 205–215.
7. Sytsma J., Thompson D., Sicoli J. Drone Ultrasonic Detection. Australian International Aerospace Congress, 2023. URL: https://search.informit.org/doi/abs/10.3316/informit.063769306863002 (last accessed 17 May 2024).

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

8. Kritsky D. N., Ovsiannik V. M., Pogudina O. K., Shevel V. V., Druzhinin, E. A. Model for intercepting targets by the unmanned aerial vehicle. *Advances in Intelligent Systems and Computing*. 2019. P. 197–206. DOI: https://doi.org/10.1007/978-3-030-25741-5_20
9. Pohudina O. et al. Assessing unmanned traffic bandwidth. *Integrated Computer Technologies in Mechanical Engineering: Synergetic Engineering*. Cham: Springer International Publishing, 2020. P. 447–458. DOI:10.1007/978-3-030-37618-5_38
10. Petersen K. Tackling air pollution with autonomous drones. MIT School of Engineering, 2021. URL: https://news.mit.edu/2021/tackling-air-pollution-with-autonomous-drones-0624 (дата звернення 17.05.2024)
11. Chu J. New traffic cop algorithm helps a drone swarm stay on task. MIT News Office, 2023. URL: https://news.mit.edu/2023/new-traffic-cop-algorithm-drone-swarm-wireless-0313 (дата звернення 17.05.2024).
12. Lizzio F. F., Capello E., Guglieri G. A Review of Consensus-based Multi-agent UAV Implementations. *Journal of Intelligent & Robotic Systems,* 2022, Vol. 106, (43). 1719 p. DOI: https://doi.org/10.1007/s10846-022-01743-9
13. Padmaja B., Moorthy Ch V K N S N Moorthy, Venkateswarulu N., Bala M.M. Exploration of issues, challenges and latest developments in autonomous cars. *Journal of Big Data*, 2023. Vol. 10(1). P. 1–24. DOI: https://doi.org/10.1186/s40537-023-00701-y
14. Enwerem C., Baras J.S. Consensus-Based Leader-Follower Formation Tracking for Control-Affine Nonlinear Multiagent Systems. *Electrical Engineering and Systems Science.* 2023. DOI: https://doi.org/10.48550/arXiv.2309.09156
15. Xu Z., Yan T., Yang S.X., Gadsden S.A. Distributed Leader Follower Formation Control of Mobile Robots based on Bioinspired Neural Dynamics and Adaptive Sliding Innovation Filter. *IEEE Transactions on Industrial Informatics*, 2023. DOI: https://doi.org/10.1109/TII.2023.3272666
16. Ye Y., Hu S., Zhu X., Sun Z. An Improved Super-Twisting Sliding Mode Composite Control for Quadcopter UAV Formation. *Machines,* 2024, 12(1), 32. DOI: https://doi.org/10.3390/machines12010032
17. Hadi B., Khosravi A., Sarhadi P. Adaptive formation motion planning and control of autonomous underwater vehicles using deep reinforcement learning. *IEEE Journal of Oceanic Engineering.* 2023. P. 1–33. URL: https://arxiv.org/ftp/arxiv/papers/2304/2304.00225.pdf (дата звернення 17.05.2024).
18. Kritskiy D., Yashin S., Koba S. Unmanned aerial vehicle mass model peculiarities. *International scientific-practical conference. Cham: Springer International Publishing*, 2020. P. 299–308. DOI: https://doi.org/10.1007/978-3-030-58124-4_29
19. Distributed Leader Follower Formation Control of Mobile Robots based on Bioinspired Neural Dynamics and Adaptive Sliding Innovation Filter. 2023. URL: https://arxiv.org/pdf/2301.01234.pdf (дата звернення 17.05.2024).

References

1. Jia, G. W., Wang, J. F. (2021), "Research review of UAV swarm mission planning method". *Systems Engineering and Electronics.* 2021. Vol. 43. №. 1. P. 99–111. DOI: 10.3969/j.issn.1001-506X.2021.01.13
2. Do, H. T. (2021), "Formation control algorithms for multiple-uavs: a comprehensive survey" / Do H. T. et all. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, 8(27), Vol. 8(27):170230. DOI:10.4108/eai.10-6-2021.170230
3. Na, S., Niu, H., Lennox, B., Arvin, F. (2022), "Bio-Inspired Collision Avoidance in Swarm Systems via Deep Reinforcement Learning", *IEEE Transactions on Vehicular Technology*, Vol. 71, No. 3, P. 2511–2526. DOI: 10.1109/TVT.2022.3145346
4. Pawełczyk, M. Ł., Wojtyra, M. (2020), "Real World Object Detection Dataset for Quadcopter Unmanned Aerial Vehicle Detection", *IEEE Access*, Vol. 8, P. 174394–174409. DOI: 10.1109/ACCESS.2020.3026192
5. Zhang, J., et al. (2019), "Perdix: A Swarm of Swarming UAVs", *Journal of Field Robotics*, Vol. 36, No. 6, P. 1240–1255.
6. Smith, J., et al. (2020), "LOCUST: Low-Cost UAV Swarm Technology for Tactical Operations," *Defense Technology*, Vol. 16, No. 3, P. 205–215.
7. Sytsma, J., Thompson, D., and Sicoli, J. (2023), "Drone Ultrasonic Detection", Australian International Aerospace Congress. available online: https://search.informit.org/doi/abs/10.3316/informit.063769306863002 (last accessed 17 May 2024).
8. Kritsky, D. N., Ovsiannik, V. M., Pogudina, O. K., Shevel, V. V., and Druzhinin, E. A. (2019), "Model for intercepting targets by the unmanned aerial vehicle", *Advances in Intelligent Systems and Computing.* P. 197–206. DOI: https://doi.org/10.1007/978-3-030-25741-5_20

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) *ISSN 2524-2296 (online)*

9. Pohudina, O., Kritskiy, D., Koba, S., and Pohudin, A. (2020), "Assessing unmanned traffic bandwidth", *Integrated Computer Technologies in Mechanical Engineering: Synergetic Engineering*. P. 447–458. DOI: https://doi.org/10.1007/978-3-030-37618-5_38
10. Petersen, K. (2021), "Tackling air pollution with autonomous drones", MIT School of Engineering. available online: https://news.mit.edu/2021/tackling-air-pollution-with-autonomous-drones-0624 (last accessed 17 May 2024).
11. Chu, J. (2023), "New traffic cop algorithm helps a drone swarm stay on task", MIT News Office. available online: https://news.mit.edu/2023/new-traffic-cop-algorithm-drone-swarm-wireless-0313 (last accessed 17 May 2024).
12. Lizzio, F. F., Capello, E., and Guglieri, G. (2022), "A Review of Consensus-based Multi-agent UAV Implementations", *Journal of Intelligent & Robotic Systems*, Vol. 106, (43). 1719 p. DOI: https://doi.org/10.1007/s10846-022-01743-9
13. Padmaja, B., Moorthy, C.H.V.K.N.S.N., Venkateswarulu, N., and Bala, M.M. (2023), "Exploration of issues, challenges and latest developments in autonomous cars", *Journal of Big Data*, Vol. 10(1). P. 1–24. DOI: https://doi.org/10.1186/s40537-023-00701-y
14. Enwerem, C. and Baras, J.S. (2023), "Consensus-Based Leader-Follower Formation Tracking for Control-Affine Nonlinear Multiagent Systems", Electrical Engineering and Systems Science. DOI: https://doi.org/10.48550/arXiv.2309.09156
15. Xu, Z., Yan, T., Yang, S.X., and Gadsden, S.A. (2023), "Distributed Leader Follower Formation Control of Mobile Robots based on Bioinspired Neural Dynamics and Adaptive Sliding Innovation Filter", *IEEE Transactions on Industrial Informatics*. DOI: https://doi.org/10.1109/TII.2023.3272666
16. Ye, Y., Hu, S., Zhu, X., and Sun, Z. (2024), "An Improved Super-Twisting Sliding Mode Composite Control for Quadcopter UAV Formation", *Machines*, 12(1), 32. DOI: https://doi.org/10.3390/machines1201003
17. Sarhadi, P., et al. (2023), "Adaptive formation motion planning and control of autonomous underwater vehicles using deep reinforcement learning", *IEEE Journal of Oceanic Engineering.* P. 1–33. available online: https://arxiv.org/ftp/arxiv/papers/2304/2304.00225.pdf (last accessed 17 May 2024).
18. Kritskiy, D., Yashin, S., and Koba, S. (2021), "Unmanned aerial vehicle mass model peculiarities". P. 299–308. DOI: https://doi.org/10.1007/978-3-030-58124-4_29
19. "Distributed Leader Follower Formation Control of Mobile Robots based on Bioinspired Neural Dynamics and Adaptive Sliding Innovation Filter" (2023), available online: https://arxiv.org/pdf/2301.01234.pdf (last accessed 17 May 2024).

Надійшла (Received) 09.05.2024

Відомості про авторів / About the Authors

Бінько Ігор Вікторович – Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", аспірант кафедри інформаційних технологій проєктування, Харків, Україна; e-mail: i.v.binko@khai.edu; ORCID ID: https://orcid.org/0009-0007-5638-4292

Шевель Володимир Вікторович – кандидат технічних наук, доцент, Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", доцент кафедри інформаційних технологій проєктування, Харків, Україна; e-mail: v.shevel@khai.edu; ORCID ID: https://orcid.org/0000-0003-0534-0242

Крицький Дмитро Миколайович – кандидат технічних наук, доцент, Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", доцент кафедри інформаційних технологій проєктування, Харків, Україна; e-mail: d.krickiy@khai.edu; ORCID ID: https://orcid.org/0000-0003-4919-0194

Binko Ihor – National Aerospace University "Kharkiv Aviation Institute" named after M. E. Zhukovsky, PhD student at the Department of Information Technology Design, Kharkiv, Ukraine.

Shevel Volodymyr – PhD (Engineering Sciences), Associate Professor, National Aerospace University "Kharkiv Aviation Institute" named after M. E. Zhukovsky, Associate Professor at the Department of Information Technology Design, Kharkiv, Ukraine.

Krytskyi Dmytro – PhD (Engineering Sciences), Associate Professor, National Aerospace University "Kharkiv Aviation Institute" named after M. E. Zhukovsky, Associate Professor at the Department of Information Technology Design, Kharkiv, Ukraine.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

A COMPREHENSIVE APPROACH TO MANAGING ROBOT GROUP FORMATION

Subject matter: Research and development of methods for controlling swarms of unmanned aerial vehicles (UAVs) based on the "master – slave" model. This includes examining existing classifications and interactions between unmanned aerial vehicles in various formations such as groups, flocks, associations, and swarms, with the goal of creating an effective management system. **Goal** To improve the quality of interaction between unmanned aerial vehicles based on the "master – slave" model during flight missions through constant control between objects. Ensuring reliable execution of flight missions by implementing new management methods that account for different modes of interaction between devices. **Tasks**: Analyze the classification of existing UAVs; analyze the parameters and model of interaction of unmanned aerial vehicles in existing groups, flocks, associations, swarms; create a scenario of interaction between two UAVs based on the "master – slave" model; develop a program for visualizing the flight of unmanned aerial vehicles based on the "master – slave" model; conduct flight testing according to the proposed model on stages with various geospatial objects. **Methods**: Simulation method for developing a UAV flight visualization subsystem; graphical modeling method for creating an aircraft-type unmanned aerial vehicle model; methods of algorithm theory for developing a scenario of interaction between two UAVs. Utilization of specialized software tools for visualization and simulation of UAV behavior in real-time conditions. **Results**: Developed a classification of unmanned aerial vehicles; created a graphical model of the Mini-Flight-M aircraft; developed a scheme for the interaction of two UAVs in "teacher" or "mentor" modes; created a program for visualizing the flight of UAVs based on the "master – slave" model; conducted flight testing according to the proposed model on stages with various geospatial objects. The results confirmed the effectiveness of the developed model and demonstrated its applicability in various fields, including environmental monitoring, rescue operations, and other autonomous missions. **Conclusions**: The proposed approach to controlling a UAV swarm based on the "master – slave" model improves the quality of interaction between the devices and ensures reliable execution of flight missions. Further research should focus on optimizing energy consumption and ensuring reliable communication between swarm agents. It is also important to develop methods for protecting UAV swarms from cyberattacks and other threats to enhance their resilience and reliability during complex missions.

Keywords: unmanned aerial vehicles; swarm; master – slave; interaction; simulation; visualization.

Бібліографічні описи / Bibliographic descriptions

Бінько І. В., Шевель В. В. Крицький Д. М. Комплексний підхід до управління формуванням групи роботів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 17–32. DOI: https://doi.org/10.30837/2522-9818.2024.2.017

Binko, I., Shevel, V. Krytskyi D. (2024) "A comprehensive approach to managing robot group formation", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 17–32. DOI: https://doi.org/10.30837/2522-9818.2024.2.017

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 004.896
DOI: https://doi.org/10.30837/2522-9818.2024.2.033

І. Бінько, В. Шевель, А. Биков, Д. Крицький

АНАЛІЗ ДЕЦЕНТРАЛІЗОВАНОЇ МОДЕЛІ УПРАВЛІННЯ ДРОНІВ І РОЗРАХУНОК ТРАЄКТОРІЇ ПЕРЕХОПЛЕННЯ

Предмет дослідження – вивчення застосування інноваційного методу *Cascade DataHub* для оптимізації управління автоматизованими рухомими системами, зокрема безпілотними літальними апаратами. У статті аналізуються теоретичні та практичні аспекти впровадження зазначеного методу в різних галузях. **Мета роботи** – всебічно проаналізувати сучасні моделі та методи керування групою дронів з огляду на децентралізовані підходи, а також розробити ефективні алгоритми для оптимізації траєкторії перехоплення. Розглянути підвищення точності та надійності управління складними автоматизованими системами завдяки збільшеній інтеграції даних у реальному часі. Дослідження спрямоване на виявлення потенційних переваг методу в контексті зменшення часу на реакцію систем і підвищення точності ухвалення рішень. **Завдання:** розробити комплексні алгоритми для швидкого оброблення та аналізу великих обсягів даних із різноманітних джерел; створити надійні комунікаційні протоколи для забезпечення стійкості зв'язку між системами в екстремальних умовах; проаналізувати інтеграцію цих розробок у практичному застосуванні, що дасть змогу збільшити їх ефективність у реальних оперативних умовах. Для досягнення поставленої мети використовуються такі методи: математичне моделювання, статистичний аналіз, машинне та глибоке навчання. Їх застосування дозволяє забезпечити високу точність і надійність роботи управлінських систем. **Результати.** У процесі дослідження встановлено, що метод *Cascade DataHub* забезпечує значне зменшення часу реакції систем на команди, підвищує точність виконання завдань і зменшує втрати даних під час їх передачі. Упровадження цього методу також сприяє ефективнішому розподілу ресурсів між автоматизованими одиницями, що є критично важливим для місій із високими вимогами до координації та часової синхронізації. **Висновки.** Усебічно проаналізовано сучасні моделі та методи керування групою дронів з огляду на децентралізовані підходи. Розроблено ефективні алгоритми оптимізації траєкторії перехоплення, спрямовані на підвищення точності та надійності управління складними автоматизованими системами з допомогою інтеграції даних у реальному часі. Дослідження виявило потенційні переваги запропонованого методу в контексті зменшення часу реакції систем та підвищення точності ухвалення рішень, що сприяє ефективнішому функціонуванню автоматизованих систем.

Ключові слова: *Cascade DataHub*; автоматизовані системи; машинне навчання; управління реальним часом; глибоке навчання.

Вступ

Протягом останніх десятиліть ро́боти стали все більш поширеними в різних галузях, зокрема військовій, цивільній та промисловій. Ефективне управління роботами вимагає швидкого збору, зберігання, оброблення та аналізу великих обсягів інформації в реальному часі. Історично методи керування роботами постійно еволюціонували, пристосовуючись до зростання вимог щодо швидкості оброблення та надійності систем. Останні роки позначені збільшенням інтересу до методів машинного навчання, особливо глибокого, що успішно застосовуються у багатьох галузях [1].

Cascade DataHub, новітня структура нейронної мережі, є одним із таких інноваційних методів, що забезпечує значне підвищення швидкості навчання та ефективності управління завдяки зменшенню кількості необхідної навчальної інформації. У статті подається система управління безпілотними апаратами, що дає змогу керувати групою роботів (роєм), і розглядаються можливості використання методу *Cascade DataHub* для управління автоматизованою рухомою системою.

Основна мета цього дослідження полягає в проведенні всебічного аналізу сучасних моделей і методів управління групою дронів, зважаючи на децентралізовані підходи. Крім того, необхідно розробити ефективні алгоритми для оптимізації траєкторії перехоплення, що підвищують точність і надійність управління складними автоматизованими системами завдяки збільшеній інтеграції даних у реальному часі. Дослідження спрямоване на виявлення потенційних переваг методу в контексті зменшення часу реакції систем і підвищення точності ухвалення рішень.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Аналіз останніх досліджень і публікацій

Управління автоматизованими рухомими системами в складних умовах, як-от керування безпілотними літальними апаратами (БПЛА), вимагає стійких адаптивних стратегій, що можуть ефективно впоратися з високою мінливістю та невизначеністю. Методи управління еволюціонували від простих ручних керувань до складних автономних систем, що використовують передові алгоритми. Спочатку системами переважно керувала безпосередньо людина. З розвитком технологій акцент змістився на автоматизацію для підвищення точності, зменшення часу реакції та мінімізації людських помилок. Це призвело до розроблення різних методологій, кожна з яких адаптована до певних оперативних контекстів:

– в ранніх БПЛА використовувалися *традиційні системи управління*; ці системи покладаються на заздалегідь задані правила й обмежені в керуванні динамічними змінами в середовищі. Вони прямолінійні, але не мають гнучкості;

– *адаптивні системи управління* коригують свої параметри в реальному часі для адаптації до змін у середовищі. Вони більш гнучкі, ніж традиційні системи, і використовуються там, де умови експлуатації не передбачувані;

– *прогностичні системи управління* застосовують моделі для прогнозування майбутніх станів системи, ефективні в сценаріях, де важливе планування та передбачення. Вони можуть виявити потенційні проблеми та проактивно коригувати поведінку системи;

– *інтелектуальні системи управління* застосовують штучний інтелект, зокрема машинне та глибоке навчання, для прийняття рішень на основі аналізу даних. Вони ефективні в складних середовищах, де численні змінні впливають на поведінку системи.

У табл. 1 наведено порівняння основних методів управління.

Таблиця 1. *Порівняльний аналіз методів управління автоматизованими рухомими системами*

Метод	Переваги	Недоліки	Ідеальні умови застосування
Традиційні методи	Надійність, простота налаштування	Обмежена адаптивність	Стабільні, передбачувані умови
Адаптивні системи	Висока адаптивність, самоналаштування	Складність програмування	Змінні умови, де потрібна гнучкість
Методи на основі машинного навчання	Швидкість, адаптивність, навчання з досвіду	Потребують великих обсягів даних для навчання	Динамічні та непередбачувані умови

Нижче подано сучасні студії та публікації, пов’язані з методами управління автоматизованими рухомими системами, особливо БПЛА, з огляду на адаптивні стратегії та складні середовища.

Dual Attention and Focus Loss Using UAV by Y Xu et al. [2] – досліджено повністю автоматизовану систему з використанням БПЛА для збору відеоданих з антен. У роботі інтегруються інноваційні методи управління для ефективного результату. *Developing Mobile Application to Program and Control Robot by D Bhole et al.* [3] – проаналізовано розроблення мобільного застосунку для управління роботизованою рукою з допомогою *Bluetooth*, розглянуто інтеграцію автоматизованих систем у стратегії управління. *Autonomous Inspection and Maintenance Missions with AI Planning and the ROSPlan Framework by J Fillan* [4] – обговорюється методика планування ШІ із застосуванням мобільних маніпуляторів та БПЛА, наголошується на потребі в автоматизованих рішеннях. *Introductory Chapter: Motion Planning for Dynamic Agents by ZA Ali* [5] – описано планування руху як критичного аспекту робототехніки та автоматизації, зокрема для БПЛА й наземних безпілотних транспортних засобів, із упровадженням методів оптимального управління.

Визначення не розв’язаних раніше частин загальної проблеми

Ідентифікація нерозв’язаних питань у царині управління автоматизованими рухомими системами є фундаментальним аспектом для розширення можливостей сучасних технологій. Особливо це стосується БПЛА, де точне та надійне керування є критично важливим. Незважаючи на значний прогрес у технологіях управління, існує низка фундаментальних питань, що залишається нерозв’язаною та потребує подальших досліджень.

1. *Інтеграція даних у реальному часі.* Попри розвиток алгоритмів, наявні системи часто зазнають труднощів, пов’язаних із забезпеченням надійної та ефективної інтеграції інформації з різних джерел

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

у реальному часі. Це особливо актуально в умовах, де динаміка зовнішнього середовища швидко змінюється.

2. *Безпека передачі даних*. Забезпечення безпеки каналів комунікації є основним викликом, особливо в контексті збільшення загроз кібербезпеці.

3. *Адаптивність до нових умов*. Необхідність адаптації систем управління до непередбачених умов без втрати продуктивності або точності.

4. *Розроблення високоефективних алгоритмів*. Створення алгоритмів, що можуть швидко обробляти великі обсяги інформації та забезпечувати точність у прийнятті рішень за динамічних умов.

Основною метою цього дослідження є проведення всебічного аналізу сучасних моделей і методів керування групою дронів, зважаючи на децентралізовані підходи. Крім того, метою статті є розроблення ефективних алгоритмів для оптимізації траєкторії перехоплення, що підвищують точність та надійність управління складними автоматизованими системами завдяки збільшеній інтеграції даних у реальному часі. Дослідження спрямоване на виявлення потенційних переваг методу в контексті зменшення часу реакції систем і підвищення точності ухвалення рішень.

Для досягнення окресленої мети визначені такі завдання:

– розроблення алгоритмів: створення нових алгоритмів, що підвищують швидкість оброблення інформації та точність ухвалення рішень;

– тестування в різних умовах: перевірка ефективності нових методів управління в лабораторних та полігонних умовах;

– аналіз результатів: оцінювання впливу застосування нових алгоритмів на загальну продуктивність і надійність систем;

– інтеграція розробок на практиці: упровадження розроблених алгоритмів і методів у практичному застосуванні, що дасть змогу збільшити їх ефективність в реальних оперативних умовах.

Цей підхід спрямований на заповнення наукових прогалин у царині керування автоматизованими системами та на підтримку розвитку більш ефективних і безпечних технологій управління, що можуть адаптуватися до швидко змінюваних умов експлуатації.

Матеріали й методи

Управління ро́ботами зазвичай здійснюється за допомогою бортового комплексу навігації та управління, що містить інтегровану систему з приймачем супутникової навігації, систему датчиків і сигналів, різні види антен і датчиків, модуль автопілота й систему накопичення та передачі інформації.

Бортова система навігації та управління забезпечує різні можливості, зокрема: рух заданим маршрутом із точністю до координат і висоти поворотних пунктів маршруту, зміна маршрутного завдання або повернення до точки старту за командою з наземного пункту керування, автосупровід обраної цілі, стабілізація кутів орієнтації робота, підтримка заданих висот і швидкості польоту, збирання та передача необхідної інформації та параметрів польоту, а також програмне управління пристроями цільового обладнання [6].

Загальна концепція системи керування дроном (рис. 1) ілюструє взаємодію між наземним програмно-апаратним комплексом і безпосередньо роботом, що обмінюються даними з допомогою захищеного каналу зв’язку. Для оброблення вхідної інформації програміст має зважати на можливі різні формати й типи даних, щоб правильно їх інтерпретувати та використовувати в подальшому обробленні. Наприклад, під час отримання текстових даних важливо перевірити їх правильність і коректність, а також переконатися, що вони містять необхідну інформацію для подальшого оброблення. У процесі отримання числових показників важливо перевірити їх діапазон значень і взяти до уваги можливі помилки округлення або неправильного форматування. Отже, коректне оброблення вхідної інформації є важливим складником ефективної програмної реалізації будь-якого завдання.

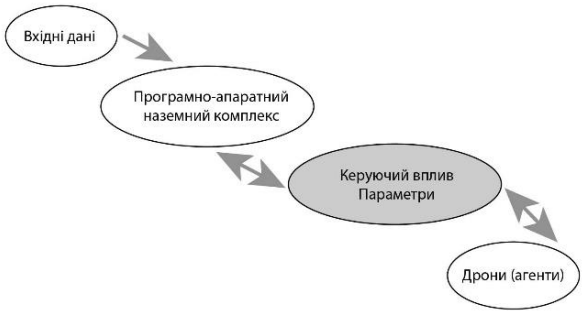


Рис. 1. Загальна концепція системи управління дронами

Система передачі інформації та зв’язку між роботом та програмно-апаратним наземним комплексом може здійснюватися з допомогою різних каналів зв’язку, таких як радіоканали, мережі зв’язку, а також інфрачервоних і лазерних засобів передачі даних.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Крім того, передача інформації може здійснюватися на великі відстані завдяки супутниковим засобам зв'язку. У разі взаємодії групи роботів між собою та з програмно-апаратним наземним комплексом з'являється потреба у використанні протоколів комунікації та алгоритмів координації дій, що дають змогу розв'язувати завдання зі збору інформації та здійснення операцій у групі [7]. На рис. 2 подано основні елементи взаємодії в системі управління роєм роботів із застосуванням генетичного програмування, що ілюструє зазначені аспекти взаємодії.

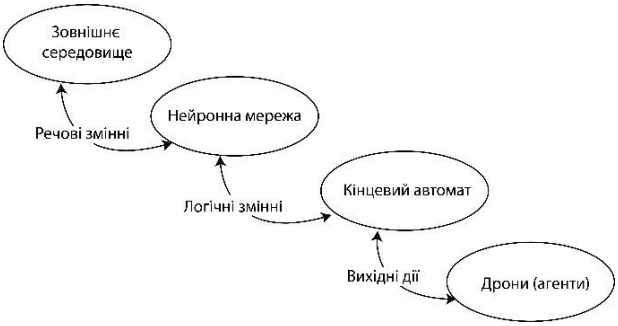


Рис. 2. Основні елементи взаємодії в системі управління роєм роботів із застосуванням генетичного програмування

Крім того, для забезпечення безпеки пересування та керування роботами може впроваджуватися система дистанційного контролю та аварійного відключення, що дозволяє операторам з наземного комплексу віддалено взаємодіяти з роботами та управляти ними в разі виникнення непередбачуваних ситуацій. Зв'язок між роботами та пілотованими об'єктами може здійснюватися завдяки мережам зв'язку та спеціальним пристроям, що дають змогу передавати інформацію між цими об'єктами. Така взаємодія може бути корисною для забезпечення координації та співпраці між пілотованими й безпілотними об'єктами. Крім того, передача інформації та зв'язок між роботами може здійснюватися за допомогою радіо- або інших комунікаційних каналів. Це дозволяє організовувати координацію дій між роєм роботів, що виконують спільне завдання, а також забезпечувати взаємодію між ними та програмно-апаратним наземним комплексом.

Окрім безпосередньої передачі інформації між роботами й програмно-апаратним наземним комплексом, також можлива передача даних за допомогою супутникових зв'язків. Це дає змогу контролювати безпілотний апарат у віддаленому режимі, наприклад, якщо робот перебуває на значній відстані від оператора. У разі використання робота для виконання місій у спеціальних умовах, таких як погана погода або висока інтенсивність радіації, можуть бути використані спеціальні комунікаційні канали, що дозволяють передавати інформацію за відповідних умов. Отже, система управління роботами може бути досить складною та динамічною. Для її ефективної роботи необхідно впроваджувати різноманітні технології передачі інформації та забезпечення зв'язку між компонентами системи.

На рис. 3 показано структуру нейронної мережі та її взаємодію з кінцевим автоматом. Символ *S* позначає нейрони з функцією активації сигмоподібної кишки, а символ *L* позначає нейрони з функцією порогової активації. Поруч вказані номери нейронів, що використовуються в описі крос-операції нейронної мережі. Кожен із трьох нейронів нейронної мережі отримує число, яке може бути як нулем, так і одиницею. Отже, існує вісім можливих комбінацій вихідних сигналів нейронної мережі (000, 001, 010, 011, 100, 101, 110, 111), що надходять на вхід кінцевого автомата. Цю структуру нейронної мережі та її взаємодію з кінцевим автоматом можна використовувати для розв'язання різноманітних завдань у сфері машинного навчання та робототехніки [8].

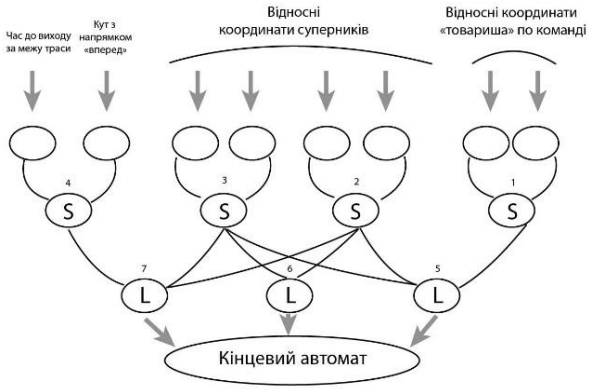


Рис. 3. Нейронна мережа та її взаємодія з кінцевим автоматом

Алгоритм генетичного програмування, що використовується системою управління, передбачає такі етапи: створення початкового покоління, мутація, схрещування (кросовер), відбір особин на формування наступного покоління, обчислення функції пристосованості. Ці процеси зображені на рис. 4, що демонструє загальну структуру нечіткої моделі системи управління автоматизованою безпілотною технікою на основі методу лінеаризації зворотним зв'язком із застосуванням нечіткого логічного висновку.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

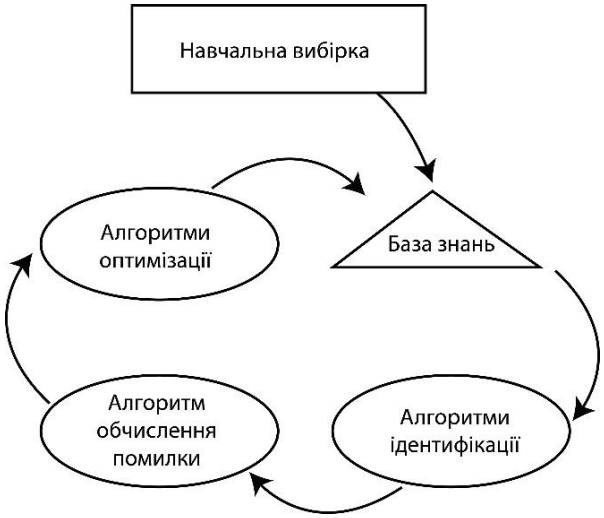


Рис. 4. Загальна структура моделі управління безпілотною технікою із застосуванням нечіткого логічного висновку

Перевагами цієї системи є наочність і відносна простота реалізації, відсутність високих вимог до продуктивності технічного обладнання, а також можливість використання системи для різних вхідних змінних. Недоліками є практична неможливість керування роєм, оскільки взаємодія із "сусідами" оцінюється лише за фактом подій. Незважаючи на значну кількість параметрів, що беруться до уваги, їх спрощення у використовуваній системі не може коректно та об'єктивно оцінити результати, наприклад, ігнорується зміна стану довкілля та параметри робота, що залежать від цього. Як метод аналізу та синтезу впроваджується лінеаризація зворотним зв'язком спільно з нечіткими системами логічного висновку. Порівняно із "звичайною" лінеаризацією, такий алгоритм можна застосовувати для суттєво нелінійних об'єктів управління. Передбачено, що метод лінеаризації зворотним зв'язком спільно з нечіткими системами логічного висновку може бути впроваджений для керування різними об'єктами, що мають суттєву нелінійність. Це може передбачати системи зі змінними параметрами, системи зі збуреннями або системи з нелінійними функціями. Застосування такого методу збільшує точність керування та зменшує вплив похибок вимірювання на якість управління. Порівняно зі звичайною лінеаризацією, метод лінеаризації зворотним зв'язком бере до уваги нелінійні ефекти в системі, тобто зберігає більш точну модель системи [9]. Завдяки цьому методу можна ефективно моделювати поведінку системи, що дає змогу більш точно прогнозувати її реакцію на різні вхідні сигнали. Отже, застосування методу лінеаризації зворотним зв'язком разом із нечіткими системами логічного висновку є ефективним способом управління складними системами з високим ступенем нелінійності.

Зазвичай використання бази знань призводить до обмеження сфери застосування систем управління, до цільової функції об'єкта, яким вона керує. Це може бути як перевагою, так і недоліком зазначеної системи.

Перевагами системи управління роботами на основі лінеаризації зворотним зв'язком та нечіткого логічного висновку є значне зниження впливу невизначеності на якість систем та підвищення якості ідентифікації завдяки оптимізації системою параметрів нечіткої моделі. Крім того, структура нечіткої моделі системи управління роботами, що використовується, дозволяє виправити недоліки нечітких систем, такі як відсутність імовірності доповнення вихідного набору правил бази знань та можливість наявності неповного набору правил, суперечливих чи ідентичних правил через людський фактор у процесі формування правил бази знань.

Недоліками системи управління безпілотними апаратами є трудомісткість обчислень і складність формування бази знань під час навчання нечіткої моделі. Залежність від якості навчання та налаштування нечіткої бази знань може вплинути на ефективність системи управління.

Управління автоматизованою технічною системою на основі розподіленої системи полягає в тому, що одна людина здатна керувати роєм та передавати агентам команди для виконання складних завдань. Кожен апарат оснащений спеціальним комп'ютером, що забезпечує автономну роботу апарата в умовах відсутності зв'язку з керувальною системою.

Управління безпілотним апаратом на основі розподіленої системи є більш ефективним, ніж традиційні методи, оскільки забезпечує більш точне та швидке виконання завдань. Управління може здійснюватися з будь-якої точки світу за допомогою мережі Інтернет.

Програмне забезпечення для керування роботами на основі розподіленої системи може виконувати такі функції: планування маршрутів, автоматичне стеження за об'єктами, зйомка відео й фото, аналіз зібраної інформації та прийняття рішень на її основі.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Управління автоматизованою безпілотною системою на основі розподіленої системи застосовують у таких галузях, як землеробство, лісове господарство, охорона довкілля, моніторинг стану доріг та інфраструктури, військова царина тощо [10].

Схема взаємодії процесів за допомогою *Cascade DataHub* зображена на рис. 5.

Запропонована система має безперечну перевагу керування роєм. Крім того, схема взаємодії процесів, що використовується, дає змогу уникнути багатьох проблем, пов'язаних з управлінням зазначеними процесами та з організацією доступу одних даних до інших.

Однак суттєвими недоліками можна виокремити складність реалізації алгоритмів керування та програмного забезпечення, а також недостатню захищеність системи управління загалом. Також необхідно звернути увагу на обладнання, що використовується системою розподіленого керування, і на ймовірні труднощі в процесі інтеграції програмного забезпечення.

Рис. 5. Схема взаємодії процесів за допомогою *Cascade DataHub* у системі розподіленого управління дронами

В організації управління роєм роботів важливо звернути увагу на такі особливості, як взаємодія роботів у групі, забезпечення отримання та передачі інформації, контроль групи загалом. Взаємодія у рою означає керування діями агентів і контроль за їх виконанням, забезпечення безпеки польоту та уникнення зіткнень.

Отримання та передача інформації передбачає організацію зв'язку між об'єктами системи управління роботами, наземний комплекс та сторонні довірені суб'єкти. Контроль групи – це насамперед визначення розташування групи та її елементів, облік кількості об'єктів системи, встановлення масштабів групи. Необхідно зауважити, що в процесі розроблення та проєктування систем управління роботами важливим етапом є визначення вимог до забезпечення безпеки відповідно до потреб конкретної місії та умов її виконання. Потрібно зважати на потенційні загрози, що можуть виникнути під час місії, і розробляти заходи для їх попередження та/або усунення.

Також важливим етапом є створення алгоритмів керування, що дають змогу забезпечити ефективну роботу системи та досягти поставлених цілей. Ці алгоритми мають бути розроблені з огляду на характеристики роботів, використовувані датчики та інші компоненти системи управління [11, 12].

Окрім того, у розробленні систем управління необхідно брати до уваги технічні недоліки роботів, зокрема обмежену максимальну швидкість та дальність переміщення, обмежену місткість батарей тощо. На всі перелічені фактори необхідно зважати в проєктуванні системи, щоб забезпечити її ефективну роботу та досягти максимальної продуктивності.

Нарешті, важливим аспектом є підтримка та обслуговування системи управління. Потрібно забезпечити доступ до необхідних ресурсів для ефективності системи, а також розробити

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

процедури та інструкції для її експлуатації та технічного обслуговування.

Метод *Cascade DataHub* – це інноваційний підхід до керування безпілотними апаратами, що дає змогу забезпечити надійний, ефективний та безпечний контроль за роботою агентів в умовах реального часу. Основна ідея методу полягає в тому, щоб об'єднати різноманітні джерела інформації, які забезпечують функціонування безпілотних апаратів, в єдину систему управління та моніторингу.

Основні компоненти методу Cascade DataHub

Датчики – забезпечують збір інформації про стан безпілотного апарата, а також про довкілля, в якому він працює. Серед датчиків, які використовуються в методі *Cascade DataHub*, можна виокремити *GPS*-навігатори, акселерометри, гіроскопи, датчики тиску, температури, вологості тощо.

Мережа передачі даних – забезпечує передачу інформації від датчиків до центральної системи керування безпілотними апаратами. Залежно від умов використовуються різноманітні мережі передачі даних, наприклад *Wi-Fi, Bluetooth, 3G/4G, LoRaWAN, NB-IoT* тощо [13].

Центральна система управління – забезпечує оброблення та аналіз інформації, що надходить від датчиків, та прийняття рішень.

Отже, збір та аналіз даних, які надходять від датчиків, та прийняття рішень на їх основі є важливими компонентами керування безпілотними апаратами. Однак для ефективного виконання цих завдань необхідно мати підтримку від відповідних програмних засобів.

Основна ідея методу *Cascade DataHub* полягає в тому, що інформація, яка надходить від датчиків, збирається на одному місці, де вона обробляється та аналізується. Після цього оброблені дані передаються на інші системи управління, що застосовують цю інформацію для контролю та управління автоматизованою безпілотною системою.

Другий етап використання *Cascade DataHub* полягає в навчанні мережі дрібних деталей керування безпілотним апаратом, таких як точне розташування та орієнтація в просторі, уникання перешкод тощо. На цьому етапі мережа вчиться робити більш точні та складні маневри, що дозволяє їй краще керувати роботом у складних умовах.

Третій етап застосування *Cascade DataHub* полягає в навчанні мережі детального управління безпілотною автоматизованою системою. На цьому етапі мережа вчиться виконувати високоточні маневри та рухи, що дає змогу досягати більшої точності та ефективності управління. Наприклад, мережа може вивчити виконання точної зйомки з певної висоти та кута для максимально якісних зображень. Крім того, можна навчити мережу розрізняти певні об'єкти на зображеннях, що допоможе в побудові більш точних карт.

Проте метод *Cascade DataHub* впроваджується не лише для вдосконалення автономної роботи безпілотних апаратів. Він може застосовуватися в багатьох інших сферах, де потрібно розв'язувати складні завдання із значною кількістю інформації.

Наприклад, зазначений метод може бути використаний у медицині для вдосконалення процесу діагностики захворювань на основі аналізу зображень або в економічній галузі для аналізу фінансових ризиків та побудови прогнозів.

Отже, метод *Cascade DataHub* є потужним інструментом для розв'язання складних завдань, що вимагають значної кількості інформації та високої точності результатів. Він може бути застосований у багатьох сферах, зокрема в авіаційній промисловості та різноманітних наукових дослідженнях. Із зростанням кількості даних і вдосконаленням технологій навчання машин метод *Cascade DataHub* стає ще більш ефективним і потужним інструментом для досягнення високих результатів у різних сферах діяльності.

Однією з основних переваг методу *Cascade DataHub* є висока швидкість оброблення інформації. Завдяки цьому дані можуть бути аналізовані в реальному часі, що дає змогу операторам здійснювати швидке та ефективне керування БПЛА.

Іншою перевагою методу *Cascade DataHub* є його достатня точність і надійність. Цей метод точно визначає параметри польоту, а саме: швидкість, висоту, напрямок та інші, що дозволяє операторам ефективно керувати автоматизованою безпілотною системою.

Cascade DataHub – це метод глибокого навчання, що дає змогу зменшити кількість навчальної інформації та збільшити швидкість навчання. Це досягається завдяки послідовному під'єднанню декількох нейронних мереж, кожна з яких відповідає за визначення різних аспектів управління системою.

Принципи роботи методу *Cascade DataHub* передбачає кілька етапів.

Перший етап полягає в тому, що використовується нейронна мережа, яка навчається на загальних даних і визначає загальне керування системою.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Другий етап полягає в під'єднанні до першої мережі наступної, яка визначає більш детальні аспекти управління системою.

Метод уже застосовувався для керування групою БПЛА в сільському господарстві, де важливо швидко реагувати на мінливі погодні умови й забезпечувати точність збору показників. *Cascade DataHub* вирізняється здатністю зменшувати кількість потрібної навчальної інформації та підвищувати швидкість її оброблення, що забезпечує високу ефективність управління.

Технічні характеристики

DataHub Manager – програма, що керує безпілотними апаратами та забезпечує збір інформації з них.

DataHub Analytics – програма, яка аналізує та візуалізує дані з БПЛА.

DataHub Connectors – компоненти, що дають змогу з'єднувати *DataHub* із різними джерелами інформації, такими як сенсори, бази даних та інші системи.

DataHub API – програмний інтерфейс, що дозволяє інтегрувати *DataHub* з іншими системами та розробляти власні застосунки на базі *DataHub*.

За допомогою *DataHub Manager* можна керувати безпілотними апаратами з одного місця, використовуючи віддалений доступ через інтернет. Крім того, програма допомагає встановлювати параметри польоту, контролювати батареї, керувати камерою та виконувати інші дії.

DataHub Analytics дає змогу аналізувати інформацію, зібрану з безпілотних апаратів, за допомогою різних інструментів аналізу даних, таких як графіки, таблиці, карти тощо. Крім того, програма дозволяє створювати власні звіти та дашборди для візуалізації інформації.

DataHub Connectors є додатковими програмними інструментами, що дають змогу отримувати інформацію з різних джерел і джерел трансляції в режимі реального часу. У цьому разі йдеться про інформацію з датчиків, мереж, баз даних, трансляцій відео та звуку. *Connectors* можуть бути налаштовані для збору інформації з різних джерел та її транслювання до аналітичних інструментів, розроблених із використанням *Cascade DataHub* [14].

Загалом *DataHub Connectors* дають змогу під'єднуватися до кількох джерел трансляції з різними протоколами зв'язку: *MQTT, OPC UA, REST API* тощо. Крім того, *Connectors* можуть бути розширені для під'єднання до нових джерел інформації за допомогою розширення програмного забезпечення.

DataHub API – це інтерфейс програмування застосувань (*API*), який дозволяє розробникам звертатися до інформації, розташованої в системі *DataHub*. За допомогою *API* можна взаємодіяти з базою даних, відправляти запити на отримання інформації та керувати нею. *DataHub API* використовує стандартні протоколи *REST API*, що дає змогу розробникам легко інтегрувати свої програмні продукти з *DataHub*.

API DataHub розроблено з огляду на масштабованість і гнучкість, що дозволяє розробникам створювати застосунки різного рівня складності та розміру. Крім того, *DataHub API* підтримує мови програмування *Python, JavaScript, Java, Ruby, PHP* та *C #*, що уможливлює їх вибір.

Метод *Cascade DataHub* допомагає створити єдину платформу для збору, оброблення та аналізу інформації з безпілотних літальних апаратів. Цей метод інтегрує різноманітні джерела даних, такі як *GPS*, камери, сенсори, інші системи управління та моніторингу, у єдину систему, що дає змогу здійснювати моніторинг, аналіз та керування апаратами в реальному часі.

Ключовою особливістю методу є можливість створення різних зв'язків між даними, що дозволяє аналізувати та контролювати роботу апаратів з різних кутів зору. Завдяки технології оброблення інформації в реальному часі оператори можуть миттєво реагувати на небезпечні ситуації, підвищуючи безпеку та надійність управління.

Крім того, метод *Cascade DataHub* зменшує витрати на керування безпілотними апаратами, оскільки дає змогу здійснювати контроль і моніторинг апаратів на віддаленій відстані зі зручного місця, зменшуючи необхідність у великій кількості спеціалізованого обладнання та персоналу.

Незважаючи на переваги методу *Cascade DataHub*, він також має недоліки. Зокрема йдеться про потребу у високошвидкісному інтернет-з'єднанні для передачі значного обсягу інформації в реальному часі. Також можуть виникнути проблеми зі стабільністю роботи системи в разі непередбачуваних ситуацій, таких як відключення одного із джерел даних.

Для забезпечення ефективного управління групою дронів у мінливих умовах важливо розробити систему взаємодії між БПЛА, яка дає змогу їм виконувати різні завдання в координації один з одним (рис. 6).

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Можливий підхід, коли використовується основний дрон (лідер), що розподіляє завдання між іншими апаратами. Взаємодія дронів у такий спосіб дає змогу ефективно реагувати на зміни в середовищі та перерозподіляти завдання у разі виходу з ладу одного з літальних апаратів.

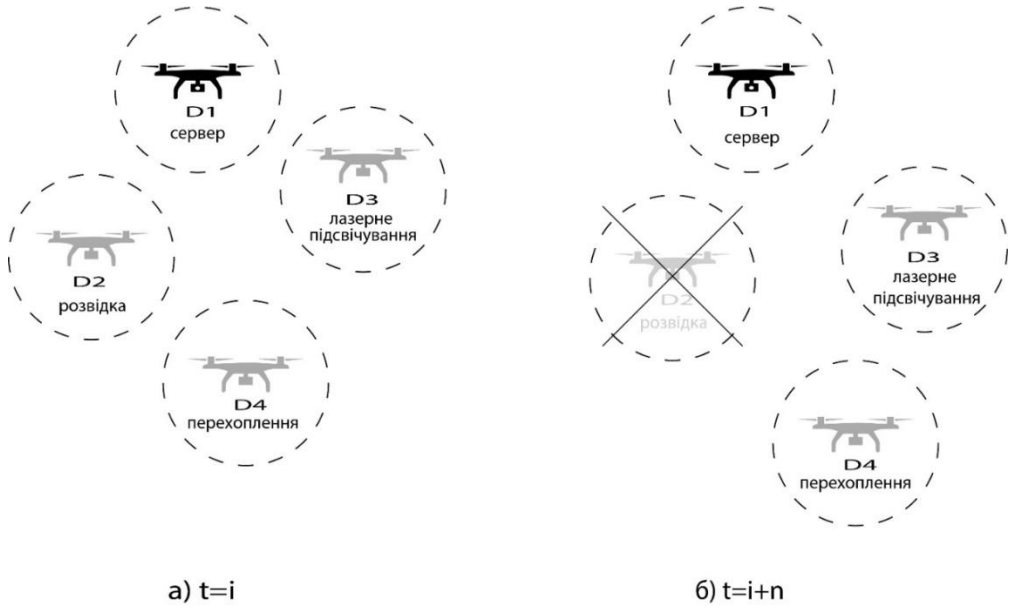


Рис. 6. Схема реалізації взаємодії дронів у певний момент часу

На рис. 6, *а* показано, як дрони виконують завдання в певний момент часу $t = i$. Основний дрон (*D1*) керує іншими апаратами: *D2* виконує розвідувальні завдання, *D3* – лазерне підсвічування, *D4* – перехоплення.

На рис. 6, *б* показано ситуацію в момент часу $t = i + n$, коли один із дронів (*D2*) був збитий або вийшов з ладу. Основний дрон (*D1*) перерозподіляє завдання між іншими апаратами, і завдання збитого дрона (*D2*) виконує допоміжний БПЛА (*D3*).

Запропонована схема взаємодії дронів забезпечує гнучкість і стійкість системи управління в умовах мінливого середовища. Перерозподіл завдань у разі виходу з ладу одного з дронів дозволяє зберегти ефективність операцій та швидко адаптуватися до нових умов.

Такий підхід значно підвищує надійність системи та мінімізує втрати функціональних можливостей у критичних ситуаціях. Основний дрон (лідер) виконує ключову роль у забезпеченні координації та управління групою дронів, що дає змогу максимально ефективно використовувати доступні ресурси.

Результати моделювання показали, що застосування цієї схеми взаємодії значно підвищує ефективність виконання завдань, зокрема в умовах складної оперативної обстановки. Перспективи подальших досліджень передбачають оптимізацію алгоритмів розподілу завдань і вивчення можливостей інтеграції додаткових функцій, таких як автоматичне виявлення загроз і реагування на них.

Упровадження зазначеного підходу в реальні системи керування безпілотними літальними апаратами може значно впливати на розвиток технологій автономного управління та підвищення ефективності виконання різних операцій.

У процесі визначення траєкторії перехоплювача використовувався метод наведення за принципом "крива погоні". Цей метод був обраний, оскільки він відносно простий у реалізації, а також його можна застосовувати як у системах наведення самонавідних ракет, так і в БПЛА.

Метод наведення за кривою погоні схожий на гонитву собаки за зайцем, тому в літературі його також називають методом "крива переслідування", або "собача крива". Існує два основні випадки застосування цього методу: переслідування цілі, що віддаляється, і переслідування цілі, що наближається (рис. 7).

У першому випадку, якщо дрон має достатню дальність польоту та швидкість більшу, ніж швидкість цілі, він може вразити ціль. У другому випадку під час наближення до цілі різко зростає швидкість повороту. Це навантаження може бути надмірним для корпусу ракети, що, імовірно,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

спричинить його руйнування. У реальності керуюча сила, створювана кермом напрямку, збільшується лише до певного значення. Отже, може настати момент, коли кермо літального апарата відхилиться до упору, але максимальна керуюча сила, що виникає в цьому разі, виявиться недостатньою для необхідної зміни напрямку руху.

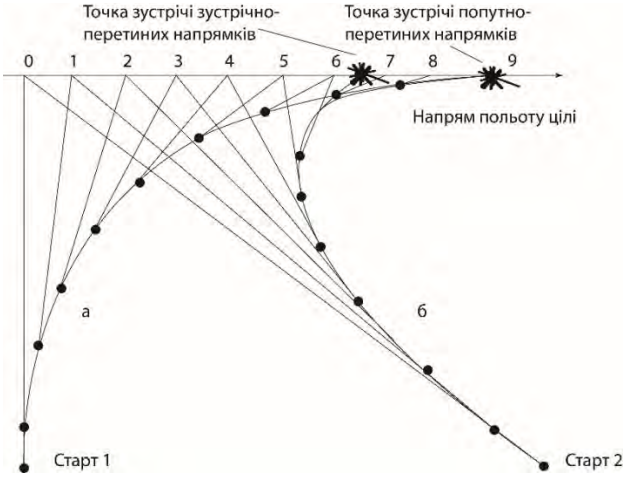


Рис. 7. Криві погоні за попутно-перетинних та зустрічно-перетинних напрямок польоту

Із цього моменту дрон рухатиметься по колу мінімального радіуса, що відповідає граничній керуючій силі. Наведення припиниться, оскільки літальний апарат не встигатиме розгортатися за ціллю. Ціль за деякий час вийде з поля зору координатора, після чого наведення стає неможливим [15].

Щоб вивести рівняння лінії, оберемо систему координат, у якій вісь абсцис проходить крізь початкове положення точок P і A, у цьому разі точка A розташована на початку системи координат xAy. Співвідношення постійних швидкостей точок позначимо літерою k.

Якщо припустити, що за нескінченно малий проміжок часу точка P пройшла відстань dS, а точка A – відстань dS_i, тоді за поставленою вище умовою матимемо співвідношення $dS = kdS_i$, або

$$\sqrt{dx^2 + dy^2} = k\sqrt{d\xi^2 + d\eta^2}. \quad (1)$$

Далі необхідно виразити $d\xi$ і $d\eta$ через x, y та їх диференціали. За умовою координати точки P мають задовольняти рівняння дотичної до шуканої кривої, тобто

$$\eta - y = \frac{dy}{dx}(\xi - x). \quad (2)$$

Додаючи до цього рівняння, задане умовою рівняння траєкторії $F(\xi,\eta)$ руху "втікача", можна визначити з отриманої системи рівняння ξ і η. Після підставлення цих значень у диференційне рівняння воно запишеться у вигляді

$$\Phi\left(x, y, \frac{dy}{dx}, \frac{d^2y}{dx^2}\right) = 0. \quad (3)$$

Постійні інтегрування можуть бути знайдені з початкових умов $(y = 0; y' = 0)$ (коли $x = 0$).

Загалом для довільно заданої кривої $F(\xi,\eta)$ знайти рішення отриманого рівняння досить складно. Задача істотно спрощується, якщо розглянути простий випадок, коли траєкторія "втікача" є прямою.

Розглянемо випадок $A0(0,0)$, $P0(0,1)$ за умови руху "втікача" вздовж осі x і якщо $k > 0$. У довільний момент часу "втікач" завжди перебуває на дотичній до кривої траєкторії руху "переслідувача". Отже,

$$\frac{dy}{dx} = \frac{-y}{a - x}. \quad (4)$$

На підставі цього запишемо диференційне рівняння

$$y + y'(a - x) = 0, \quad (5)$$

де $y > 0$.

З умови $a = V \cdot t$ випливає $\frac{y}{y'} + Vt = x$, після диференціювання за часом $\dot{y} = y' \cdot \dot{x}$ і $\dot{y}' = y'' \cdot \dot{x}$, на підставі яких знаходимо

$$\dot{x} = \frac{dx}{dy} = \frac{V \cdot y'^2}{y \cdot y''}. \quad (6)$$

Запишемо вираз для визначення довжини кривої

$$l = Wt = k\int_0^x \sqrt{1 + (y')^2}\, dx. \quad (7)$$

З виразів $dx^2 + dy^2 = W^2 dt^2$ і $\omega^2 = \frac{dx^2}{dt^2} + \frac{dy^2}{dt^2} = \dot{x}^2 + (y' \cdot \dot{x})^2$ випливає $\dot{x} = \frac{W}{\sqrt{1 + y'^2}}$.

Аналогічно проводимо диференціювання за y:

$$y'' - k \cdot \frac{y'^2}{y} \cdot \sqrt{1 + y'^2} = 0. \quad (8)$$

Рішення з підставленням $u = x' = \frac{1}{y'}$, $y'' = \frac{-1}{u^3}\frac{du}{dx}$ за умови поділу змінних приводить до $\frac{-du}{\sqrt{1 + u^2}} = k \cdot \frac{dy}{y}$, після інтегрування отримуємо $\arcsin u = k \cdot \ln y + C$.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Далі після використання формального визначення $\sin h$ з $C_1 = e^C$ маємо $x' = \frac{dx}{dy} = \frac{1}{2}\left[(C_1 \cdot y)^k - (C_1 \cdot y)^{-k}\right]$. Ще раз інтегруємо з визначенням постійної інтегрування C_2. З початкових умов $\left.\frac{dx}{dy}\right|_{y=1} = 0$ випливає $C_1 = 1$, а також $x\big|_{y=1} = 0$. Отримуємо $C_2 = \frac{k}{1-k^2}$ або

$$x(y) = \frac{1}{2}\left(\frac{y(1+k)}{(1+k)} - \left\{\frac{y^{(1+k)}}{(1-k)} \atop \ln|y|\right\}\right) + \left\{\frac{\frac{k}{1-k^2}}{-\frac{1}{4}} \right\} \left\{k \neq 1 \atop k = 1\right\}.$$

Результати досліджень та їх обговорення

У статті проаналізовано ефективність застосування методу *Cascade DataHub* у контексті керування безпілотними літальними апаратами. Практична реалізація цього методу демонструє, як інтеграція різноманітних джерел інформації в реальному часі може значно покращити якість і швидкість управлінських рішень у динамічних умовах. Встановлено, що *Cascade DataHub* забезпечує більшу стабільність зв'язку між апаратами, що важливо для місій із високими вимогами до координації та часової синхронізації.

На мапі (рис. 8) зображені отримані траєкторії польоту. У нижній частині форми розташовуються елементи керування швидкістю відтворення анімаційної моделі та старту / паузи демонстрації. Під час моделювання користувачеві надається спрощений візуальний процес переслідування, нейтралізації та повернення на базу перехоплювача. На формі червоним кольором позначено перехоплювач, а зеленим – ціль. У процесі демонстрації користувач може змінити масштаб часу за допомогою керувального елемента в нижній лівій ділянці форми, за замовчуванням візуалізація відбувається в реальному часі. Після нейтралізації ціль забарвлюється в чорний колір, візуалізація призупиняється та виводиться повідомлення про успішність перехоплення. Після цього перехоплювач починає повертатися назад на базу запуску.

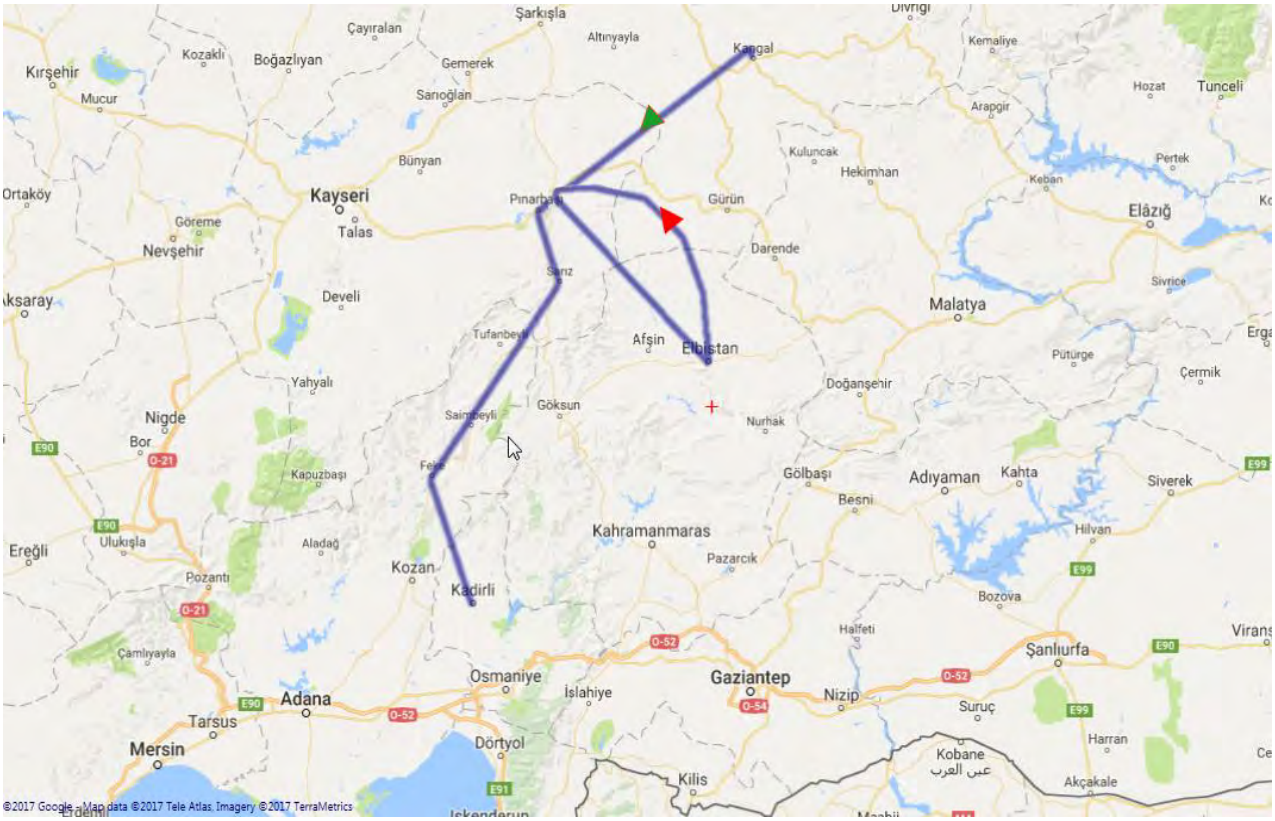
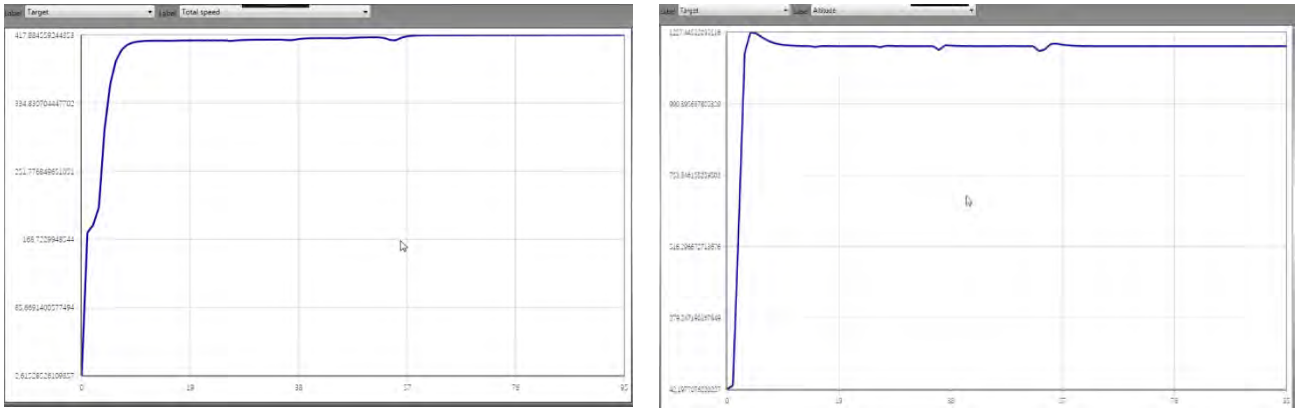


Рис. 8. Демонстрація перехоплення цілі

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Переходячи до форми графічних звітів (рис. 9), користувач може переглянути показники метрики, отримані в процесі моделювання способом вибору відповідного покажчика та його належності (ціль або перехоплювач).



а) б)

Рис. 9. Графіки: а) висота польоту дрона; б) швидкість польоту дрона

Основні результати передбачають зниження часу реакції системи на команди, підвищення точності виконання завдань і зменшення втрат інформації під час її передачі. Також було зазначено, що застосування цього методу сприяє ефективнішому розподілу ресурсів між апаратами, що дозволяє використовувати їх у більш складних місіях без збільшення витрат на обслуговування та управління.

Обговорення цих результатів вказує на значні переваги методу *Cascade DataHub* перед традиційними підходами, особливо в контексті масштабованих операцій із застосуванням безпілотних літальних апаратів. Подальші дослідження мають зосередитися на розвитку алгоритмів адаптації до змінних умов експлуатації та інтеграції з іншими системами управління для створення більш гнучких і робастних систем.

Висновки

Унаслідок проведених досліджень підтверджено ефективність методу *Cascade DataHub* для керування безпілотними літальними апаратами в децентралізованій моделі. Основні результати показали, що метод забезпечує значне скорочення часу реакції системи на команди, підвищення точності виконання завдань і зменшення втрат інформації під час її передачі. Упровадження методу *Cascade DataHub* дає змогу ефективно розподіляти ресурси між автоматизованими одиницями, що є критично важливим для місій із високими вимогами до координації та часової синхронізації.

Усебічно проаналізовано сучасні моделі та методи управління групою дронів з огляду на децентралізовані підходи. Розроблені алгоритми оптимізації траєкторії перехоплення спрямовані на підвищення точності та надійності керування складними автоматизованими системами завдяки інтеграції даних у реальному часі. Дослідження виявило потенційні переваги запропонованого методу в контексті зменшення часу реакції систем і підвищення точності ухвалення рішень, що сприяє ефективнішому функціонуванню автоматизованих систем.

Подальші дослідження мають зосередитися на оптимізації алгоритмів адаптації до змінних умов експлуатації та інтеграції з іншими системами управління для створення більш гнучких і робастних систем. Упровадження окресленого підходу в реальні системи керування безпілотними літальними апаратами може значно впливати на розвиток технологій автономного управління та підвищення ефективності виконання різних операцій.

Список літератури

1. Taye M. M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow. *Applications and Future Directions. Computers. 2023.* Vol. 12. P. 91. DOI: https://doi.org/10.3390/computers12050091

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

2. Xu Y., Ke Q., Jiang Z., Zhai Y., Genovese A., Piuri V. Dual attention and focus loss using UAV. Piuri Labs. 2023. URL: https://piurilabs.di.unimi.it/Papers/tai23.pdf
3. Bhole D., Domle M., Motghare M., Vaidya A. P. P. M. Developing mobile application to program and control robot. IRJMETS. 2023. URL: https://www.irjmets.com/uploadedfiles/paper/issue_12_december_2023/47953/final/fin_irjmets1704292742.pdf
4. Fillan J. Autonomous inspection and maintenance missions with AI planning and the ROSPlan framework. NTNU Open. 2023. URL: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3094654
5. Ali Z. A. Introductory chapter: Motion planning for dynamic agents. InTechOpen. 2024. URL: https://www.intechopen.com/chapters/1178552
6. Rati B., Rajendra P., Parvez F., Jyoti G. Blockchain-enabled secure and efficient data sharing scheme for trust management in healthcare smartphone network. *The Journal of Supercomputing. 2023.* Vol. 79. P. 16233–16274. DOI: https://doi.org/10.1007/s11227-023-05272-6
7. Куценко Л. М., Семків О. М., Калиновський А. Я., Піксасов М. М., Сухарькова О. І. Геометричне моделювання мобільної установки для запуску безпілотних літальних апаратів. *Технічні науки.* 2017. №12(41). С. 117–120. DOI: 10.15587/2313-8416.2017.117920
8. Majumdar A. K. Fundamentals of Free-Space Optical Communications Systems, Optical Channels, Characterization, and Network/Access Technology. *Optical Wireless Communications for Broadband Global Internet Connectivity.* 2019. P. 87–118. Elsevier. DOI: https://doi.org/10.1016/B978-0-12-813365-1.00004
9. Kabir H., Tham M.-L., Chang Y. C. Internet of robotic things for mobile robots: Concepts, technologies, challenges, applications, and future directions. *Digital Communications and Networks. 2023.* Advance online publication. P. 1–39. DOI:10.1016/j.dcan.2023.05.006
10. Gielis J., Shankar A., Prorok A. A Critical Review of Communications in Multi-robot Systems. *Current Robot Reports.* 2022. Vol. 3(3). P. 213–225. DOI: 10.1007/s43154-022-00090-9
11. Gielis J., Shankar A., Prorok A. A Critical Review of Communications in Multi-robot Systems // Current Robot Reports. 2022. Vol. 3(3). С. 213–225. DOI: 10.1007/s43154-022-00090-9
12. Pathak R., Barzin R., Bora G. C. Data-driven precision agricultural applications using field sensors and Unmanned Aerial Vehicle. *International Journal of Precision Agriculture and Aviation.* 2018. Vol. 1(1). P. 19–23. DOI: https://doi.org/10.33440/j.ijpaa.20180101.0004
13. Tahir A. Formation Control of Swarms of Unmanned Aerial Vehicles. Doctoral Dissertation. University of Turku, Turku, Finland. 2023. URL: https://urn.fi/URN:ISBN:978-951-29-9411-3. ISBN: 978-951-29-9411-3.
14. Kong X., Yuhan W., Wang H. Edge Computing for Internet of Everything: A Survey. *IEEE Internet of Things Journal.* 2022, December. Advance online publication. P. 23472–23485. DOI: https://doi.org/10.1109/JIOT.2022.3200431
15. Liao S.-l., Zhu R.-m., Wu N.-q., Shaikh T. A., Sharaf M., Mostafa A. M. Path planning for moving target tracking by fixed-wing UAV. Defence Technology. 2020. Vol. 16(4). P. 811–824. DOI: 10.1016/j.dt.2019.10.010. URL: https://www.sciencedirect.com/science/article/pii/S2214914719304817

References

1. Taye, M.M. (2023), "Understanding of Machine Learning with Deep Learning: Architectures, Workflow". *Applications and Future Directions. Computers 2023.*, Vol. 12, 91 p. DOI: https://doi.org/10.3390/computers12050091
2. Xu, Y., Ke, Q., Jiang, Z., Zhai, Y., Genovese, A. and Piuri, V. (2023), 'Dual attention and focus loss using UAV', *Piuri Labs*. available at: https://piurilabs.di.unimi.it/Papers/tai23.pdf
3. Bhole, D., Domle, M., Motghare, M. and Vaidya, A.P.P.M. (2023), "Developing mobile application to program and control robot", IRJMETS. available at: https://www.irjmets.com/uploadedfiles/paper/issue_12_december_2023/47953/final/fin_irjmets1704292742.pdf
4. Fillan, J. (2023), "Autonomous inspection and maintenance missions with AI planning and the ROSPlan framework". NTNU Open. available at: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3094654
5. Ali, Z.A. (2024), "Introductory chapter: Motion planning for dynamic agents", InTechOpen. available at: https://www.intechopen.com/chapters/1178552
6. Rati, B., Rajendra, P., Parvez, F. and Jyoti, G. (2023), "Blockchain-enabled secure and efficient data sharing scheme for trust management in healthcare smartphone network", *The Journal of Supercomputing*, Vol. 79, P. 16233–16274. DOI: https://doi.org/10.1007/s11227-023-05272-6

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

7. Kutsenko, L.M., Semkiv, O.M., Kalinovskyi, A.Y. and Piksasov, M.M. (2017), "Geometric Modeling of a Mobile Installation for Launching Unmanned Aerial Vehicles". *Technical Sciences*, №12(41). P. 117–120. DOI: 10.15587/2313-8416.2017.117920
8. Majumdar, A.K. (2019), "Fundamentals of Free-Space Optical Communications Systems, Optical Channels, Characterization, and Network/Access Technology", *Optical Wireless Communications for Broadband Global Internet Connectivity*, P. 87–118. Elsevier. DOI: https://doi.org/10.1016/B978-0-12-813365-1.00004
9. Kabir, H., Tham, M.-L. and Chang, Y.C. (2023), "Internet of robotic things for mobile robots: Concepts, technologies, challenges, applications, and future directions", *Digital Communications and Networks*. Advance online publication. P. 1–39. DOI:10.1016/j.dcan.2023.05.006
10. Gielis, J., Shankar, A. and Prorok, A. (2022), "A Critical Review of Communications in Multi-robot Systems", *Current Robot Reports*, Vol. 3(3), P. 213–225. DOI: 10.1007/s43154-022-00090-9
11. Gielis J., Shankar A., Prorok A. A Critical Review of Communications in Multi-robot Systems // Current Robot Reports. 2022. Vol. 3(3). C. 213–225. DOI: 10.1007/s43154-022-00090-9
12. Pathak, R., Barzin, R. and Bora, G.C. (2018), "Data-driven precision agricultural applications using field sensors and Unmanned Aerial Vehicle", *International Journal of Precision Agriculture and Aviation*, Vol. 1(1), P. 19–23. DOI: https://doi.org/10.33440/j.ijpaa.20180101.0004
13. Tahir, A. (2023), "Formation Control of Swarms of Unmanned Aerial Vehicles", Doctoral Dissertation, University of Turku, Turku, Finland. available at: https://urn.fi/URN:ISBN:978-951-29-9411-3. ISBN: 978-951-29-9411-3.
14. Kong, X., Yuhan, W. and Wang, H. (2022, December), "Edge Computing for Internet of Everything: A Survey", *IEEE Internet of Things Journal*. Advance online publication. P. 23472–23485. DOI: https://doi.org/10.1109/JIOT.2022.3200431
15. Liao, S.-l., Zhu, R.-m., Wu, N.-q., Shaikh, T. A., Sharaf, M. and Mostafa, A. M. (2020), "Path planning for moving target tracking by fixed-wing UAV", Defence Technology, Vol. 16(4), P. 811–824. DOI: 10.1016/j.dt.2019.10.010. available at: https://www.sciencedirect.com/science/article/pii/S2214914719304817

Надійшла (Received) 09.05.2024

Відомості про авторів / About the Authors

Бінько Ігор Вікторович – Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", аспірант кафедри інформаційних технологій проєктування, Харків, Україна; e-mail: i.v.binko@khai.edu; ORCID ID: https://orcid.org/0009-0007-5638-4292

Шевель Володимир Вікторович – кандидат технічних наук, доцент, Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", доцент кафедри інформаційних технологій проєктування, Харків, Україна; e-mail: v.shevel@khai.edu; ORCID ID: https://orcid.org/0000-0003-0534-0242

Биков Андрій Миколайович – Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", аспірант кафедри інформаційних технологій проєктування, Харків, Україна; e-mail: a.bykov@khai.edu; ORCID ID: https://orcid.org/0000-0002-7184-4994

Крицький Дмитро Миколайович – кандидат технічних наук, доцент, Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", доцент кафедри інформаційних технологій проєктування, Харків, Україна; e-mail: d.krickiy@khai.edu; ORCID ID: https://orcid.org/0000-0003-4919-0194

Binko Ihor – National Aerospace University "Kharkiv Aviation Institute" named after M. E. Zhukovsky, PhD student at the Department of Information Technology Design, Kharkiv, Ukraine.

Shevel Volodymyr – PhD (Engineering Sciences), Associate Professor, National Aerospace University "Kharkiv Aviation Institute" named after M. E. Zhukovsky, Associate Professor at the Department of Information Technology Design, Kharkiv, Ukraine.

Bykov Andrii – National Aerospace University "Kharkiv Aviation Institute" named after M. E. Zhukovsky, PhD student at the Department of Information Technology Design, Kharkiv, Ukraine.

Krytskyi Dmytro – PhD (Engineering Sciences), Associate Professor, National Aerospace University "Kharkiv Aviation Institute" named after M. E. Zhukovsky, Associate Professor at the Department of Information Technology Design, Kharkiv, Ukraine.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

ANALYSIS OF DECENTRALIZED DRONE CONTROL MODEL AND INTERCEPTION TRAJECTORY CALCULATION

Subject matter: This article is devoted to the study of applying the innovative Cascade DataHub method for optimizing the management of automated mobile systems, especially unmanned aerial vehicles. The work analyzes both theoretical and practical aspects of implementing this method across various application sectors. **Goal:** The objective of the study is to conduct a comprehensive analysis of contemporary models and methods for managing a group of drones, focusing on decentralized approaches. Additionally, the study aims to develop effective algorithms for optimizing the interception trajectory, with the goal of enhancing the accuracy and reliability of managing complex automated systems through increased real-time data integration. The research is directed towards identifying the potential advantages of this method in reducing system response times and improving decision-making accuracy. **Tasks:** The main tasks of the research include the development of comprehensive algorithms for rapid processing and analysis of large volumes of data from various sources, creating reliable communication protocols to ensure connection stability under extreme conditions. Another important task is the integration of these developments into practical applications, which will increase their effectiveness in real operational conditions. **Methods**: To achieve the set goals, advanced techniques of mathematical modeling, statistical analysis, machine learning, and deep learning are used. The application of these techniques ensures high accuracy and reliability of the management systems. **Results:** During the research, it was found that the Cascade DataHub method significantly reduces the response time of systems to commands, increases the accuracy of task execution, and decreases data loss during their transmission. The implementation of this method also contributes to a more efficient distribution of resources among automated units, which is critically important for missions requiring high coordination and time synchronization**. Conclusions:** A comprehensive analysis of contemporary models and methods for managing a group of drones with a focus on decentralized approaches has been conducted. Effective algorithms for optimizing the interception trajectory have been developed, aimed at enhancing the accuracy and reliability of managing complex automated systems through real-time data integration. The study revealed the potential advantages of the proposed method in reducing system response times and improving decision-making accuracy, contributing to the more efficient functioning of automated systems.

Keywords: Cascade DataHub; automated systems; machine learning; real-time management; deep learning.

Бібліографічні описи / Bibliographic descriptions

Бінько І. В., Шевель В. В., Биков А. М., Крицький Д. М. Аналіз децентралізованої моделі управління дронів і розрахунок траєкторії перехоплення. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 33–47. DOI: https://doi.org/10.30837/2522-9818.2024.2.033

Binko, I., Shevel, V., Bykov, A., Krytskyi, D. (2024), "Analysis of decentralized drone control model and interception trajectory calculation", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 33–47. DOI: https://doi.org/10.30837/2522-9818.2024.2.033

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

УДК 004.8 DOI: https://doi.org/10.30837/2522-9818.2024.2.048

В. Волоховський

АНАЛІЗ МЕТОДІВ ТРЕНУВАННЯ ВУЗЬКОСПРЯМОВАНИХ МОВНИХ МОДЕЛЕЙ У СФЕРІ ГЕНЕРАЦІЇ ДОГОВОРІВ

Предметом дослідження є моделі та методи машинного навчання для генерації договорів в умовах обмежених ресурсів і способи порівняння та оцінювання їх ефективності. **Мета роботи** – аналіз підходів до розроблення вузькоспрямованих великих мовних моделей та визначення оптимального методу створення незалежних спеціалізованих систем, що дають змогу генерувати договори різними мовами в різних правових системах. У статті розв'язуються такі **завдання**: визначення наявних компаній та рішень, виявлення підходів до створення текстів природною мовою, аналіз способів оцінювання та порівняння таких систем, виявлення обмежень і недоліків сучасних рішень і підходів, пошук оптимального методу розроблення систем за умови обмежених ресурсів. **Досягнуті результати**: досліджено підходи до генерації текстів природною мовою та їх особливості; визначено архітектуру "Трансформер" як сучасний стандарт у сфері генерації текстової інформації; розглянуто види моделей на основі зазначеної архітектури; проаналізовано джерела даних для їх тренування; розглянуто методи адаптації моделей у вузькоспрямованих галузях; виявлено способи порівняння та оцінювання ефективності виконання різних завдань мовними моделями; виявлено недоліки наявних спеціалізованих мовних моделей і неповноту наборів метрик оцінювання завдання генерації договорів. Унаслідок аналітичного експерименту було визначено, що метод пошуково-доповненої генерації є найбільш оптимальним для розв'язання поставленого завдання в заданих умовах. Проведений експеримент та його результати можуть бути використані як основа для подальших досліджень у сфері розроблення вузькоспрямованих мовних моделей за умови обмежених ресурсів. **Висновки**. У статті проаналізовано методи генерації текстової інформації природною мовою за допомогою сучасних підходів машинного навчання. Виокремлено їх переваги й недоліки для невеликих компаній та наукових установ, які мають обмежені матеріальні та людські ресурси. Як приклад у роботі розглянуто спеціалізовану юридичну галузь і проблему генерації договорів та визначено найбільш оптимальний метод її розв'язання.

Ключові слова: велика мовна модель; генерація природної мови; договір; юридичний документ.

Вступ

Договори відіграють важливу роль у повсякденному житті людей та в ефективному функціонуванні компаній. Їх правильне формулювання та оформлення відповідно до чинного законодавства та інтересів зацікавлених сторін потребує вузькоспрямованих знань у галузі права. Складання угод вручну та внесення змін потребує чимало часу навіть для досвідчених юристів і може призвести до помилок під час копіювання його частин з інших документів та до низького рівня повторного використання цього договору в інших ситуаціях. Витрачаючи час на ці завдання, фахівці приділяють менше уваги клієнтам та розумінню їхніх потреб, що призводить до погіршення якості наданих послуг і неефективного використання часу й навичок.

Перелічимо проблеми, з якими найчастіше стикаються юристи та компанії, що працюють з договорами [1]:

– значна кількість часу та зусиль на складання та розуміння договорів;

– залежність неюридичних відділів компанії, таких як відділ кадрів і продажів, від команди юристів для укладання угод;

– неузгодженість між документами всередині компанії через різноманітність їх видів і формулювань.

Автоматизація процесу генерації та аналізу договорів за допомогою сучасних технологій може подолати ці проблеми. Методи машинного навчання, зокрема великі мовні моделі, показали себе найкраще у виконанні завдань генерації природної мови. Однак наявні комерційні рішення зазвичай перебувають у приватній власності компаній, а розроблення систем на основі мовних моделей потребує значних обсягів вузькоспрямованих тренувальних даних і чималих обчислювальних ресурсів, що обмежує потенціал використання цього методу невеликими компаніями, науковими інститутами та окремими дослідниками.

Аналіз останніх досліджень і публікацій

У сфері розуміння, оброблення та генерації природної мови в останні роки проведено чимало

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

досліджень. Автори роботи [2] запропонували архітектуру "Трансформер", яка замінила наявні рекурентні нейронні мережі, а в праці [3] її вдосконалено з метою отримання кращих результатів у галузі генерації тексту. Базові моделі, створені внаслідок багатьох досліджень, використовуються в різних галузях [4, 5]. Розроблення та вивчення вузькоспрямованих моделей показали переваги адаптації моделей до певної галузі порівняно із загальними моделями [6, 7]. Автори студій [8, 9] описують, як сучасні великі мовні моделі, що мають сотні мільярдів параметрів, без додаткового тренування можуть виконувати нові завдання в різних спеціалізованих галузях, використовуючи тільки інформацію та інструкції, які отримують з контексту. У роботі [10] сформовано великий корпус юридичних документів різними мовами для різних юрисдикцій, що спрощує для інших дослідників доступ до вузькоспрямованої інформації окресленої галузі. Автори праці [11] пропонують додати нову метрику оцінювання виконаних завдань з оброблення природної мови. Розроблені набори метрик дають змогу комплексно оцінювати знання моделей, їх здатність виконувати різні завдання та дотримуватися правил [12, 13]. У фахових дослідженнях сформовано аналогічні набори метрик для оцінювання ефективності моделей в обробленні та розумінні юридичних документів [11, 14, 15].

Визначення не розв'язаних раніше частин загальної проблеми. Мета роботи й завдання

Оброблення та генерація текстів природною мовою є нелегким завданням. Наявні підходи та моделі складні в розробленні, потребують значних обсягів обчислювальних, матеріальних і людських ресурсів. Ці фактори зумовили обмежене використання таких систем.

Поява архітектури "Трансформер", що є більш ефективною за попередні види нейронних мереж, відкрила нові можливості її застосування в галузях науки, виробництва та бізнесу. Попри це тренування й розроблення таких моделей стали складнішими, відповідно потребують ще більше ресурсів, які мають тільки корпорації або великі дослідницькі установи. Створення чималої кількості базових мовних моделей, що є у відкритому доступі, дали змогу багатьом невеликим компаніям і науковим інститутам досліджувати та розробляти нові системи.

Оскільки базові моделі натреновані на загальнодоступній інформації, їх можливості та навички в спеціалізованих галузях є обмеженими. Достатня кількість робіт присвячена дослідженню способів створення та адаптації моделей до вузькоспрямованих галузей. Інші праці присвячені розробленню моделей для певної царини та їх адаптації до вузького набору завдань і знань.

Автори студій у сфері оброблення юридичних документів і договорів приділяють увагу використанню тільки одного підходу – безпосередньому тренуванню параметрів моделі. Такі дослідження, як зазначалось раніше, потребують значних ресурсів і часу. У них зазвичай беруть участь багато науковців із різних університетів і компаній. Тому для невеликих комерційних і наукових організацій можливості використання новітніх технологій та підходів усе ще залишаються досить обмеженими. З огляду на окреслену проблему визначимо мету дослідження.

Метою роботи є аналіз підходів до розроблення вузькоспрямованих великих мовних моделей та визначення оптимального методу створення незалежних спеціалізованих систем, що уможливлюють генерацію договорів різними мовами в різних правових системах.

Сформулюємо *завдання*, що необхідно виконати для досягнення поставленої мети:

– визначення наявних компаній та рішень у цій сфері;

– виявлення підходів до створення текстів природною мовою;

– аналіз способів оцінювання та порівняння таких систем;

– виявлення обмежень і недоліків наявних рішень та підходів;

– пошук оптимального методу розроблення систем за умови обмежених ресурсів.

Перелічені завдання спрямовані на глибокий аналіз методів генерації текстових даних у вузькоспрямованих галузях та виявлення способів вирішення окресленої проблеми на прикладі генерації договорів у юридичній сфері.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Матеріали та методи

Огляд ринку

Перелічимо наявні компанії на ринку автоматизації юридичних документів (*legal document automation*):

– *Juro* – використовує велику мовну модель *GPT (Open AI)* для створення угод і уможливлює взаємодію із системою за допомогою *AI*-чату, підтримує англійську мову [16];

– *Luminance* – застосовує власну модель *Legal Inference Transformation Engine*, уможливлює автоматичне ведення переговорів щодо змісту договору (*Autopilot*) та підтримує понад 80 мов [17];

– *Icertis* – упроваджує систему *Icertis ExploreAI Service*, що поєднує можливості великих мовних моделей *Open AI*, власні *AI*-системи та *Icertis Data Lake* для аналізу та генерації документів, має підтримку декількох мов;

– *Oneflow* – використовує *GPT (Open AI)* і надає велику кількість готових шаблонів договорів, підтримує 10 європейських мов.

Серед основних функцій сервісів, що застосовують методи машинного навчання, можна виокремити такі:

– автоматичне складання договорів різних типів;

– аналіз і резюмування;

– підтримка правил та обмежень під час складання документів.

Для генерації договорів компанії найчастіше використовують нейронні мережі, а саме великі мовні моделі (*LLM*), розроблені спеціально для автоматизації документів, як у разі *Luminance*, або моделі загального призначення – *GPT* із додатковими модифікаціями для роботи в цій сфері.

З відкритих джерел встановлено, що модель *Legal Inference Transformation Engine* була натренована на понад 150 мільйонах перевірених юридичних документах. Більшість інших систем основані на моделі *OpenAI GPT*, тому ключові розбіжності між ними полягають у додатковому налаштуванні з використанням інформації в юридичній галузі. Однак щодо цих систем не вдалось отримати більш детальної інформації про підходи та дані, що використовуються для розроблення, оскільки вони є закритими та належать до інтелектуальної власності компаній.

Отже, бачимо, що сфера автоматизації юридичних документів активно адаптує та використовує сучасні підходи до оброблення природної мови, надаючи нові можливості формулювання угод, значно спрощуючи та прискорюючи цей процес. Автоматизація зазначеного процесу зменшує необхідну кількість юридичного персоналу компанії, даючи змогу неюридичним відділам складати угоди.

Аналіз методів генерації природної мови

Розглянемо підходи до розв'язання проблеми генерації природної мови (*NLG*). Найбільш ранні методи моделювання та генерації тексту використовували модульні архітектури, побудовані на наборі модулів, об'єднаних послідовно в системі. На зміну цим методам прийшли підходи, основані на плануванні, що визначали послідовність одного або декількох кроків для досягнення конкретної мети. Мета розбивалася на менші завдання, які виконувалися за допомогою дій, що мали певний набір умов та ефектів, які впливали на кінцевий результат. Наступним етапом розвитку були стохастичні підходи, що активно використовували набори даних для виявлення статистичних залежностей у природній мові для подальшої генерації тексту.

Нейронні мережі були найбільш популярними та ефективними підходами для розв'язання проблем моделювання тексту й машинного перекладу. Розглянемо деякі з найбільш використаних моделей:

– рекурентні нейронні мережі (*RNN*);

– мережі з довгою короткочасною пам'яттю (*LSTM*);

– вентильні рекурентні вузли (*GRU*);

– варіаційні автокодувальники (*VAE*);

– згорткові нейронні мережі (*CNN*);

– генеративні змагальні мережі (*GAN*) [18].

Попри широке використання окреслених підходів, вони мають певні недоліки, які обмежують їх ефективність для виконання завдання генерації текстів.

Рекурентні нейронні мережі мають тенденцію зазнавати проблеми зникнення градієнта на довгих послідовностях, що може обмежувати їх здатність до генерації довгих текстів зі складною структурою. Їх можна використовувати для генерації коротких фрагментів тексту або в завданнях, де контекст не потребує глибокого аналізу.

LSTM та *GRU* краще працюють із довгими залежностями в тексті, порівняно зі звичайними *RNN*, завдяки своїм механізмам керування пам'яттю та можуть бути ефективнішими для генерації більш складних текстів. Проте через свою рекурентну природу навчання моделей відбувається послідовно, а можливості паралелізації є обмеженими.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Варіаційні автокодувальники використовуються для генерації нових зразків зазвичай із деякою змінністю відповідно до вхідної інформації та застосовуються для створення різноманітності у тексті. Основною проблемою цього підходу є згортання розходження Кульбака – Лейблера, що призводить до генерації вихідних даних незалежно від вхідних.

Хоча згорткові нейронні мережі зазвичай використовуються для оброблення зображень, вони можуть бути адаптовані для роботи з текстом та бути ефективними для його генерації з огляду на локальні шаблони та структури. Вони широко не використовувалися через проблеми з вибором архітектури та оптимального значення гіперпараметрів.

GAN-моделям властива здатність генерувати реалістичні дані, зокрема природну мову, і вони можуть бути застосовані для створення тексту з високою якістю та натуралізмом. Проте навчання таких моделей є більш складним через недиференційну природу дискретних символів, а також модель може генерувати поверхневі, повторювані та недалекоглядні відповіді.

Сучасним підходом до генерації тексту є архітектура "Трансформер", що має структуру "кодувальник-декодувальник" [2]. Основним принципом запропонованого підходу є механізм самоуваги, що визначає важливість частин вхідної послідовності токенів до інших слів у цій послідовності. Зазначений підхід, на відміну від попередніх, дає змогу:

– виконувати обчислення паралельно, зменшуючи необхідний час тренування;

– обробляти послідовності тексту більшої довжини без втрати контексту.

З огляду на отримані метрики (*benchmarks*) ця архітектура демонструє кращі результати виконання певних завдань оброблення природної мови, наприклад, машинного перекладу [2].

Для генерації тексту використовується варіація оригінальної архітектури, що має тільки декодувальник та генерує тестову послідовність на основі початкового вхідного тексту *(prompt*).

Ці характеристики зумовили стрімкий розвиток моделей на основі архітектури "Трансформер". Нові можливості генерації природної мови викликали значний інтерес різних сфер бізнесу, що сприяло впровадженню систем на основі мовних моделей у багатьох індустріях за останні декілька років.

Аналіз джерел даних

Для навчання моделей використовують великі за обсягом набори даних, що можуть містити терабайти корпусів тексту.

Зміст тренувальних даних *загального* призначення не обмежується однією галуззю, що робить їх більш придатними для навчання загальних моделей. Ці дані можна поділити на декілька основних класів [19]:

– текст вебсторінок, що отримується за допомогою сканування великої кількості вебсторінок в інтернеті та визначається чималим обсягом, динамічністю змісту, наявністю різних мов та багатьох тем, високим рівнем неперевіреної інформації (*Common Crawl, C4, mC4*);

– книги, яким властива висока якість змісту, граматична та лексична точність, значна довжина тексту, наявність складних мовних зворотів, термінів і фразеологізмів (*Anna's Archive, BookCorpusOpen*);

– академічні матеріали, які мають високий рівень професіоналізму та знань, що зумовлює виняткову якість їх робіт (*arXiv, PubMed Central*);

– програмний код, що містить приклади використання мов програмування для розв’язання різних завдань (*BIG-QUERY* та *phi-1*);

– дані соціальних медіа, що охоплюють створені користувачами дописи, коментарі та діалоги й визначаються потенційною присутністю шкідливої інформації, такої як упередження, дискримінація та насильство (*Pushshift Reddit* та *OpenWebText*);

– дані енциклопедій, які є в онлайн-енциклопедіях або інших базах знань і яким властивий певний рівень надійності інформації (*Wikipedia*).

Вузькоспрямовані моделі потребують навчальної інформації, яка містить знання, особливі для певної галузі.

У юриспруденції *Pile of Law* та *MultiLegalPile* є найбільшими наборами даних [10]. *Pile of Law* містить близько 256 ГБ правової та адміністративної інформації. Для її формування використано 35 різних джерел, зокрема юридичні документи, судові висновки, публікації державних установ, контракти, статути, нормативні акти, журнали справ тощо. *MultiLegalPile* об’єднує 689 ГБ юридичних документів 24 мовами з 17 різних юрисдикцій. Він містить набір *Pile of Law*, а також декілька натренованих моделей на основі *RoBERTa* та *Longformer*. Ці набори даних можна використовувати для тренування як одномовних, так і багатомовних моделей і адаптувати їх під законодавство різних країн.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Базові моделі

Розроблення мовних моделей потребує значних ресурсів, часу та навичок. Більшість компаній та дослідницьких установ не мають достатньо матеріальних і людських ресурсів для тренування нових моделей. А час, необхідний для досягнення бажаних результатів, може виявитися занадто тривалим.

Для тренування *LLaMA* – базової моделі з відкритим доступом, розробленої компанією *Meta*, з 65 мільярдами параметрів – використано 2048 *Nvidia* A100 *GPU*, кожен з яких мав 80 ГБ пам'яті *RAM* [20]. Тренування на наборі даних обсягом 1,4 мільярда токенів тривало приблизно 21 день. Вартість застосування таких ресурсів за оцінками може становити приблизно 2,4 мільйона доларів [21].

Щоб зробити великі мовні моделі доступнішими та відкрити перспективи ширших досліджень навіть для окремих фахівців, корпорації та великі компанії, що спеціалізуються на мовних моделях, публікують у відкритий доступ базові, попередньо натреновані моделі (*foundational models*):

– *Google: LaMDA, Chinchilla, Gemma* [22, 23, 24];

– *Meta: LLaMA 2* [4];

– *Mistral: Mixtral* [5].

Зазвичай перелічені моделі містять загальні знання та можуть виконувати різні завдання, проте їх використання для вузькоспрямованих завдань та в специфічних сферах є обмеженим. Подальше налаштування може дати змогу моделі набувати нових знань та виконувати нові завдання, потребуючи значно менше ресурсів для навчання, порівняно з розробленням нової моделі.

Вузькоспрямовані комерційні моделі

Розроблення вузькоспрямованих моделей у певних галузях показало їх перевагу над моделями загального призначення.

Компанія *Microsoft* розробила продукт *Microsoft Sales Copilot*, що допомагає менеджерам з продажу збільшувати ефективність роботи та створювати персоналізовані пропозиції клієнтам, використовуючи *GPT*-моделі від *OpenAI* з додатковим налаштуванням у сфері продажу [25]. Система дає змогу створювати персоналізовані листи на основі інформації про клієнта, деталей угоди та попередньої комунікації із замовниками. Також можна аналізувати наради й зустрічі, додаючи виділення ключових слів, теми розмов, конкурентів, ключові метрики оцінювання ефективності та запропоновані завдання.

Компанія *Luminance* спільно з дослідниками з Кембриджського університету розробили модель *Legal Pre-Trained Transformer*, що була натренована на понад 150 мільйонах перевірених юридичних документів [26]. Ця модель дає змогу генерувати документи, зокрема контракти й договори, аналізувати та резюмувати їх зміст (*summarization*). Чат-бот на основі зазначеної моделі може вести переговори щодо змісту контракту, перевіряти компанії на відповідність вимогам та обмеженням, виявляти зміни в нових версіях [17].

Основним недоліком розглянутих моделей є те, що вони належать компаніям і доступ до них є закритим. Це обмежує можливості їх використання, покращення та незалежного оцінювання іншими дослідниками. Тому постає питання розроблення власних спеціалізованих систем.

Методи тренування та налаштування вузькоспрямованих моделей

Для розроблення вузькоспрямованих мовних моделей упроваджують різні підходи, що відрізняються за кількістю необхідних ресурсів і часу, ефективністю та обсягом тренувальних даних.

Тренування нової вузькоспрямованої моделі, наприклад *Legal Pre-Trained Transformer*, є найбільш складним підходом. Цей процес схожий до того, як розробляються базові моделі або моделі загального призначення.

Для налаштування параметрів моделі на основі архітектури "Трансформер" застосовується механізм самоуваги [2]:

$$Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \ , \qquad (1)$$

де d – розмірність моделі;

n – кількість запитів;

m – кількість пар "ключ – значення";

$Q \in \mathbb{R}^{n\times d_k}$ – матриця запитів;

$K \in \mathbb{R}^{m\times d_k}$ – матриця ключів;

$V \in \mathbb{R}^{m\times d_v}$ – матриця значень;

$\sqrt{d_k}$ – коефіцієнт масштабування.

Механізм багатоголової самоуваги паралельно обчислює функції самоуваги (1) для отримання

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

інформації з декількох підпросторів подання на різних позиціях:

$$MultiHead\left(Q,K,V\right)=Concat\left(head_1,\ldots,head_h\right)W^o, \quad (2)$$

$$head_i=Attention\left(QW_i^Q,KW_i^K,VW_i^V\right),$$

де h – кількість голів самоуваги;

W^o – фінальна матриця ваг, отримана від усіх голів самоуваги;

W_i^Q – матриця ваг запитів;

W_i^K – матриця ваг ключів;

W_i^V – матриця ваг значень.

Кожен шар моделі додатково містить нейронну мережу прямого зв'язку:

$$FFN\left(x\right)=\mathrm{Re}LU\left(xW_1+b_1\right)W_2+b_2, \quad (3)$$

де x – вхідний вектор;

$W_1\in\mathbb{R}^{d\times d_m}$ і $W_2\in\mathbb{R}^{d_m\times d}$ – матриці ваг лінійних трансформацій.

Основною розбіжністю тренування вузькоспрямованих моделей від моделей загального призначення є використання властивих для певної галузі даних, що навчають модель нових знань і завдань. Перевагами цього методу є висока ефективність моделі, повний контроль над тренувальною інформацією та можливість адаптації до різних завдань і потреб. Контроль над навчальними даними також дає змогу визначити мови, які підтримуватимуться системою. Як і в ситуації тренування базових моделей, недоліками окресленого підходу є необхідність у значних обсягах тренувальної інформації, а також обчислювальних ресурсах та часі, що зумовлюють вартість розроблення моделі. Збір вузькоспрямованих тренувальних наборів може викликати труднощі, оскільки дані можуть бути недоступні для широкого використання, належати до інтелектуальної власності інших компаній або їх якість та обсяг можуть бути недостатніми для навчання великих систем.

У процесі роботи з формування набору даних *MultiLegalPile* було натреновано нові вузькоспрямовані моделі в галузі права *Legal-XLM-R* за допомогою зібраної інформації. Вони демонструють значно вищі показники під час оцінювання на наборах метрик *LEXTREME* та *LexGLUE*, якщо порівнювати з моделями загального застосування (*DeBERTa* та *RoBERTa*), а також мають переваги щодо інших вузькоспрямованих моделей у цій галузі (*Legal-BERT* та *CaseLaw-BERT*) [10].

Тонке налаштування (*fine-tuning*) великих мовних моделей для конкретних галузей передбачає адаптацію попередньо навченої моделі для кращого розуміння та генерації тексту, що відповідає цьому домену. Розглянемо основні підходи тонкого налаштування.

Під час *повного тонкого налаштування* (*full-fine tuning*) попередньо навчена мовна модель налаштовується для роботи з інформацією, властивою для певної галузі [23]. Цей підхід передбачає оновлення всіх параметрів моделі під час тренування, що дає змогу їй отримати знання в окресленій сфері та вивчити нові закономірності. Проте налаштування вимагає великого обсягу інформації для ефективного виявлення та розуміння особливостей цільової галузі. Хоча цей підхід може бути ефективним, він також потребує значних обчислювальних ресурсів та часу на тренування, насамперед для великих моделей, що мають мільярди параметрів. Нещодавні дослідження у сфері оптимізації тренувального процесу показали, що за допомогою нових методів, зокрема *LOMO (LOw-Memory Optimization)*, можна виконувати налаштування моделей навіть із десятками мільярдів параметрів лише на декількох *GPU*-процесорах [27].

Для розв'язання проблем попереднього підходу розроблено *параметро-ефективні методи* тонкого налаштування (*PEFT*) [28]. Вони передбачають упровадження різних методів глибокого навчання для зменшення кількості параметрів, які треба навчити, і водночас зберігають схожий рівень ефективності повного тонкого налаштування. *PEFT*-методи оновлюють лише незначну кількість додаткових параметрів або підмножину попередньо навчених, зберігаючи наявні знання моделі та адаптуючи їх до нового завдання, що зменшує ризик катастрофічного забування. Застосування повного тонкого налаштування на специфічних тренувальних даних, обсяг яких зазвичай набагато менший, ніж набір даних, що використовувався для тренування базової моделі, може призвести до перенавчання. *PEFT*-методи дають змогу розв'язати цю проблему способом вибіркового оновлення попередньо навчених параметрів або їх повного "заморожування".

Виокремлюють декілька основних груп параметро-ефективних методів тонкого налаштування.

Підхід *адитивного* тонкого налаштування (*Additive Fine-tuning*) додає нові параметри під час налаштування моделі для нового завдання, поділяється на методи на основі адаптерів (*adapter-based*), м'якого налаштування на основі підказок (*soft prompt-based*) тощо, наприклад *AttentionFusion* і *AdapterFusion*. Метод

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

послідовного адаптера додає мережі адаптерів після шарів самоуваги та нейронної мережі прямого зв'язку [28]. Кожен адаптер є модулем нижчого рангу, що містить низхідну проєкцію, нелінійну функцію активації та висхідну проєкцію, а також залишкового з'єднання. Для вхідної інформації h результат обчислюється таким чином:

$$h = h + \left(\mathrm{Re}\,LU\left(hW_{down}\right)\right)W_{up}\ , \qquad (4)$$

де h – вхідний вектор;

$W_{down} \in \mathbb{R}^{d\times r}$ – низхідна проєкція;

$W_{up} \in \mathbb{R}^{r\times d}$ – висхідна проєкція;

$\mathrm{Re}\,LU\left(x\right) = \max\left(0, x\right)$ – нелінійна функція активації.

Підхід *часткового* тонкого налаштування (*Partial Fine-tuning*) зменшує кількість налаштовуваних параметрів способом обрання критичних попередньо навчених ваг і відкидання неважливих, поділяється на методи оновлення зміщення (*Bias Update*), маскування попередньо натренованих ваг (*Pretrained Weight Masking*) та дельта-маскування ваг (*Delta Weight Masking*). Метод порогового маскування використовує визначене значення порогу τ для побудови матриці двійкових масок M з метою вибору попередньо навчених ваг W шарів самоуваги та нейронної мережі прямого зв'язку з допомогою множення елементів матриці:

$$W' = W \odot M\ , \qquad (5)$$
$$M = \mathbb{I}_{s_{i,j}>\tau},$$

де $\odot$ – добуток Адамара;

τ – визначений поріг;

$W \in \mathbb{R}^{d\times k}$ – ваги моделі;

$S \in \mathbb{R}^{d\times k}$ – матриця, що ініціалізується випадковими рівномірно розподіленими дійсними числами; якщо елемент матриці перевищує поріг τ, відповідній позиції в матриці двійкових масок присвоюється значення 1, інакше – 0.

Підхід *перепараметризованого* тонкого налаштування (*Reparameterized Fine-tuning*) застосовує перетворення низького рангу, щоб зменшити кількість параметрів для навчання. Він поділений на методи розкладання низького рангу (*Low-rank Decomposition*), наприклад *Intrinsic SAID*, і методи на основі *LoRA* (*Delta-LoRA*). Метод *LoRA* (*Low-Rank Adaptation*) додає дві матриці низького рангу, що оновлюються в процесі тренування [29]. Низхідна та висхідна матриці проєкцій використовуються паралельно з матрицями запитів Q, ключів K і значень V у шарі самоуваги моделі. Під час навчання оновлюються тільки матриці W_{down} та W_{up}. Нові ваги обчислюються таким чином:

$$\Delta W = W_{down}W_{up}\ , \qquad (6)$$

де $W_{down} \in \mathbb{R}^{d\times r}$ – низхідна проєкція;

$W_{up} \in \mathbb{R}^{r\times k}$ – висхідна проєкція;

$r \ll \left\{k, d\right\}$ – обрана розмірність.

Під час генерації результату моделлю вагові коефіцієнти ΔW об'єднуються з вихідною матрицею ваг

$$h = W_0 x + \Delta W x\ , \qquad (7)$$

де x – вхідний вектор;

$W_0 \in \mathbb{R}^{d\times k}$ – ваги моделі;

$\Delta W \in \mathbb{R}^{d\times k}$ – додатково натреновані ваги за допомогою методу *LoRA*.

Підхід *гібридного* тонкого налаштування (*Hybrid Fine-tuning*) поєднує різні підходи *PEFT* для посилення їх переваг і зменшення недоліків. Розрізняють такі підходи до комбінації методів: ручні, що вимагають складного дизайну (*MixAnd-Match Adapter, Compacter*), та автоматичні, які використовують структурний пошук (*AutoPEFT*). Метод *Compacter* розроблено на основі підходів адаптера, розкладання низького рангу та параметризованого гіперкомплексного множення (*parameterized hypercomplex multiplication*) [30]. Він має схожу до адаптерів структуру, однак замінює низхідну та висхідну проєкції шаром параметризованого гіперкомплексного множення низького рангу (*low-rank parameterized hypercomplex multiplication*). Значення ваг обчислюються таким чином:

$$W = \sum_{i=1}^{n} A_i \otimes B_i\ , \qquad (8)$$

де $A_i \in \mathbb{R}^{n\times n}$ – загальна матриця ваг для всіх шарів адаптера;

$B_i \in \mathbb{R}^{\frac{k}{n}\times\frac{d}{n}}$ – матриця ваг, притаманна для окремих шарів адаптера;

$\otimes$ – добуток Кронекера.

Підхід *уніфікованого* тонкого налаштування (*Unified Fine-tuning*) є єдиною структурою для додавання різноманітних методів налаштування в єдину архітектуру (*AdaMix* та *ProPETL*) і зазвичай використовує один *PEFT*-метод, а не комбінацію різних. Метод *AdaMix* містить M модулів адаптації, що долучаються в кожен шар моделі:

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

$A_{ij}: i \in [1, L]1, L, j \in [1, M]$ – j-й модуль адаптації в i-му шарі [31]. Під час тренування на кожному кроці випадковим способом обирається пара матриць проєкцій для i-го шару:

$$A_i = \{W_{ij}^{up}, W_{ik}^{down}\}, \quad (9)$$

$$B_i = \{W_{ij'}^{up}, W_{ik'}^{down}\},$$

де W_{ij}^{down} – низхідна проєкція i-го шару j-го модуля адаптації;

W_{ij}^{up} – висхідна проєкція i-го шару j-го модуля адаптації.

Отже, вся вхідна інформація обробляється тим самим набором модулів. Використовуючи матриці, наведені у (9), відбувається така трансформація:

$$h = h + f\left(h \cdot W^{down}\right) W^{up} \quad (10)$$

де h – вхідний вектор;

W^{down} – низхідна проєкція;

W^{up} – висхідна проєкція;

$f(x)$ – функція активації.

Серед основних переваг параметро-ефективних методів можна виокремити необхідність у значно меншому обсязі вузькоспрямованих тренувальних даних і обчислювальних ресурсів, оскільки цей метод тренує тільки незначну частину параметрів. Недоліками підходу є менша ефективність моделі та необхідність додаткового тренування параметрів для виконання різних завдань, властивих для обраної сфери. Набір мов, з якими модель може взаємодіяти, обмежується тими, які підтримує базова модель. Додавання нових мов потребує значних обсягів інформації, щоб навчити модель різних аспектів та особливостей її використання. Зауважимо, що це майже неможливо за допомогою цього методу через незначну кількість ваг, які оновлюються.

Метод *пошуково-доповненої генерації* (*Retrieval-Augmented Generation*) ефективно застосовується в галузях, що потребують доступу до актуальних і точних знань, які постійно оновлюються та змінюються [19]. Він оснований на використанні зовнішніх джерел, що містять перевірену та точну інформацію. Під час генерації відповіді система шукає та отримує додаткові релевантні дані із цих джерел. На підставі запиту, контексту й додаткової інформації модель дає більш точні та аргументовані відповіді, основані на достовірних фактах.

RAG-модель описується таким чином: на основі вхідної послідовності x система отримує текстові документи z і використовує їх як додатковий контекст під час генерації цільової послідовності y. Основними елементами системи є пошуковий компонент і генератор. Пошуковий компонент $p_\eta(z|x)$ знаходить k найбільш релевантних фрагментів тексту на основі x. *Dense Passage Retriever* використовує щільний кодувальник, який перетворює фрагменти тексту в d-вимірні вектори дійсних чисел та створює індекс для всіх фрагментів тексту, що застосовуються для пошуку [32]. Компонент генератора $p_\theta(y_i|x, z, y_{1:i-1})$, який є великою мовною моделлю, генерує токени тестової послідовності на основі контексту попередніх $i-1$ токенів $y_{1:i-1}$, вхідних даних x та фрагментів z, отриманих від пошукового компонента.

RAG-модель на основі токенів дозволяє компоненту генератора обирати вміст із кількох документів під час генерації відповіді. Найбільш релевантні k-документи отримуються за допомогою пошукового компонента. Наступний токен обчислюється таким чином:

$$p(y|x) \approx \prod_i^N \sum_{z \in top-k(p(\cdot|x))} p_\eta(z|x) p_\theta(y_i|x, z, y_{1:i-1}), \quad (11)$$

де $top-k\left(p(\cdot|x)\right)$ – k-документи з набору z найвищої попередньої ймовірності $p_\eta(z|x)$.

Можна виокремити три групи підходів, що використовуються для реалізації зазначеного методу [33]:

– наївний (*naive*) – найпростіший підхід, оснований на індексації даних, отриманні їх зі сховища та генерації відповіді; можуть траплятися галюцинації, а отримана інформація може бути нерелевантною та повторюваною;

– розширений (*advanced*) – зосереджується на покращенні якості отриманої інформації способом додаткового оброблення перед пошуком і після нього; використовує індексування за допомогою підходів ковзного вікна, дрібної сегментації та додавання метаданих;

– модульний (*modular*) – містить різноманітні стратегії для вдосконалення компонентів способом додавання нових модулів: пошукових, злиття (*fusion*), пам'яті, маршрутизації, передбачення та адаптації до завдання.

Метод пошуково-доповненої генерації має низку переваг. Він значно дешевший для впровадження

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

та використання, оскільки застосовує наявні моделі як основу та не потребує додаткового тренування. Завдяки отриманню даних із баз знань система має доступ до інформації, яка з'явилась після тренування моделі та не була додана до її тренувального набору. Звернення до перевірених та достовірних джерел інформації зменшує рівень галюцинацій.

Основними викликами під час його використання можна назвати отримання якісної інформації з баз знань, їх оцінювання та впорядкування за актуальністю. Метод не передбачає тренування та зміни параметрів моделі, тому загальні знання та завдання, які модель може виконувати, залишаються незмінними. Так само підтримка різних мов обумовлюється обраною базовою моделлю.

Метод *навчання з контексту* (*in-context learning*) дає змогу мовним моделям виконувати нові завдання на основі інструкцій природної мови та кількох прикладів, що демонструють виконання нового завдання та надаються моделі через вхідний текст (*prompt*). Зазначений метод оснований на можливостях великих мовних моделей розвивати широкий набір навичок і здібностей під час навчання, а потім використовувати їх під час генерації відповідей, швидко адаптуватися до нового завдання та його розпізнавання.

У межах цього підходу можна виокремити три категорії інструкцій, що допомагають моделі навчитися нового завдання:

– навчання на основі декількох прикладів (*few-shot learning*) – модель отримує інструкції, що описують завдання, та демонстрації його виконання (зазвичай від 10 до 100);

– навчання на основі одного прикладу (*one-shot learning*) – передбачає інструкції та лише одну демонстрацію;

– навчання без прикладів (*zero-shot learning*) – надає тільки інструкції за допомогою природної мови, проте не наводить жодних прикладів.

Дослідження на прикладі *GPT*-3 та інших моделей меншого розміру виявили таке: що більша модель, то кращі результати можна отримати, застосовуючи навчання з контексту [8]. Додаткові експерименти з використання цього методу та *GPT*-4 моделі в галузі медицини продемонстрували можливості досягнення кращих результатів, якщо порівнювати з вузькоспрямованою моделлю *Med-PaLM* 2, яка пройшла тонке налаштування на декількох спеціалізованих наборах даних [7, 9]. Під час цього дослідження розроблено підхід *Medprompt*, що формує вхідний текст, використовуючи декілька різних технік: динамічний вибір декількох прикладів (*few-shot*), автоматично створений ланцюжок думок (*chain of thought*) та ансамблевий метод вибору (*choice shuffling ensemble*). Його було застосовано до інших галузей знань, зокрема до права. Ефективність оцінювалася за допомогою набору метрик *MMLU (Massive Multitask Language Understanding)* [27]. Експеримент показав, що без додаткового налаштування *GPT*-4 отримав оцінку 68,3. Унаслідок додаткового застосування підходу *Medprompt* досягнуто значно вищий результат – 72,9 бала.

Основною перевагою зазначеного методу є відсутність налаштування параметрів, що унеможливлює витрати на обчислювальні ресурси, необхідні для тренування та зміни ваг. Також він потребує значно меншого обсягу спеціалізованої інформації порівняно з тонким налаштуванням. Складання прикладів та інструкцій є інтуїтивним процесом, схожим на те, як люди взаємодіють між собою під час опису та визначення завдань. Це дає змогу експертам певної галузі, не маючи знань про машинне навчання та мовні моделі, впроваджувати метод для виконання нових вузькоспрямованих завдань.

Недоліками навчання з контексту є обмеження розміру вхідного тексту, що лімітує кількість прикладів і розмір інструкцій, які можна навести. Ефективність цього методу зазвичай є меншою порівняно з тренуванням нової моделі та тонким налаштуванням. Набір підтримуваних мов зазвичай не може бути розширеним.

Аналіз ефективності моделей

Для оцінювання ефективності та надійності роботи мовних моделей використовують різні метрики. Найпростішими з них є *ROUGE* та *BERTScore*.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) вимірює повноту відповіді та застосовується для оцінювання якості резюмування та генерації тексту. Існує декілька варіантів цієї метрики, зокрема *ROUGE-N* (вимірює перекриття n-грам), *ROUGE-L* (вимірює найдовшу спільну підпослідовність) та *ROUGE-W* (вимірює зважену найдовшу спільну підпослідовність).

BERTScore обчислює схожість між згенерованим та еталонним текстом [34]. Обидві послідовності подаються за допомогою векторів вбудовувань (*embeddings*). Схожість між ними обчислюється за допомогою косинуса подібності (*cosine similarity*).

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Зі зростанням необхідності оцінювати більші та складніші моделі розроблено набори тестів (*benchmarks*), що оцінюють їх рівень знань і можливості виконання завдань.

SuperGLUE – фреймворк для оцінювання здібностей моделей розуміти тексти природною мовою [13]. Він містить різні групи завдань англійською мовою: відповіді на запитання з одним чи декількома правильними варіантами, класифікація тексту та визначення причинно-наслідкових зв'язків.

MMLU (Massive Multitask Language Understanding) – набір тестів загального призначення, розроблений для перевірки знань, набутих на етапі навчання моделі із застосуванням методу навчання з контексту [27]. Він охоплює 57 різних предметів, як загальних (математика та історія), так і більш фахових (право та медицина). Складність завдань коливається від початкового до професійного рівня, що перевіряють як знання про світ, так і здатність вирішувати проблеми.

Для юриспруденції також розроблено фахові набори тестів.

LexGLUE (Legal General Language Understanding Evaluation) – набір тестів для юридичної галузі, що містить сім різних наборів даних англійською мовою та перевіряє здібності моделей у виконанні завдань класифікації тексту та відповіді на запитання з декількома варіантами [14].

LegalBench – набір тестів, що містить 162 метрики [11]. Вони охоплюють шість типів юридичних завдань, які по-різному оцінюють модель: виявлення проблем, згадування правил, їх застосування, прийняття рішення на основі правил, тлумачення тексту, розуміння риторики.

LEXTREME – інший фреймворк, що передбачає 11 наборів даних, які охоплюють 24 мови [15]. Він містить три групи завдань: класифікація з визначення одного та декількох класів і розпізнавання іменованих об'єктів у юридичних документах.

Можна помітити, що більшість загальних і спеціалізованих наборів метрик зосереджуються на аналітичних здібностях моделі та її розумінні природної мови: класифікація та маркування тексту, резюмування, прогнозування висновків судових справ, розпізнавання іменованих сутностей. Однак вони майже не приділяють увагу завданню генерації нового контенту та його оцінюванню.

Пошук оптимального підходу

Визначимо критерії порівняння різних підходів тренування великих мовних моделей для використання у сфері генерації договорів.

Обсяг обчислювальних ресурсів визначає кількість серверів, що оптимізовані для машинного навчання та мають відповідні *GPU*-процесори, необхідні для навчання моделі за прийнятний час. Критерій є категоріальним, оскільки абсолютне порівняння не буде ефективним для оцінювання через значну розбіжність для різних методів. Використаємо такі категорії: великий – кількість обчислювальних серверів перевищує 10, малий – менше ніж 10 серверів, відсутній – не потребує додаткових обчислювальних ресурсів для тренування.

Обсяг тренувальних даних визначає розмір вузькоспрямованого тренувального набору, що дасть змогу моделі набути знань і навичок у новій сфері й ефективно виконувати поставлені завдання. Критерій є категоріальним, оскільки абсолютне порівняння не буде ефективним для оцінювання через значну варіацію необхідних обсягів для різних методів. Використаємо такі категорії: великий – кількість тренувальних даних понад 10 ГБ, середній – від 10 МБ до 10 ГБ, малий – менше ніж 10 МБ. Варто зауважити, що для методу навчання з контексту обсяг інформації обмежується розміром вхідного тексту, який модель може обробити [35]. Аналогічно й метод пошуково-доповненої генерації може застосовувати обмежений обсяг інформації для кожного запиту. Проте загальний набір даних, що використовується для пошуку найбільш релевантної інформації, може бути значно більшим.

Час тренування визначає обсяг необхідних часових ресурсів для налаштування моделі з використанням спеціалізованих даних. Критерій є категоріальним, оскільки абсолютне порівняння не буде ефективним для оцінювання через значну розбіжність значень для різних методів. Використаємо такі категорії: великий – тренування потребує понад 24 год, середній – від 1 до 24 год, малий – менше ніж 10 хв. Варто зауважити, що критерій має одноразовий ефект на тренування моделі за допомогою методів, що змінюють ваги системи. Для методів пошуково-доповненої генерації, оскільки вони не модифікують параметри, значення цього критерію визначає час, необхідний для пошуку й оброблення релевантної інформації, і впливає на кожен запит до моделі в процесі генерації відповіді. У разі використання

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

методу навчання з контексту тренувальні дані є частиною запиту, тому навчання моделі за допомогою цих даних є частиною генерації відповіді.

Контроль тренувальних даних визначає можливість значно впливати на інформацію, що застосовується для навчання моделі. Критерій є категоріальним. Використаємо такі категорії: повний – розробники самостійно визначають тип даних, їх походження та обсяги, частковий – можливість визначати тільки частину тренувальної інформації, зазвичай вузькоспрямовані набори даних, відсутній – неможливість значно вплинути на інформацію, яку модель застосовує під час генерації. Зауважимо, що для методів пошуково-доповненої генерації та навчання з контексту тренувальною є інформація, що модель отримує через вхідний текст. Хоча ці дані впливають на процес генерації відповіді, вони не запам'ятовуються моделлю та не будуть використані для наступних запитів.

Підтримка великих документів визначає можливість обробляти значну за обсягом інформацію на десятки сторінок. Критерій має два значення: так чи ні. Зауважимо, що методи тренування та тонкого налаштування обробляють такі документи під час налаштування параметрів. За умови впровадження інших методів розмір документа обмежується розміром вхідного тексту моделі. У разі застосування методу пошуково-доповненої генерації можна зменшити вплив цього фактора завдяки використанню тільки найбільш релевантних і значущих частин документа.

Підтримка нових мов визначає можливість упровадження моделі в різних країнах та юрисдикціях. Критерій має два значення: так – існує можливість навчити систему нової мови та ефективно її використовувати; ні – набір мов обмежується базовою моделлю, а впровадження вузькоспрямованої інформації іншими мовами не матиме значного ефекту.

Обсяг обчислювальних ресурсів для актуалізації знань визначає кількість серверів, необхідних для оновлення даних моделі у разі зміни законодавства та правил. Критерій є категоріальним, оскільки абсолютне порівняння не буде ефективним для оцінювання через значну розбіжність для різних методів. Використаємо такі категорії: великий – кількість обчислювальних серверів перевищує 10, малий – менше ніж 10 серверів, відсутній – не потребує додаткових обчислювальних ресурсів. У разі застосування підходів тренування нової моделі, повного та параметро-ефективного тонкого налаштування модель набуває нових знань унаслідок повторного процесу налаштування ваг. За умови використання інших методів модель набуває актуальних знань під час генерації кожної відповіді та не потребує додаткового тренування.

Рівень підтримки уваги до бізнес-правил визначає здатність моделі дотримуватися заданих правил і обмежень. Критерій є категоріальним. Використаємо такі категорії: високий – можливість підтримки багатьох правил різної складності, середній – базові нескладні правила, низький – базові нескладні правила в обмеженому обсязі. У разі застосування методів тренування та тонкого налаштування модель під час тренування навчається виконувати завдання аналізу та уваги до правил. Підхід навчання з контексту може використовувати шаблони вхідного тексту, що містять обмеження, а здатність системи дотримуватися їх визначається можливостями базової моделі. Аналогічно й підхід пошуково-доповненої генерації, крім отримання інформації з джерел на запит користувача, може визначити релевантні правила та додати їх до контексту запиту.

Значення критеріїв для кожного методу налаштування моделі наведено в табл. 1.

Таблиця 1. *Критерії оцінювання методів тренування вузькоспрямованих мовних моделей*

Критерій	Тренування моделі	Повне тонке налаштування	Параметро-ефективне тонке налаштування	Пошуково-доповнена генерація	Навчання з контексту
Обсяг обчислювальних ресурсів	Великий	Великий	Малий	Відсутній	Відсутній
Обсяг тренувальних даних	Великий	Великий	Середній	Середній	Малий
Час тренування	Тривалий	Тривалий	Нетривалий	Відсутній	Відсутній
Контроль тренувальних даних	Повний	Частковий	Частковий	Відсутній	Відсутній
Підтримка великих документів	Так	Так	Так	Так	Ні
Підтримка нових мов	Так	Так	Ні	Ні	Ні
Обсяг обчислювальних ресурсів для оновлення знань	Великий	Великий	Малий	Відсутній	Відсутній
Рівень підтримки уваги до бізнес-правил	Високий	Високий	Високий	Середній	Низький

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Використовуючи розглянуті вище критерії, визначимо найбільш оптимальний підхід за допомогою методу адитивного згортання з ваговими коефіцієнтами.

Спочатку перетворимо категоріальні критерії до числових значень:

– обсяг обчислювальних ресурсів: великий – 2, малий – 1, відсутній – 0;

– обсяг тренувальних даних: великий – 2, середній – 1, малий – 0;

– час тренування: тривалий – 2, нетривалий – 1, відсутній – 0;

– контроль тренувальних даних: повний – 2, частковий – 1, відсутній – 0;

– підтримка великих документів: так – 1, ні – 0;

– підтримка нових мов: так – 1, ні – 0;

– обсяг обчислювальних ресурсів для оновлення знань: великий – 2, малий – 1, відсутній – 0;

– рівень підтримки уваги до бізнес-правил: високий – 2, середній – 1, низький – 0.

З метою спрощення подання інформації використаємо абревіатури для методів налаштування моделей: тренування моделі – ТМ, повне тонке налаштування – ПТН, параметро-ефективне тонке налаштування – ПЕТН, пошуково-доповнена генерація – ПДГ, навчання з контексту – НК.

Далі виконаємо нормування кожного з них на проміжку [0, 1], використовуючи мінімальне й максимальне значення, та подамо результати в табл. 2.

Таблиця 2. *Нормовані критерії*

Критерій	ТМ	ПТН	ПЕТН	ПДГ	НК
Обсяг обчислювальних ресурсів	1	1	0,5	0	0
Обсяг тренувальних даних	1	1	0,5	0,5	0
Час тренування	1	1	0,5	0	0
Контроль тренувальних даних	1	0,5	0,5	0	0
Підтримка великих документів	1	1	1	1	0
Підтримка нових мов	1	1	0	0	0
Обсяг обчислювальних ресурсів для оновлення знань	1	1	0,5	0	0
Рівень підтримки уваги до бізнес-правил	1	1	1	0,5	0

Оскільки наведені критерії потребують як мінімізації, так і максимізації, перетворимо їх таким чином, щоб більше значення мало більшу корисність відповідно до поставленого завдання. Перетворенню підлягають такі критерії: обсяг обчислювальних ресурсів, обсяг тренувальних даних, час тренування, обсяг обчислювальних ресурсів для оновлення знань. Значення оновлених критеріїв подані в табл. 3.

Таблиця 3. *Критерії максимізації корисності*

Критерій	ТМ	ПТН	ПЕТН	ПДГ	НК
Обсяг обчислювальних ресурсів	0	0	0,5	1	1
Обсяг тренувальних даних	0	0	0,5	0,5	1
Час тренування	0	0	0,5	1	1
Контроль тренувальних даних	1	0,5	0,5	0	0
Підтримка великих документів	1	1	1	1	0
Підтримка нових мов	1	1	0	0	0
Обсяг обчислювальних ресурсів для оновлення знань	0	0	0,5	1	1
Рівень підтримки уваги до бізнес-правил	1	1	1	0,5	0

Визначимо вагові коефіцієнти для кожного критерію. Було проведено експертне дослідження серед фахівців в Україні та за кордоном, які спеціалізуються на великих мовних моделях. За допомогою опитування було визначено такі вагові коефіцієнти:

– обсяг обчислювальних ресурсів: 0,15;

– обсяг тренувальних даних: 0,15;

– час тренування: 0,15;

– контроль тренувальних даних: 0,08;

– підтримка великих документів: 0,15;

– підтримка нових мов: 0,07;

– обсяг обчислювальних ресурсів для оновлення знань: 0,1;

– рівень підтримки уваги до бізнес-правил: 0,15.

Результати досліджень

Наведемо результати виконання завдання з оптимізації за допомогою адитивного згортання з ваговими коефіцієнтами та визначимо максимально корисний метод тренування вузькоспрямованих моделей (табл. 4).

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Таблиця 4. *Корисність методів тренування вузькоспрямованих мовних моделей*

Метод	Корисність
ТМ	0,45
ПТН	0,41
ПЕТН	0,615
ПДГ	**0,7**
НК	0,55

З огляду на досягнуті результати аналізу можемо зробити висновок, що метод пошуково-доповненої генерації є найбільш оптимальним за заданих умов. Методи ТМ та ПТН, що потребують тренування всіх параметрів моделі, виявилися менш корисними через значний обсяг необхідних ресурсів для їх налаштування. Метод ПЕТН є другим за корисністю та використовує значно менше ресурсів для налаштування, якщо порівнювати з ТМ та ПТН, проте все ж таки потребує певного тренування параметрів. Метод НК є менш корисним за ПЕТН, але для виконання деяких завдань простота його використання та швидкість налаштування можуть бути вагомими факторами. Хоча ПДГ має певні недоліки, зокрема залежність від можливостей базової моделі та ефективності алгоритму пошуку релевантної інформації, вони компенсуються відсутністю необхідності тренування моделі та простотою актуалізації знань.

Висновки

У процесі дослідження проаналізовано підходи до розроблення вузькоспрямованих великих мовних моделей, виявлено їх переваги, недоліки та обмеження, визначено найбільш оптимальний метод створення незалежних спеціалізованих систем, що дають змогу генерувати договори різними мовами в різних правових системах.

Аналіз попередніх досліджень та ринку виявив, що більшість моделей у відкритому доступі, натренованих для роботи в юридичній галузі, мають архітектуру кодувальника, яка не є ефективною для завдань генерації тексту. А наявні моделі з архітектурою декодувальника є або закритими, або подані моделями загального призначення, що потребують додаткової адаптації в обраній галузі.

Щоб порівняти підходи між собою, було сформовано набір критеріїв і наведено значення для кожного з методів налаштування моделей. Для визначення найбільш оптимального та корисного підходу впроваджено метод лінійного адитивного згортання з ваговими коефіцієнтами.

Унаслідок аналітичного експерименту виявлено, що метод пошуково-доповненої генерації є найбільш оптимальним за заданих умов, хоча програє більш складним підходам у гнучкості налаштування. Значно менший обсяг тренувальних ресурсів і наявний набір можливостей дають змогу ефективно адаптувати цей підхід для вузькоспрямованих галузей. Водночас метод параметро-ефективного тонкого налаштування за наявності додаткових часових і обчислювальних ресурсів може бути так само ефективним для адаптації в юридичній галузі.

Додатково визначено, що більшість спеціалізованих наборів метрик зосереджуються на аналітичних завданнях класифікації, резюмування та розуміння тексту, однак майже не приділяють уваги оцінюванню якості генерації нового контенту.

Перспективи подальшого розвитку

У подальших дослідженнях плануємо приділити увагу тренуванню моделей на основі деяких розглянутих методів і порівняти їх ефективність для виконання завдань аналізу, розуміння та генерації договорів у юридичній галузі. Також плануємо дослідити можливості поєднання декількох методів у межах однієї системи таким чином, щоб підсилити переваги кожного підходу та позбавитися недоліків для підвищення ефективності системи у виконанні поставлених завдань.

Зважаючи на те, що наявні набори метрик не мають повноцінних можливостей для оцінювання ефективності генерації юридичних документів, цей напрям також потребує подальших досліджень і розвитку.

Подяка

Автор висловлює подяку Збройним силам України за можливість написати повноцінну роботу під час повномасштабного вторгнення Російської Федерації на територію України. Також дякує науковому керівникові О. С. Назарову за підтримку та допомогу під час написання роботи.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Список літератури

1. Generative AI for Legal Contracts. Nasdaq. URL: https://www.nasdaq.com/articles/generative-ai-for-legal-contracts (дата звернення: 27.05.2024).
2. Vaswani A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. Attention is all you need. Advances in neural information processing systems. *31st Conference on Neural Information Processing Systems*. 2017. 30. DOI: https://doi.org/10.48550/arXiv.1706.03762
3. Devlin J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. 2018. DOI: https://doi.org/10.48550/arXiv.1810.04805
4. Touvron H. та ін. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*. 2023. DOI: https://doi.org/10.48550/arXiv.2307.09288
5. Jiang A. Q. та ін. Mixtral of experts. *arXiv preprint arXiv:2401.04088*. 2024. DOI: https://doi.org/10.48550/arXiv.2401.04088
6. Wu S. та ін. BloombergGPT: A large language model for finance. *arXiv preprint arXiv:2303.17564*. 2023. DOI: https://doi.org/10.48550/arXiv.2303.17564
7. Singhal K. та ін. Towards expert-level medical question answering with large language models. *arXiv preprint arXiv:2305.09617*. 2023. DOI: https://doi.org/10.48550/arXiv.2305.09617
8. Brown T. та ін. Language models are few-shot learners. *Advances in neural information processing systems*. 2020. № 33. P. 1877–1901. DOI: https://doi.org/10.48550/arXiv.2005.14165
9. Nori H. та ін. Can generalist foundation models outcompete special-purpose tuning? Case study in medicine. *arXiv preprint arXiv:2311.16452*. 2023. DOI: https://doi.org/10.48550/arXiv.2311.16452
10. Niklaus J., та ін. Multilegalpile: A 689gb multilingual legal corpus. *arXiv preprint arXiv:2306.02069*. 2023. DOI: https://doi.org/10.48550/arXiv.2306.02069
11. Guha N. Legalbench: A collaboratively built benchmark for measuring legal reasoning in large language models. *Advances in Neural Information Processing Systems*. 2024. № 36. DOI: https://doi.org/10.48550/arXiv.2308.11462
12. Hendrycks D. та ін. Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*. 2020. DOI: https://doi.org/10.48550/arXiv.2009.03300
13. Wang A., та ін. Superglue: A stickier benchmark for general-purpose language understanding systems. *Advances in neural information processing systems*. 2019. № 32. DOI: https://doi.org/10.48550/arXiv.1905.00537
14. Chalkidis I., та ін. LexGLUE: A Benchmark Dataset for Legal Language Understanding in English. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. 2022. С. 4310–4330. DOI: https://aclanthology.org/2022.acl-long.297
15. Niklaus J., та ін. LEXTREME: A Multi-Lingual and Multi-Task Benchmark for the Legal Domain. *Findings of the Association for Computational Linguistics: EMNLP 2023*. 2023. С. 3016–3054. DOI: https://aclanthology.org/2023.findings-emnlp.200
16. Mabey R. Unveiling our legal AI Assistant. Juro. URL: https://juro.com/blog/legal-ai-assistant (дата звернення: 10.03.2024).
17. Browne R. An AI just negotiated a contract for the first time ever and no human was involved. CNBC. URL: https://www.cnbc.com/2023/11/07/ai-negotiates-legal-contract-without-humans-involved-for-first-time.html (дата звернення: 10.03.2024).
18. Ian G. та ін. Generative adversarial nets. *Advances in neural information processing systems*. 2014. № 27. DOI: https://doi.org/10.48550/arXiv.1406.2661
19. Lewis P. та ін. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*. 2020. № 33. P. 9459–9474. DOI: https://doi.org/10.48550/arXiv.2005.11401
20. Touvron H., та ін. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*. 2023. DOI: https://doi.org/10.48550/arXiv.2302.13971
21. Vanian J., Leswing K. ChatGPT and generative AI are booming, but the costs can be extraordinary. CNBC. URL: https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html (дата звернення: 27.05.2024).
22. Thoppilan R. та ін. Lamda: Language models for dialog applications. *arXiv preprint arXiv:2201.08239*. 2022. DOI: https://doi.org/10.48550/arXiv.2201.08239
23. Hoffmann J. та ін. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*. 2022. DOI: https://doi.org/10.48550/arXiv.2203.15556
24. Mesnard T. та ін. Gemma: Open models based on gemini research and technology. *arXiv preprint arXiv:2403.08295*. 2024. DOI: https://doi.org/10.48550/arXiv.2403.08295

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

25. Microsoft Copilot for Sales. Microsoft. URL: https://www.microsoft.com/en-us/ai/microsoft-sales-copilot (дата звернення: 27.05.2024).
26. Luminance's Legal Pre-Trained Transformer. Luminance. URL: https://www.luminance.com/technology.html (дата звернення: 27.05.2024).
27. Lv K. та ін. Full parameter fine-tuning for large language models with limited resources. *arXiv preprint arXiv:2306.09782*. 2023. DOI: https://doi.org/10.48550/arXiv.2306.09782
28. Xu L. та ін. Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment. *arXiv preprint arXiv:2312.12148*. 2023. DOI: https://doi.org/10.48550/arXiv.2312.12148
29. Hu Edward J., та ін. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*. 2021. DOI: https://arxiv.org/abs/2106.09685
30. Karimi Mahabadi, R., Henderson, J., Ruder, S. Compacter: Efficient low-rank hypercomplex adapter layers. *Advances in Neural Information Processing Systems*. 2021. № 34. P. 1022–1035. DOI: https://doi.org/10.48550/arXiv.2106.04647
31. Wang Y., та ін. AdaMix: Mixture-of-Adaptations for parameter-efficient model tuning. *arXiv preprint arXiv:2205.12410*. 2022. DOI: https://doi.org/10.48550/arXiv.2205.12410
32. Karpukhin V., та ін. Dense passage retrieval for open-domain question answering. *arXiv preprint arXiv:2004.04906*. 2020. DOI: https://doi.org/10.48550/arXiv.2004.04906
33. Gao Y. та ін. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*. 2023. DOI: https://doi.org/10.48550/arXiv.2312.10997
34. Zhang T., та ін. BERTScore: Evaluating Text Generation with BERT. *International Conference on Learning Representations*. 2020. DOI: https://doi.org/10.48550/arXiv.1904.09675
35. GPT-4o. OpenAI. URL: https://platform.openai.com/docs/models/gpt-4o (дата звернення: 27.05.2024).

References

1. "Generative AI for Legal Contracts", available at: https://www.nasdaq.com/articles/generative-ai-for-legal-contracts (last accessed 27.05.2024).
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017), "Attention is all you need", *Advances in neural information processing systems*, № 30. DOI: https://doi.org/10.48550/arXiv.1706.03762
3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K. (2018), "Bert: Pre-training of deep bidirectional transformers for language understanding", *arXiv preprint arXiv:1810.04805*. DOI: https://doi.org/10.48550/arXiv.1810.04805
4. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D. (2023), "Llama 2: Open foundation and fine-tuned chat models", *arXiv preprint arXiv:2307.09288*. DOI: https://doi.org/10.48550/arXiv.2307.09288
5. Jiang, A.Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D.S., Casas, D.D.L., Hanna, E.B., Bressand, F., Lengyel, G. (2024), "Mixtral of experts", *arXiv preprint arXiv:2401.04088.* DOI: https://doi.org/10.48550/arXiv.2401.04088
6. Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., Mann, G. (2023), "Bloomberggpt: A large language model for finance", *arXiv preprint arXiv:2303.17564*. DOI: https://doi.org/10.48550/arXiv.2303.17564
7. Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Hou, L., Clark, K., Pfohl, S., Cole-Lewis, H., Neal, D., Schaekermann, M. (2023), "Towards expert-level medical question answering with large language models", *arXiv preprint arXiv:2305.09617*. DOI: https://doi.org/10.48550/arXiv.2305.09617
8. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S. (2020), "Language models are few-shot learners", *Advances in neural information processing systems*, № 33, P. 1877–1901. DOI: https://doi.org/10.48550/arXiv.2005.14165
9. Nori, H., Lee, Y.T., Zhang, S., Carignan, D., Edgar, R., Fusi, N., King, N., Larson, J., Li, Y., Liu, W., Luo, R. (2023), "Can generalist foundation models outcompete special-purpose tuning? case study in medicine", *arXiv preprint arXiv:2311.16452*. DOI: https://doi.org/10.48550/arXiv.2311.16452
10. Niklaus, J., Matoshi, V., Stürmer, M., Chalkidis, I., Ho, D.E. (2023), "Multilegalpile: A 689gb multilingual legal corpus", *arXiv preprint arXiv:2306.02069*. DOI: https://doi.org/10.48550/arXiv.2306.02069
11. Guha, N., Nyarko, J., Ho, D., Ré, C., Chilton, A., Chohlas-Wood, A., Peters, A., Waldon, B., Rockmore, D., Zambrano, D., Talisman, D. (2024), "Legalbench: A collaboratively built benchmark for measuring legal reasoning in large language models", *Advances in Neural Information Processing Systems*, № 36. DOI: https://doi.org/10.48550/arXiv.1904.09675

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

12. Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., Steinhardt, J. (2020), "Measuring massive multitask language understanding", *arXiv preprint arXiv:2009.03300*. DOI: https://doi.org/10.48550/arXiv.2009.03300
13. Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O., Bowman, S. (2019), "Superglue: A stickier benchmark for general-purpose language understanding systems", *Advances in neural information processing systems*, № 32. DOI: https://doi.org/10.48550/arXiv.1905.00537
14. Chalkidis, I., Jana, A., Hartung, D., Bommarito, M., Androutsopoulos, I., Katz, D., Aletras N. (2022), "LexGLUE: A Benchmark Dataset for Legal Language Understanding in English", *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, P. 4310–4330. DOI: https://aclanthology.org/2022.acl-long.297
15. Niklaus, J., Matoshi, V., Rani, P., Galassi, A., Stürmer, M., Chalkidis I. (2023), "LEXTREME: A Multi-Lingual and Multi-Task Benchmark for the Legal Domain", *Findings of the Association for Computational Linguistics: EMNLP 2023*, P. 3016–3054. DOI: https://aclanthology.org/2023.findings-emnlp.200
16. Mabey, R. "Unveiling our legal AI Assistant", available at: https://juro.com/blog/legal-ai-assistant (last accessed 27.05.2024).
17. Browne, R. "An AI just negotiated a contract for the first time ever – and no human was involved", available at: https://www.cnbc.com/2023/11/07/ai-negotiates-legal-contract-without-humans-involved-for-first-time.html (last accessed 27.05.2024).
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017), "Attention is all you need", *Advances in neural information processing systems*, № 30. DOI: https://doi.org/10.48550/arXiv.1706.03762
19. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T., Riedel, S. (2020), "Retrieval-augmented generation for knowledge-intensive nlp tasks", *Advances in Neural Information Processing Systems*, № 33, P. 9459–9474. DOI: https://doi.org/10.48550/arXiv.2005.11401
20. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A. (2023), "Llama: Open and efficient foundation language models", *arXiv preprint arXiv:2302.13971*. DOI: https://doi.org/10.48550/arXiv.2302.13971
21. Vanian, J., Leswing, K. "ChatGPT and generative AI are booming, but the costs can be extraordinary", available at: https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html (last accessed 27.05.2024).
22. Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.T., Jin, A., Bos, T., Baker, L., Du, Y., Li, Y. (2022), "Lamda: Language models for dialog applications", *arXiv preprint arXiv:2201.08239*. DOI: https://doi.org/10.48550/arXiv.2201.08239
23. Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D.D.L., Hendricks, L.A., Welbl, J., Clark, A., Hennigan, T. (2022), "Training compute-optimal large language models", *arXiv preprint arXiv:2203.15556*. DOI: https://doi.org/10.48550/arXiv.2203.15556
24. Team, G., Mesnard, T., Hardin, C., Dadashi, R., Bhupatiraju, S., Pathak, S., Sifre, L., Rivière, M., Kale, M.S., Love, J., Tafti, P. (2024), "Gemma: Open models based on gemini research and technology", *arXiv preprint arXiv:2403.08295*. DOI: https://doi.org/10.48550/arXiv.2307.09288
25. "Microsoft Copilot for Sales", available at: https://www.microsoft.com/en-us/ai/microsoft-sales-copilot (last accessed 27.05.2024).
26. "Luminance's Legal Pre-Trained Transformer, available at: https://www.luminance.com/technology.html (last accessed 27.05.2024).
27. Lv, K., Yang, Y., Liu, T., Gao, Q., Guo, Q., Qiu, X. (2023), "Full parameter fine-tuning for large language models with limited resources", *arXiv preprint arXiv:2306.09782*. DOI: https://doi.org/10.48550/arXiv.2306.09782
28. Xu, L., Xie, H., Qin, S.Z.J., Tao, X., Wang, F.L. (2023), "Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment", *arXiv preprint arXiv:2312.12148*. DOI: https://doi.org/10.48550/arXiv.2312.12148
29. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W. (2021), "Lora: Low-rank adaptation of large language models", *arXiv preprint arXiv:2106.09685*. DOI: https://arxiv.org/abs/2106.09685
30. Karimi Mahabadi, R., Henderson, J., Ruder, S. (2021), "Compacter: Efficient low-rank hypercomplex adapter layers", *Advances in Neural Information Processing Systems*, № 34, P. 1022–1035. DOI: https://doi.org/10.48550/arXiv.2106.04647
31. Wang, Y., Agarwal, S., Mukherjee, S., Liu, X., Gao, J., Awadallah, A.H., Gao, J. (2022), "AdaMix: Mixture-of-Adaptations for parameter-efficient model tuning", *arXiv preprint arXiv:2205.12410*. DOI: https://doi.org/10.48550/arXiv.2205.12410
32. Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., Yih, W.T. (2020), "Dense passage retrieval for open-domain question answering", *arXiv preprint arXiv:2004.04906*. DOI: https://doi.org/10.48550/arXiv.2004.04906
33. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H. (2023), "Retrieval-augmented generation for large language models: A survey", *arXiv preprint arXiv:2312.10997*. DOI: https://doi.org/10.48550/arXiv.2312.10997

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

34. Zhang, T., Kishore, V., Wu, F., Weinberger, K., Artzi Y. (2020), "BERTScore: Evaluating Text Generation with BERT", *International Conference on Learning Representations*. DOI: https://doi.org/10.48550/arXiv.1904.09675
35. "GPT-4o", available at: https://platform.openai.com/docs/models/gpt-4o (last accessed 27.05.2024).

Надійшла (Received) 29.05.2024

Відомості про авторів / About the Authors

Волоховський Віталій Євгенович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії, Харків, Україна; e-mail: vitalii.volokhovskyi@nure.ua; ORCID ID: https://orcid.org/0009-0006-5682-1889

Volokhovskyi Vitalii – Kharkiv National University of Radio Electronics, PhD student at the Department of Software Engineering, Kharkiv, Ukraine.

ANALYSIS OF METHODS FOR TRAINING DOMAIN-SPECIFIC LANGUAGE MODELS IN THE AREA OF LEGAL CONTRACTS GENERATION

The **subject** of the research is machine learning models and methods for generating legal contracts with limited resources and performance evaluation benchmarks. The **goal** of the work is to analyse approaches of domain-specific Large Language Models development and to find the optimal method of creating independent specialized systems that can generate contracts in different languages and legal systems. The article addresses the following **tasks**: identification of existing companies and solutions in this area, exploring approaches to create texts in natural language, analysis of evaluation and comparison methods of such systems, inspecting limitations and shortcomings of existing solutions and approaches, finding the optimal method of developing systems with limited resources. The following **results** were obtained: approaches of natural language generation and their features were investigated; the "Transformer" architecture was defined as a modern standard in the field of text information generation; different model types which are based on this architecture were considered; data sources for training were analysed; methods of adapting models in specialized areas were considered; model evaluating benchmarks for various tasks were reviewed; shortcomings of the existing specialized language models and the incompleteness of existing benchmarks for contract generation task evaluation were revealed. As a result of the analytical experiment, it was determined that the Retrieval-Augmented Generation method is the most optimal for solving the given task under the given conditions. The conducted experiment and its results can be used as a basis for further research of domain-specific language models development with limited resources. **Conclusions**: the article provides an overview of natural language generation methods using modern machine learning techniques, considers their advantages and disadvantages for small companies and scientific institutions that have limited resources. The work examinates a specialized legal domain and the problem of contract generation and determines the most optimal method to solve it.

Keywords: large language model; natural language generation; contract; legal document.

Бібліографічні описи / Bibliographic descriptions

Волоховський В. Є. Аналіз методів тренування вузькоспрямованих мовних моделей у сфері генерації договорів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 48–64. DOI: https://doi.org/10.30837/2522-9818.2024.2.048

Volokhovskyi, V. (2024), "Analysis of methods for training domain-specific language models in the area of legal contracts generation", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 48–64. DOI: https://doi.org/10.30837/2522-9818.2024.2.048

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 004.94
DOI: https://doi.org/10.30837/2522-9818.2024.2.065

Д. Гольдінер

ЗАСТОСУВАННЯ МОВИ ПРОГРАМУВАННЯ *GO* ДЛЯ МОДЕЛЮВАННЯ ПРОЦЕСІВ МАСОВОГО ОБСЛУГОВУВАННЯ

Предметом дослідження статті є методи та підходи до програмного моделювання систем масового обслуговування на прикладі багатоканальної системи з обмеженою чергою та відмовами в разі її переповнення. **Мета роботи** – обґрунтування доцільності застосування сучасних комп'ютерних інформаційних технологій, а саме мови програмування *Go* для моделювання систем масового обслуговування. У статті вирішуються такі **завдання**: формулювання досліджуваної системи масового обслуговування; визначення компонентів, критеріїв спорідненості систем масового обслуговування з їх програмними моделями; загальний огляд конкаренсі як математичної моделі; опис підходів та інструментарію мови програмування *Go*. Упроваджуються такі **методи**: мова програмування *Go* та її інструментарій, конкаренсі, паралельне виконання. **Досягнуті результати**: сформульовано досліджуване завдання масового обслуговування; визначено критерії порівняння компонентів систем масового обслуговування з інструментарієм мови програмування *Go*; проаналізовано доцільність використання мови програмування *Go* для моделювання систем масового обслуговування; подальшого розвитку набув інструментарій для комп'ютерного моделювання систем масового обслуговування; запропоновано застосування підходів конкаренсі, їх імплементації в мові програмування *Go* до моделювання систем масового обслуговування. **Висновки**. Мова програмування *Go* є дуже вдалою технологією для моделювання систем масового обслуговування. Її філософія, спосіб роботи, а також вбудований інструментарій має широкі можливості для моделювання різноманітних систем масового обслуговування. Використання зазначеної мови є доцільним і дозволяє наблизити поведінку програми до модельованого процесу, спростити імплементацію та зменшити час, необхідний на оброблення даних. Визначено значні перспективи щодо подальшого впровадження програмного продукту, реалізованого мовою *Go* у сфері моделювання процесів масового обслуговування.

Ключові слова: комп'ютерне моделювання; система масового обслуговування; мова програмування *Go*; конкаренсі.

Вступ

Останнім часом кількість навантажених систем, що обслуговують потоки заявок, які описуються випадковими законами, неперервно зростає. Така тенденція спостерігалася завжди, але тривалий час постійне вдосконалення одноядерних процесорів завдяки збільшенню кількості транзисторів перекривало потреби. Однак після певного періоду збільшувати продуктивність окремих ядер процесорів стало вкрай проблематично – закон Мура перестав діяти. Логічним рішенням стала ідея паралельного виконання завдань на окремих ядрах процесора та збільшення загальної продуктивності пристрою таким чином. Унаслідок цього виробники почали звертати увагу конструкторів на збільшенні енергоефективності ядер, а також на одночасному розміщенні більшої їх кількості на одному чипі. Одноядерні процесори були майже повністю витіснені з ринку багатоядерними аналогами.

Утім, сам по собі багатоядерний процесор не має жодних переваг, оскільки для ефективного використання його потенціалу необхідна підтримка багатопотоковості тим програмним забезпеченням, що будуть на ньому виконуватись, разом з операційною системою. І під час спроби запустити на такому процесорі програму, що використовує лише один потік, спостерігатимемо такі самі показники швидкодії, як і в ситуації з одноядерним процесором.

З появою багатоядерних процесорів обсяги запитів почали зростати із значно більшою швидкістю, ніж можливості з масштабування ресурсів, що використовуються в обробленні даних. Водночас поновлюється запит на оптимізацію програмного продукту для більш ефективного застосування наявних ресурсів.

Одним із розділів математики, що описує поведінку процесів з оброблення значної кількості запитів із подальшим паралельним опрацюванням, є теорія масового обслуговування. Її активний розвиток припав на 50–70 рр. XX ст., що було пов'язано з індустріалізацією та розвитком телефонії. Завдання, зумовлені потребами виробництва та комунікації, спонукали до створення математичного апарату, що дасть змогу оптимізувати процеси. Можна стверджувати, що з поширенням технології

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Інтернет експоненційним зростанням кількості мережних запитів та обсягів даних, які постійно потребують оброблення, теорія масового обслуговування може набути нового розвитку. Вона буде застосована до подолання чергових викликів, сформованих потребами XXI ст.

Аналіз наявних підходів до розв'язання завдання

Система масового обслуговування (СМО) – є складником теорії ймовірностей та математичної статистики, що використовуються для аналізу процесів, в яких у випадкові моменти часу надходять вимоги для подальшого їх опрацювання наявними каналами обслуговування. Математичне моделювання систем, які мають велику інтенсивність надходження вимог, що є непропорційною продуктивності їх оброблення, дає змогу оцінити потенціальну пропускну здатність, а також її збільшити. Застосування стохастичного аналізу дозволяє аналізувати час очікування заявок і визначати оптимальну кількість каналів оброблення для задоволення потреб [1].

Основні елементи СМО:

– вхідний потік вимог – це послідовна сукупність вимог (заявок) на надання певної послуги, що надходить до системи. Вхідний потік може мати різні види ймовірнісного розподілу залежно від потреб та особливостей процесу, що моделюється. Зазвичай параметризується через інтенсивність прибуття λ заявок, що впливає на завантаженість системи, час очікування та ймовірність відмови в обслуговуванні;

– черга – це механізм для тимчасового зберігання заявок, які очікують на обслуговування. Черга може мати обмежену або необмежену ємність. Структура та політика черги щодо новоприбулих заявок відіграють важливу роль у визначенні порядку опрацювання заявок. Це значно впливає на ймовірність відмови в разі надходження в переповнену чергу, а також на розподілення навантаження поміж каналами;

– канали обслуговування – технічні пристрої або люди, які забезпечують обслуговування вимог. Кожен канал може обслуговувати не більше ніж одну заявку за раз. Час, необхідний для обслуговування, частіше буває випадковим із певним розподілом імовірності. Кількість каналів обслуговування є обмеженою та визначає її пропускну здатність. Має вплив на час перебування вимог у системі та на ймовірність відмови;

– вихідний потік вимог – це потік вимог, що залишають систему, отримавши чи не отримавши замовлену послугу. Результатом оброблення вимоги є один із ключових критеріїв оцінювання ефективності СМО [2].

Важливим чинником у програмному моделюванні систем масового обслуговування є те, якою мірою механізми мови програмування відповідають критеріям математичного апарату. Одна з основних розбіжностей між мовами програмування, що впливає на доцільність використання для моделювання СМО, полягає в підході до паралельного виконання. Будемо розглядати тільки популярні мови програмування, адже важливу роль відіграє можливість розвивати та підтримувати програмний продукт у довгостроковій перспективі. Відповідно до рейтингу, побудованого способом аналізу вакансій на роль програмного інженера за 2023 р., у табл. 1 наведено десять мов, придатних для написання серверів, що мають долю понад 1%.

Таблиця 1. *Класифікація популярних мов програмування*

Назва	Доля	Підхід до багатопотоковості
JavaScript / TypeScript	29.8%	асинхронність
Python	19.64%	багатопотоковість
Java	17.78%	багатопотоковість + бібліотека для підтримки конкаренсі
C#	12.21%	багатопотоковість
PHP	9.39%	асинхронність
C / C++	9.14%	багатопотоковість + бібліотека для підтримки конкаренсі
Ruby	4.37%	асинхронність
Go	1.91%	SCP, конкаренсі на рівні ядра мови

Серед наведених технологій можна виокремити кілька груп:

– підтримка асинхронності в межах одного потоку операційної системи;

– підтримка роботи з потоками операційної системи напряму;

– підтримка конкаренсі за допомогою розширень, доданих нещодавно як стороння бібліотека;

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

– повна підтримка конкаренсі та інтеграція ідей *CSP* на рівні ядра мови.

Мови програмування, що донедавна використовуються для написання програм, зокрема й для моделювання систем масового обслуговування, не мали підтримки *CSP* та конкаренсі. Зараз упроваджується часткова підтримка, яка не надає повноти функціоналу [3]. Така ситуація дає змогу застосувати нові підходи до програмного моделювання процесів із використанням сучасних мов програмування.

Мета роботи

Метою статті є обґрунтування доцільності впровадження сучасних комп'ютерних інформаційних технологій, а саме – підходу *CSP* із використанням мови програмування *Go* для моделювання систем масового обслуговування. Також передбачено проаналізувати спорідненість інструментарію паралельного виконання зазначеної технології до математичних узагальнень. Це надалі може відкрити шлях до значного вдосконалення інструментарію та покращення ефективності програмних моделей.

Постановка завдання

Розглядатимемо класичну задачу про багатоканальну систему масового обслуговування з обмеженою чергою вимог та відмовою в разі переповнення черги [1], де n – кількість однакових каналів обслуговування, до яких надходить пуассонівський потік заявок інтенсивності λ. Якщо на момент надходження нової заявки є хоча б один вільний канал, він негайно починає оброблення. Якщо всі канали зайняті – вимога стає останньою до загальної черги ємності k. Заявки покидають чергу для подальшого обслуговування в тій самій послідовності, у якій вони надходили на очікування. Канал, що звільнюється від виконання, відразу починає оброблення першої в черзі вимоги. У цьому разі кожна заявка обслуговується тільки одним каналом і кожен канал може обслуговувати не більше ніж одна вимога одночасно. Якщо вільними є декілька каналів обслуговування, для оброблення буде обрано канал випадковим чином. Час, необхідний на оброблення однієї вимоги, є випадковою величиною з експоненційним законом розподілу ймовірності:

$$F(x) = 1 - e^{-\nu x}, \text{ де } \nu > 0. \qquad (1)$$

Причинами такого рішення є відносна простота самого процесу, наявність потужного математичного апарату аналітичного дослідження [4], а також значна поширеність процесів, що підпадають під цю модель. Отже, задана система підпадає під умовне визначення: $M/M/n/m$, де

– перша M вказує на марковський вхідний потік вимог;

– друга M вказує на те, що процес оброблення вимог також є марковським;

– n визначає, що система є багатоканальною та задає кількість каналів;

– m описує обмежену ємність системи та визначає кількість місць для очікування $m > 0$.

Вхідний потік вимог має задовольняти такі вимоги:

– стаціонарність потоку;

– відсутність післядії;

– ординарність.

Імовірність надходження нової вимоги до системи за проміжок часу t може бути визначений таким чином:

$$P_i(t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t}, \qquad (2)$$

де λ – це інтенсивність прибуття вимог, що надходять до системи за одиницю часу; i – кількість вимог, наявних у системі, разом із наступною в момент t.

Зважаючи на обмеження щодо кількості вимог, які можуть очікувати на обслуговування в черзі, обчислимо ймовірність відмови у виконанні новій вимозі через переповнення черги. Оскільки майбутнє обслуговування не залежить у контексті теорії ймовірностей від того, що відбувалося до моменту часу t_0 у зв'язку з особливостями ймовірнісного розподілу. Для описаної системи ймовірність відмови дорівнює ймовірності розташування в системі рівно i вимог. Маємо такий вираз [1]:

$$\begin{cases} 1 \le i \le n & P_i = \dfrac{\rho^i n^i}{i!} P_0 \\ n < i \ge n + m & P_i = \dfrac{\rho^i n^n}{n!} P_0 \end{cases}, \qquad (3)$$

$$P_0 = \left[\sum_{i=0}^{n-1} \frac{\rho^i n^i}{i!} + \frac{n^n}{n!} \cdot \frac{\rho^n \left(1 - \rho^{m+1}\right)}{1 - \rho} \right]^{-1}, \qquad (4)$$

де $\rho = \dfrac{\lambda}{n\nu}$ – середня продуктивність системи. Нас найбільш цікавить ситуація, коли $\rho > n$, оскільки

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

саме в цьому разі можемо бачити наповнення черги й відмови через її переповнення [1]. Надалі досліджуватимемо методи зменшення ймовірності відмови черговій вимозі. Отже, ймовірність відмови з причини переповнення черги може бути обчислена за допомогою виразу

$$P_{vidm} = P_{n+m} = \frac{\rho^{n+m} n^n}{n!} P_0, \quad (5)$$

$$P_{vidm} = \frac{\rho^{n+m} n^n}{n!} \cdot \left[\sum_{i=0}^{n-1} \frac{\rho^i n^i}{i!} + \frac{n^n}{n!} \cdot \frac{\rho^n \left(1 - \rho^{m+1}\right)}{1-\rho} \right]^{-1}. \quad (6)$$

Наведений вираз означає, що запит отримає відмову щодо обслуговування, якщо всі лінії та місця очікування будуть зайняті.

Розв'язання проблеми

Конкаренсі в математиці та інформатиці визначається як властивість систем, у яких кілька обчислень виконуються одночасно та потенційно взаємодіють один з одним [5]. Термін, запропонований Ентоні Хоаром у межах його роботи з *Communicating Sequential Processes* (*CSP*), став основоположним для розуміння комплексних систем, де багато процесів відбуваються одночасно й асинхронно. *CSP* моделює поведінку системи за допомогою алгебри процесів, даючи змогу описати взаємодію між паралельними процесами через події комунікації.

Розвиток теорії *CSP* сприяв упровадженню поняття "конкаренсі" у високорівневих мовах програмування та архітектурах систем. Поняття паралелізму в *CSP* та його математична модель дозволяють проєктувати та аналізувати складні системи, що містять паралельні процеси та взаємодію за допомогою обміну повідомленнями. Теорія конкаренсі застосовується в багатьох галузях разом із розробленням операційних систем, розподіленими обчисленнями та проєктуванням мікросхем. Вона допомагає забезпечити взаємну незалежність, синхронізацію та уникнення взаємоблокувань між паралельними процесами. Окреслений аспект математики та комп'ютерних наук постійно розвивається, оскільки нові парадигми та методології надходять для кращого розуміння й ефективнішого використання паралелізму в обчислювальних системах [6].

Розглянемо застосування підходів конкаренсі до розв'язання поставленого завдання. За термінологією *CSP* багатоканальна система масового обслуговування з обмеженою чергою та відмовами може бути подана у вигляді мережі процесів, що комунікують між собою. Кожен етап, або стан, у якому перебуватиме заявка під час оброблення, є процесом. Перехід між станами відбуватиметься під впливом відповідних подій або сигналів. У цьому разі, згідно із визначенням конкаренсі, усі ці процеси можуть відбуватися асинхронно для кількох заявок водночас. Для початку виокремимо незалежні процеси, що супроводжують оброблення заявки в СМО:

– надходження заявок може мати різні розподілення ймовірностей та інтенсивність;

– утримання в черзі може мати додатковий функціонал визначення пріоритетів;

– оброблення каналом обслуговування може мати випадкову тривалість, а також імовірність помилки;

– покидання заявкою системи, що не залежить від результату обслуговування.

Формалізуємо взаємодію між зазначеними процесами через обмін повідомленнями за допомогою подій, які описуватимуть переходи між станами заявки:

– надходження (*arrive*): подія, що означає надходження чергової заявки до системи та є початковою подією;

– взяття до черги (*onqueue*): за умови, якщо в черзі є вільні місця, заявка успішно стає на очікування;

– покидання черги (*dequeuer*): у разі звільнення хоча б одного каналу обслуговування заявка, що стоїть наступною в черзі, може покидати чергу, звільнюючи місце для подальших надходжень;

– початок оброблення (*start*): канал розпочинає обслуговування заявки;

– успішне завершення оброблення (*success*): опрацювання заявки пройшло в штатному режимі;

– помилка під час оброблення (*fail*): з певних причин канал не зміг надати очікувану послугу;

– відмова (*reject*): у черзі на момент надходження не було вільних місць;

– базова подія завершення оброблення для кінцевих алгоритмів (*STOP*).

Користуючись визначеною множиною подій, сформулюємо алфавіт процесу:

$$aSMO = \{arrive, onqueue, dequeue, start, success, fail, reject\}. \quad (7)$$

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Тоді процес оброблення заявки матиме такий вигляд [7]:

$$SMO = arrive \to \left(reject \to STOP \middle| onqueue \to dequeue \to start \to \left(success \to STOP \middle| fail \to STOP \right) \right). \quad (8)$$

Отже, бачимо, що системи масового обслуговування, які розглядаються в межах дослідження, можуть бути цілісно описані за допомогою *CSP* та конкаренсі. Крім цього, завдяки абстракціям, а також комунікації через події, ми маємо змогу гнучко масштабувати та розширювати взаємодію компонентів без необхідності змінювати самі процеси. Цей процес може бути запущено паралельно, що не змінить його загальну будову.

Go – це сучасна компільована мова програмування із жорсткою системою типів, що була створена корпорацією *Google* 2012 р. для задоволення нагальних потреб і як відповідь на виклики у сфері комп'ютерної інженерії, що постали перед розробниками програмного продукту в XXI ст., а саме:

– неперервне надходження значної кількості одночасних запитів із подальшим паралельним їх обробленням;

– необхідність підтримки великої різноманітності процесорів, платформ та операційних систем;

– потреба швидко реалізовувати бізнес-ідеї та раніше отримувати відгук про відповідність очікувань;

– бажання спростити написання, розширення, підтримку, а також читання коду програм;

– ідея децентралізованої спільноти розробників і популяризації пакетів із відкритим кодом.

Одним із ключових етапів у формуванні мови *Go* є реалізація ідеї конкаренсі [8]. За цим терміном ховається кілька особливостей рантайму, а саме:

– розбиття складного процесу на ланцюг незначних послідовних завдань;

– взаємодія між процесами відбувається з допомогою сигналів;

– заблоковані завдання стають на очікування, та виконання переходить до інших процесів у черзі;

– здатність легко збільшити кількість обробників для простих завдань.

Кожна із згаданих вище властивостей позначається в теорії масового обслуговування. Тож розглянемо всі особливості підходу та їх вплив на моделювання систем масового обслуговування.

Більшість систем у комп'ютерній інженерії або є дуже складними із самого початку, або стають такими з часом. Їх спрощення є непростим завданням, яке потребує значних зусиль, що пізніше – то більше. Одним з ефективних способів запобігання ускладненню систем є декомпозиція складних завдань на низку простих, маленьких операцій, що взаємодіють між собою. Водночас таке розбиття дає змогу більш ефективно будувати математичні моделі систем, оскільки, замість одного надскладного завдання, будемо моделювати незначні, значно простіші процеси.

Спосіб взаємодії між процесами є дуже важливим, адже самі по собі вони не ефективні. Існує декілька способів комунікації:

– спільний доступ до пам'яті (м'ютекси);

– неблокувальні алгоритми;

– обмін повідомленнями крізь канали.

З-поміж зазначених підходів саме обмін повідомленнями за допомогою каналів найкраще вписується в контекст систем масового обслуговування. Оскільки в цьому підході, на відміну від інших, наголошується не на доступі до даних, а на взаємодії між компонентами [9]. Як наслідок, від проблеми оброблення даних переходимо до питання моделювання взаємодії та визначення станів компонентів системи, що добре описується математичним апаратом.

Хоча *Go* є молодою мовою та має запозичення ідей з інших, старших мов програмування, вона має і свої унікальні властивості, що роблять програми, написані на *Go* ефективнішими та відмінними за характером від програм, написаних спорідненими мовами. Це єдина сучасна технологія, що реалізує принципи конкаренсі, описані в роботах Ч. Е. Р. Хоара "CSP" на рівні ядра [10]. Тому прямий переклад з таких мов, як, наприклад, *Java* або *C++*, малоймовірно дасть очікуваний результат. Відповідно, для ефективного використання мови програмування *Go* важливо розуміти її особливості, ідіоми, а також практики, що дозволяють розкрити потенціал. У *Go* чітко сформульовані стандарти щодо форматування, найменування, будови програми, які дають змогу уніфікувати рішення, а також зробити їх більш зрозумілими для інших розробників і стійкими до змін.

Також ця технологія перебуває в активній фазі розроблення та вдосконалення, і до неї регулярно надходять оновлення, надаючи додаткові можливості та переваги.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Ядром програми є її рантайм – набір процесів, що відповідають за оброблення та виконання алгоритму програми, а також забезпечення допоміжних операцій, що відбуваються за замовчуванням на рівні мови [11]. Рантайм у мові *Go* має кілька особливостей, що виділяють її на фоні інших, а саме:

– чітка типізація з нещодавно доданою підтримкою параметризованих типів;

– високий рівень абстракції взаємодії з операційною системою;

– збірник сміття, що працює за методологією *mark & sweep*;

– власний менеджер потоків із вродженою підтримкою конкаренсі.

Ці особливості будуть описані більш детально надалі.

Інструментарій мови програмування *Go* дозволяє дуже гнучко відтворювати різноманітні навантажені системи, багато в чому завдяки спорідненості ідей, закладених у філософію мови, з теорією масового обслуговування [12]. Розглянемо, яким чином можна програмно змоделювати процес масового обслуговування, описаний раніше, за допомогою мови *Go*. Будемо рухатися послідовно, відповідно до життєвого циклу заявки в системі.

Спочатку визначимося з тим, як відтворити вхідний потік вимог. Для цього можемо скористатися функцією, що буде викликатися в циклі. Її відповідальність полягатиме в надсиланні вхідних параметрів далі в чергу для подальшого оброблення. Ця функція буде єдиним публічним контрактом нашої програми, і саме з її допомогою відбуватиметься подання вхідних даних. У користувача програми не буде взаємодії з подальшими етапами.

Наступним кроком буде передача новоствореної вимоги до черги очікування. Для стандартизації логіки виконання ми відмовляємось від прямої комунікації з каналами оброблення вимог і завжди передаватимемо їм роботу через чергу. Такий крок дасть змогу задовольнити вимоги підходу конкаренсі. Це зі свого боку підвищить стійкість програми до помилок, а також спростить її будову й подальшу підтримку. Роль черги в мові програмування *Go* відіграватиме канал з буфером, що дорівнює розміру черги. У разі спроби записати значення до каналу із заповненим буфером рантайм блокуватиме виконання функції. Отже, за необхідності моделювати систему з відмовами скористаємося можливістю додати дію за замовчуванням і повертатимемо помилку переповненої черги, що буде відповідати відмові у виконанні вимоги. Сам по собі буфер каналу в мові програмування *Go* має політику оброблення елементів *FIFO*. Це означає, що вимоги обслуговуватимуться в тій самій послідовності, у якій вони надходили. У разі успішного розташування вимоги в черзі вона має дочекатися, коли всі вимоги, що надійшли раніше, будуть розібрані. Якщо на момент готовності до виконання в наявності понад одна вільна горутина, то вимога буде взята однією з них випадковим чином.

Роль каналу обслуговування в мові *Go* відіграватиме горутина – паралельний процес, що виконує вказану функцію. Вхідні дані для виконання операції горутина вичитує з каналу. Оброблення вхідної інформації відбувається штатним чином і може привести як до досягнення успішного результату, так і спричинити помилку. Коли оброблення вимоги завершене, цикл горутини повертається до спроби читати з каналу та блокується до надходження чергового завдання.

У цьому разі розмір каналу, кількість горутин, функція, що виконуватиме розрахунок, а також структура даних для вимоги задаються на етапі компіляції й не можуть бути змінені в процесі роботи програми. Організувати необхідний режим синхронізації між горутинами та каналами дозволяє директива *select*, роль якої полягає в оркестрації. Час виконання вимоги залежить від функції, заданої для виконання в горутинах, й інженер має змогу визначити її таким чином, щоб задовольнити вимогу до експоненційного випадкового розподілу.

Отже, можемо розглянути базову реалізацію програмного забезпечення для моделювання СМО з обмеженою чергою, без відмов, написану мовою *Go*. Першою компонентою буде функція з назвою *feed*, завданням якої є лише подання заявок до системи масового обслуговування (рис. 1).

```
func feed(inCh chan<- inputData) {  1 usage
    for i := 0; i < requestsNumber; i++ {
        inCh <- inputData{id: i, duration: time.Millisecond * time.Duration(i%3)}
    }
    close(inCh)
}
```

Рис. 1. Функція подання заявок

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Коли всі запити надіслано до черги, закриваємо канал заявок, сигналізуючи іншим компонентам про цю подію. Другою операцією розберемо функцію *service*, що описує поведінку каналу обслуговування та виконує примітивну логіку: повернути помилку, якщо значення *id* рівняється по модулю 3, а інакше – повернути успіх (рис. 2). Функція *newSMO* запускає канали обслуговування та відповідальна за закриття каналу відповідей після завершення оброблення даних (рис. 3). Усі раніше згадані операції будуть викликані у функції *main* у момент запуску програми (рис. 4), також у ній читаємо результати оброблення заявок з каналу й виводимо їх у командний рядок, продовжуючи ці дії до закриття каналу.

```
func service(inCh <-chan inputData, outCh chan<- outputData, wg *sync.WaitGroup) {  1 usage
    for in := range inCh {
        if in.duration == 0 {
            outCh <- outputData{id: in.id, err: errors.New( text: "service duration can not be 0")}
            continue
        }

        time.Sleep(in.duration)
        outCh <- outputData{id: in.id}
    }

    wg.Done()
}
```

Рис. 2. Функція обслуговування

```
func newSMO(numberOfServices int, inCh <-chan inputData, outCh chan<- outputData) {  1 usage
    wg := sync.WaitGroup{}
    wg.Add(numberOfServices)
    for i := 0; i < numberOfServices; i++ {
        go service(inCh, outCh, &wg)
    }

    wg.Wait()
    close(outCh)
}
```

Рис. 3. Функція ініціалізації СМО

```
func main() {
    inCh := make(chan inputData, numberOfServices)
    outCh := make(chan outputData, queueSize)

    go feed(inCh)
    go newSMO(numberOfServices, inCh, outCh)

    for res := range outCh {
        if res.err != nil {
            log.Printf( format: "Service for the inpput id=%d failed because: %v\n", res.id, res.err)
            continue
        }
        log.Printf( format: "Service for the inpput id=%d successfuly finshed", res.id)
    }
}
```

Рис. 4. Основна функція виклику та оброблення результатів

Результати дослідження

Мова програмування *Go* перезавантажує підхід до паралельного виконання програм унаслідок інтеграції методології конкаренсі у філософію та інструментарій мови. Вона є першопрохідницею, оскільки саме в *Go* вперше було реалізовано концепцію *CSP* на рівні базових механізмів.

Горутини дуже добре відтворюють поведінку незалежних каналів оброблення вимог, оскільки так само не поділяють між собою стан і не впливають напряму на взаємне виконання вимог. За потреби

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

можна вказати кількість горутин як будь-яке ціле додатне число. Однак важливо не забувати, що навіть дуже легкі потоки можуть стати проблемою за умови відсутності контролю за їх виконанням та вчасним коректним завершенням. Відповідно, необхідно наслідувати певні практики для зменшення ймовірності неочікуваної поведінки, а також застосовувати ефективні методики для виявлення, пошуку джерела та усунення помилок, пов'язаних із конкаренсі в програмі [13].

Канал у *Go* повністю відповідає ролі черги, він надає послідовність очікування вимог, синхронізацію доступу, а також випадковість підбору горутини для виконання. Якщо потрібно, можна прибрати буфер у каналу, що призведе до трансформації на систему з відмовами та без черги. У разі необхідності змоделювати систему без відмов у інструментарії розробника є слайс, що є масивом із здатністю автоматично збільшувати свою ємність. Отже, він може відігравати роль умовно необмеженої черги.

Функції, що виконуватимуть розрахунки, є дуже гнучким механізмом і дозволяють задовольнити будь-які вимоги до часу розв'язання задачі залежно від потреб експерименту.

Однією з проблем, що постає під час дизайну мови програмування, є задача про багатопотоковість. Оскільки логіка виконуватиметься на потоках операційної системи, популярним підходом є пряме поєднання програмних потоків із потоками ОС. Серед переваг такого рішення – його відносна простота. Однак програми стають залежними від того, яким чином операційна система виділяє їм процесорний час. У мові *Go* розробники взяли за основу ідею про легкі, маленькі програмні потоки, які не мають прямого виходу на рівень операційної системи [14]. Значною особливістю мови програмування *Go* є її скедулер (планувальник) – механізм аркестрації горутин. Система планування багатопотоковості має такі компоненти:

– потік ядра процесора (*Core*);

– потік операційної системи (*M*), що взаємодіє з потоком процесора;

– програмний потік (*P*) – абстракція рівня рантайму *Go*, що зазвичай відповідає одному потоку операційної системи;

– горутина (*G*) – маленький автономний процес, що описує бізнес-логіку;

– допоміжні процеси рантайму – операції, що є горутинами самі по собі й забезпечують життєдіяльність програми;

– скедулер (планувальник) – алгоритм, що обирає, коли та якій горутині дати змогу виконання, а також який тред (потік) операційної системи її обслуговуватиме.

Опишемо роботу механізму взаємодії компонентів планувальника (рис. 5).

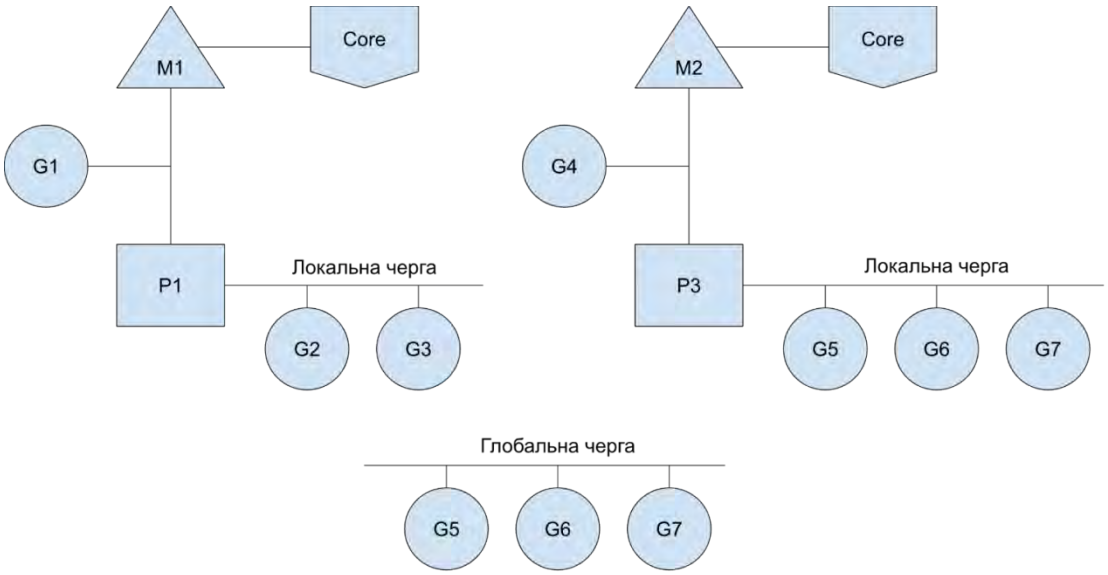


Рис. 5. Компонентна схема планувальника

1. У процесі використання в програмному коді директиви *Go* із подальшим визначенням, яка функція має виконуватись, рантайм створює маленьку горутину з власним стеком пам'яті мінімального розміру.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

2. Планувальник визначає локальну чергу програмного потоку, до якої потрапить горутина на очікування.

3. Коли потік буде готовий взяти на виконання чергову горутину, він бере її в роботу.

4. Є перелік умов, за яких виконання горутини може бути тимчасово призупинене з міркувань надання часу іншим горутинам. У цьому разі горутина, що була попередньо в обробленні, "паркується" до наступного вікна можливостей.

5. У разі закінчення очікувальних горутин у локальній черзі програмного потоку він виконає спробу забору половини черги горутин у іншого, більш завантаженого програмного потоку.

6. Раз за певний час або якщо всі локальні черги пусті, забір відбувається з глобальної черги.

Основна ідея такого підходу – надання додаткового рівня абстракції для щільнішого планування роботи потоків операційної системи. Цей механізм забезпечує дію підходу конкаренсі, навіть за наявності лише одного ядра процесора [15].

Найважливішою перевагою в сукупності факторів є загальна простота взаємодії елементів, масштабування та потенціал до паралельного виконання на багатоядерних процесорах [16]. В аналогічний спосіб можна розширювати можливості [17].

Висновки

Мова програмування *Go* – дуже вдала технологія для моделювання систем масового обслуговування. Її філософія, спосіб роботи, а також вбудований інструментарій має широкі можливості для моделювання різноманітних систем масового обслуговування. Методологія конкаренсі, що була взята за основу ядра *Go*, є основною відмінністю від інших мов програмування. Запропонована та описана в межах наукової праці Ч. Е. Р. Хоара "CSP", вона описує новий підхід до побудови абстракцій навколо взаємодії асинхронних процесів. Це дуже вдало розширює інструментарій інженера механізмами взаємодії складників системи. У цьому разі виконання програмної реалізації моделі значною мірою наближається до досліджуваного процесу. Черговою перевагою застосування *Go*, порівняно з іншими поширеними мовами програмування, є її підвищена швидкодія. Важливу особливість також становить простота у використанні інструментарію та подальша підтримка готового продукту, що особливо помітно в роботі з багатопотоковістю.

Отже, використання мови програмування *Go* є доцільним і дає змогу наблизити поведінку програми до модельованого процесу, спростити імплементацію та зменшити час, необхідний для оброблення даних.

Наукова новизна

Набув подальшого розвитку інструментарій для комп’ютерного моделювання систем масового обслуговування. Запропоновано застосування підходів конкаренсі, їх імплементації в мові програмування *Go* до моделювання систем масового обслуговування. Це матиме такі наслідки:

– підвищить ефективність розрахунків;

– спростить програмну реалізацію;

– наблизить програмну реалізацію до модельованого процесу.

Перспективи

Окреслений напрям розвитку методів програмного моделювання систем масового обслуговування має значний потенціал. Надалі необхідно спроєктувати та реалізувати програмний продукт мовою *Go*, що даватиме змогу моделювати різноманітні процеси. На базі такого програмного рішення буде можливо перевіряти на практиці різноманітні гіпотези щодо оптимізації в системах масового обслуговування. Ключовою вимогою для цього завдання буде універсальність і забезпечення можливостей щодо моделювання широкого спектра систем масового обслуговування.

Список літератури

1. Литвинов А.Л. Теорія систем масового обслуговування. Харків: ХНУМГ ім. О.М. Бекетова, 2018. 141 с.
2. Borovkov A. Stochastic Processes in Queueing Theory. Translate by K.Wickwire. *New York: Springer-Verlag*, 1976. 280 p. DOI: https://doi.org/10.1007/978-1-4612-9866-3

3. Chabbi M., Ramanathan M.K. A study of real-world data races in Golang. *PLDI 2022: Proceedings of the 43rd ACM SIGPLAN International Conference on Programming Language Design and Implementation,* San Diego, 13–17 June 2022 / Special Inter. Group on Program. Lang. SIGPLAN. New York, 2022. P. 474–489. DOI: https://doi.org/10.1145/3519939.3523720
4. Zeifman A.I. Quasi-Markov Processes and Description of Some Models of Queueing Theory. *IFAC Proceedings Volumes.* 1986. Vol. 19. No. 5. P. 445–449. DOI: https://doi.org/10.1016/S1474-6670(17)59840-7
5. Hoare A. Communicating sequential processes. New Jersey: Prentice Hall, 1985. 235 p. DOI: https://doi.org/10.1145/359576.359585
6. Yunfei Liu EMsFEM based concurrent topology optimization method for hierarchical structure with multiple substructures/ Yunfei Liu et al. *Computer Methods in Applied Mechanics and Engineering.* 2024. Vol. 418. Part A. 116549. P. 1–26. DOI: https://doi.org/10.1016/j.cma.2023.116549
7. C. A. R. Hoare. Communicating sequential processes. *Communications of the ACM*. 1978. Vol. 21. No 8. P. 666–677. DOI: https://doi.org/10.1145/359576.359585
8. Sufyan bin U. Mastering GoLang: A Beginner's Guide. Boca Raton: CRC Press, 2022. 298 p. DOI: https://doi.org/10.1201/9781003310457
9. Fava D., Steffen M. Ready, set, Go!: Data-race detection and the Go language. *Science of Computer Programming.* 2020. Vol. 195. 102473. P.1–23. DOI: https://doi.org/10.1016/j.scico.2020.102473
10. Sottile M., Mattson T., Rasmussen C. Introduction to Concurrency in Programming Languages. New York: Chapman and Hall/CRC, 2009. 344 p. DOI: https://doi.org/10.1201/b17174
11. Sufyan bin U. GoLang: The Ultimate Guide. Boca Raton: CRC Press, 2022. 366 p. DOI: https://doi.org/10.1201/9781003309055
12. Bowman H., Gomez R. Concurrency Theory: Calculi an Automata for Modelling Untimed and Timed Concurrent Systems. London: Springer-Verlag, 2006. 422 p. DOI: https://doi.org/10.1007/1-84628-336-1
13. Zhang D., Qi P., Zhang Y. GoDetector: Detecting Concurrent Bug in Go. *IEEE Access.* 2021. Vol. 9. P. 136302–136312. DOI: https://doi.org/10.1109/ACCESS.2021.3116027
14. Donovan A. A. A., Kernighan B. W. The Go programming language. New York: Addison-Wesley Professional, 2015. 400 p.
15. Pontelli E., Gupta G. On the duality between or-parallelism and and-parallelism in logic programming. *1st International EURO-PAR Conference on Parallel Processing:* EURO-PAR 199. Stockholm, 29–31 aug. 1995. Lecture Notes in Computer Science. Vol 966. Springer, Berlin. 2005. P. 43–54. DOI: https://doi.org/10.1007/BFb0020454
16. Komendantskaya E., Schmidt M., Heras J. Exploiting Parallelism in Coalgebraic Logic Programming. *Electronic Notes in Theoretical Computer Science.* 2014. Vol. 303. P. 121–148. DOI: https://doi.org/10.1016/j.entcs.2014.02.007
17. Whitney J., Gifford C., Pantoja M. Distributed execution of communicating sequential process-style concurrency: Golang case study. *The Journal of Supercomputing.* 2019. Vol. 75. No 3. P. 1396–1409. DOI: https://doi.org/10.1007/s11227-018-2649-2

References

1. Lytvynov, A. (2018), *Queuing Theory System, [Teoriia system masovoho obsluhovuvannia]*, KhNUMH im. O.M. Beketova, Kharkiv, 141 p.
2. Borovkov, A. (1976), *Stochastic Processes in Queueing Theory. Translate by K.Wickwire*, Springer-Verlag, New York, 280 p. DOI: https://doi.org/10.1007/978-1-4612-9866-3
3. Chabbi, M., Ramanathan, M. (2022), "A study of real-world data races in Golang", *PLDI 2022: Proceedings of the 43rd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, 13–17 June 2022, San Diego. Special Inter. Group on Program. Lang. SIGPLAN, New York, P. 474–489. DOI: https://doi.org/10.1145/3519939.3523720
4. Zeifman, A. (1986), "Quasi-Markov Processes and Description of Some Models of Queueing Theory", *IFAC Proceedings Volumes*, Vol. 19, No. 5, P. 445–449. DOI: https://doi.org/10.1016/S1474-6670(17)59840-7
5. Hoare, A. (1985), *Communicating sequential processes*, Prentice Hall, New Jersey. 235 p. DOI: https://doi.org/10.1145/359576.359585
6. Liu, Y. (2024), "EMsFEM based concurrent topology optimization method for hierarchical structure with multiple substructures"/ Liu, Y. et al., *Computer Methods in Applied Mechanics and Engineering,* Vol. 418, Part A, 116549, P. 1–26. DOI: https://doi.org/10.1016/j.cma.2023.116549
7. Hoare, A. (1978), "Communicating sequential processes", *Communications of the ACM*, Vol. 21, No 8, P. 666–677. DOI: https://doi.org/10.1145/359576.359585
8. Sufyan, bin U. (2022), *Mastering GoLang: A Beginner's Guide*, CRC Press, Boca Raton, 298 p. DOI: https://doi.org/10.1201/9781003310457
9. Fava, D., Steffen, M. (2020), "Ready, set, Go!: Data-race detection and the Go language", *Science of Computer Programming*, Vol. 195, 102473, P.1–23. DOI: https://doi.org/10.1016/j.scico.2020.102473
10. Sottile, M., Mattson, T., Rasmussen, C. (2009), *Introduction to Concurrency in Programming Languages*, Chapman and Hall/CRC, New York, 344 p. DOI: https://doi.org/10.1201/b17174
11. Sufyan, bin U. (2022), *GoLang: The Ultimate Guide*, CRC Press, Boca Raton, 366 p. DOI: https://doi.org/10.1201/9781003309055

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

12. Bowman, H., Gomez, R. (2006), *Concurrency Theory: Calculi an Automata for Modelling Untimed and Timed Concurrent Systems*, Springer-Verlag, London, 422 p. DOI: https://doi.org/10.1007/1-84628-336-1
13. Zhang, D., Qi, P., Zhang, Y. (2021), "GoDetector: Detecting Concurrent Bug in Go", *IEEE Access*, Vol. 9, P. 136302–136312. DOI: https://doi.org/10.1109/ACCESS.2021.3116027
14. Donovan, A., Kernighan, B. (2015), *The Go programming language*, Addison-Wesley Professional, New York, 400 p.
15. Pontelli, E., Gupta, G. (2005), "On the duality between or-parallelism and and-parallelism in logic programming", *1st International EURO-PAR Conference on Parallel Processing: EURO-PAR 199*, Stockholm, 29–31 aug. 1995. Lecture Notes in Computer Science, vol 966, Springer, Berlin, P. 43-54. DOI: https://doi.org/10.1007/BFb0020454
16. Komendantskaya, E., Schmidt, M., Heras, J. (2014), "Exploiting Parallelism in Coalgebraic Logic Programming", *Electronic Notes in Theoretical Computer Science,* Vol. 303, P. 121–148. DOI: https://doi.org/10.1016/j.entcs.2014.02.007
17. Whitney, J., Gifford, C., Pantoja, M. (2019), "Distributed execution of communicating sequential process-style concurrency: Golang case study", *The Journal of Supercomputing*, Vol. 75, No 3, P. 1396–1409. DOI: https://doi.org/10.1007/s11227-018-2649-2

Надійшла (Received) 09.05.2024

Відомості про авторів / About the Authors

Гольдінер Денис Ігорович – Харківський національний університет радіоелектроніки, аспірант, Харків, Україна; e-mail: denys.holdiner@nure.ua; ORCID ID: https://orcid.org/0000-0002-1456-1867

Goldiner Denys – Kharkiv National University of Radio Electronics, PhD Student, Kharkiv, Ukraine.

APPLICATION OF GO PROGRAMMING LANGUAGE FOR SIMULATION OF MASS SERVICE PROCESSES

The **subject matter** of the article is methods and approaches to software modeling of System of Mass Services on the example of a multi-channel system with a limited queue and failures in case of its overflow. The **goal** of the work is to justify the feasibility of using modern computer information technologies, namely the Go programming language for modeling System of Mass Services. The following **tasks** were solved in the article: formulation of the researched System of Mass Services; determination of components, criteria of kinship of System of Mass Services with their software models; general overview of concurrency as a mathematical model; a description of the approaches and tools of the Go programming language. The following **methods** are used: Go programming language and its tools, concurrency, parallel execution. The following **results** were obtained: the researched task of mass service was formulated; criteria for comparing the components of System of Mass Services with the Go programming language toolkit were formed; an analysis of the feasibility of using the Go programming language for modeling System of Mass Services was carried out; received further development of tools for computer simulation of System of Mass Services; the application of concurrency approaches, their implementation in the Go programming language, to the modeling of System of Mass Services is proposed. **Conclusions**: The Go programming language is a very successful technology for modeling System of Mass Services. Its philosophy, way of working, as well as the built-in toolset provide ample opportunities for modeling various System of Mass Services. The use of this language is appropriate and allows to bring the behavior of the program closer to the simulated process, simplify implementation, and reduce the time required for data processing. There are great prospects for the further implementation of a software product implemented in the Go language in the field of mass service process modeling.

Keywords: computer simulation; System of Mass Services; Go programming language; concurrency.

Бібліографічні описи / Bibliographic descriptions

Гольдінер Д. І. Застосування мови програмування *GO* для моделювання процесів масового обслуговування. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 65–75. DOI: https://doi.org/10.30837/2522-9818.2024.2.065

Goldiner, D. (2024), "Application of GO programming language for simulation of mass service processes", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 65–75. DOI: https://doi.org/10.30837/2522-9818.2024.2.065

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

UDC 004.8 DOI: https://doi.org/10.30837/2522-9818.2024.28.076

N. HULIIEV

CHOICE OF MACHINE LEARNING MODELS FOR PREDICTING THE DEVELOPMENT OF PSYCHOLOGICAL DISORDERS IN PEOPLE WITH HYPOTHIREOSIS AND HYPERTHIREOSIS

The **subject** of this article is endocrinological diseases, namely, the analysis of complications in people with hypothyroidism and hyperthyroidism. It is known that these diseases occur asymptomatically or in a way that may indicate other possible diseases, so people do not suspect what exactly they are suffering from. Later, the diseases develop to the point where complications occur in the body, some of the most dangerous of which are psychological disorders: depression, mania, aggression, etc. Therefore, the **aim** of this work is to develop methods for predicting the occurrence of neurological deterioration in people who have already been diagnosed with endocrinological diseases. The article solves the **problem** of choosing the best models for predicting the occurrence of psychological disorders in people with endocrinological problems. Machine learning methods that are widespread in the medical field were analyzed and one of them was chosen that more optimally solves all the tasks of the task. The selection of criteria took into account potential problems with medical and psychological data. The **method** used was linear additive convolution, which is used to select the best alternatives according to the results, with the Pareto principle, which aims to exclude less suitable alternatives because all the features have lower values than in other options. For the experiment, all features were converted into quantitative ones to calculate convolution values. The evaluation criteria are given in the paper. The following **results** were obtained: the forecasting model in further study of this problem will be a random forest. **Conclusions**: the forecasting methods were studied and a more optimal model was chosen using linear additive convolution, namely, the random forest algorithm, its advantages and disadvantages were considered. A more detailed analysis of its development will be presented in the following articles. A mathematical description of the chosen forecasting method is provided, which includes potential ways of implementation and steps for building an algorithm for one of these methods.

Keywords: hypothyroidism; hyperthyroidism; psychological disorders; forecasting; linear additive convolution; Pareto principle; random forest; decision tree; Gini index.

Introduction

Hypothyroidism and hyperthyroidism are among the most common endocrine diseases of the thyroid gland, the factors of which are

– environment
– bad habits;
– genetics;
– unhealthy diet;
– allergies;
– iodine deficiency;
– stress.

According to the World Health Organization, they rank second among endocrinopathies, and diabetes mellitus ranks first.

According to research, the total number of people who develop manifest hypothyroidism ranges from 3% to 8%, and if we add cases of subclinical hypothyroidism, it is 10% to 12%. The consequences of this disease increase over time, as it has many effects on various organs of the patient. Most often, the cardiovascular and nervous systems are affected. The disease negatively affects physical, sexual, cognitive, and intellectual functioning, and therefore may have atypical symptoms that force patients to see different doctors because of concerns about the heart, nerves, stomach, and reproductive function, without realizing that the problem is different, and local treatment will not help overcome a global disease.

A significant number of observations of endocrine disorders are devoted to the study of patients' psychological health, as it is known that deterioration can range from passive or increased exhaustion to unexpected aggressive and dangerous actions. And a long course of the disease can cause parts of a person's personality to be removed, such as states of affect and memory impairment.

All patients develop complications that lead to a deterioration in the psycho-emotional state, the nature of which may vary depending on the severity of the disease. M. Bleuler studied mental syndromes of a narrow circle, namely deviant behavior that occurs

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

due to endocrine diseases. His systematization is that the scientist combined them into one structure, which includes the following components:

– mood deterioration;
– lack of motivation;
– decrease in activity;
– change in instincts;
– change in urges [1].

Hyperthyroidism is a disease in which the thyroid gland produces excessive amounts of hormones, which can cause thyrotoxicosis. The disease occurs due to the following causes: toxic denoma, toxic multinodular goiter, and Graves' syndrome. Diagnosis of the disease involves the use of imaging methods, ultrasound, monitoring of iodine absorption, and biochemical tests. Thyroid dysfunction negatively affects the skin, reproductive, cardiovascular, immune, and gastric systems. Treatment options include antithyroid drugs, surgery, and radioactive iodine.

The disease can occur suddenly or develop in the body over time, go away on its own or in remission. The following symptoms are subtle: ventricular arrhythmias, tachycardia. Dangerous signs are disorders of the skin, eyes, musculoskeletal system, and neck. Other signs of the disease include insomnia, anxiety, irritability and depression, memory and attention impairment, delirium, and apathy. One of the less common complications is psychosis.

The relevance of the analysis of these two diseases is that their rapid spread among the population, asymptomatic or ambiguous development (i.e., the very signs that may indicate abnormal endocrine behavior of the body) make us believe that other problems have arisen. After all, due to the many forms they can take, it is very difficult to diagnose them in time, which causes further complications, namely psychological disorders. The patient's psycho-emotional state deteriorates, ranging from various syndromes to severe disorders. Therefore, after the diagnosis of hypothyroidism and hyperthyroidism in a patient, it is necessary to immediately analyze his or her mental health in order to prevent its noticeable deterioration or the occurrence of diseases associated with the nervous and mental systems of the body in a timely manner.

Analysis of recent research and publications

Machine learning is becoming a widespread means of prediction and diagnosis in medicine. There are many methods aimed at determining the probability of certain signs occurring given certain circumstances, which is the most common purpose of their use in this area. Therefore, the study of diseases using data mining methods is currently appropriate and relevant [2, 3].

When discussing diseases that are common in the world, endocrinological diseases are still the ones that affect humanity. It is known that hypothyroidism and hyperthyroidism, which are thyroid disorders that cause abnormal regulation of thyroid hormones, can develop asymptomatically or with signs that indicate completely different diseases, becoming factors in other complications due to late diagnosis [4, 5].

Some of these deteriorations are psychological disorders, such as depression, mania, aggression, etc. Neurological problems increase the damage to the entire body, which makes it impossible to treat a patient with hypothyroidism or hyperthyroidism normally [6]. Currently, many observations are aimed at analyzing the course of endocrinological diseases or at options for mitigating symptoms, so the problem of preventing the development of psychological disorders due to hypothyroidism and hyperthyroidism remains relevant. This paper focuses on selecting one of the most optimal machine learning methods for predicting whether neurological problems may occur due to endocrinological diseases [7].

Currently, a large number of models are used in the medical field to improve diagnostic and preventive treatment. Unfortunately, there are still cases when it is extremely difficult to determine what exactly is affected in the body, what are the causes of the disease, neoplasms, complications, and even the patient's condition. In these difficult tasks that doctors around the world have to solve, it has become advisable to use prognostic models in order to act in a timely manner to find the appropriate and correct treatment for a person as soon as possible, because it is easier to solve problems before they occur.

One of the cases of building predictive models is the development of a mathematical model to predict intrauterine infection among newborns, as the obstetrics and gynecology department warns of an increase in the number of intrauterine infections that disrupt pregnancy and increase the likelihood of prenatal death. A projective and retrospective study was conducted among women with viral, bacterial, and combined infections. The analysis included clinical and obstetric examinations, and the prediction was made on a two-point scale using multivariate discriminant analysis using the *Statistica* statistical environment based on 105 indicators that were

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

the result of a medical examination. As a result, the most likely and influential factors were identified by the factor structure matrix of the discriminant analysis protocol.

Pregnancy is known for many unpredictable problems, so this topic is common among scientists. Another study was aimed at creating a model for predicting the course of pregnancy depending on women's laboratory and instrumental parameters to reduce the likelihood of preterm birth or antenatal fetal death. Initially, all women underwent obstetric and gynecological and somatic examinations to collect materials for analysis. Means and standard deviations were calculated using standard methods, but if there was a significant discrepancy between the values, median and quartile values were also calculated. Pearson and Mann–Whitney's tests of agreement were used to calculate the reliability of sample differences. Full statistical analysis was performed using the *Statistica* 6.0 software package. The predictive model was developed using fuzzy logic based on the Takagi-Sugeno fuzzy framework. As a result, four models were built to predict the term of labor, which is not likely to be preterm and possible threats that, on the contrary, can cause it.

Hypothyroidism and hyperthyroidism are also being studied using machine learning. For example, representatives of Kharkiv National Medical University analyzed the course of primary hypothyroidism in Ukrainians who were forced to leave their homes due to the war, which became a factor in cognitive and anxiety disorders. Changes in the life of internally displaced persons are accompanied by a modification of social and psychological relations, which negatively affects the mental state of a person, exacerbating the development of depression. IDPs and persons permanently residing in the Kharkiv region with a diagnosis of primary hypothyroidism participated in the study. After clinical-neurological and clinical-psychopathological analyzes, the results were processed using a mathematical and statistical approach using *Statistica* 6.0 and Student's *t*-test. As a result, the average values with a possible arithmetic mean error were obtained, and the dependencies between the values were determined using correlation analysis. It was found that among IDPs there are more people with primary hypothyroidism manifested in depression of varying degrees than in the second group of people. Also, thyroid hormone deficiency caused anxiety in all, but in those who did not change their place of residence during the war, it was found to be the most common [8–10].

It is worth mentioning studies on the treatment of hypothyroidism. The Ivano-Frankivsk Medical University considered the question of the greatest effectiveness of alpha-lipoic acid Dialipon or the drug Vitaxon. The medical parameters of 42 middle-aged patients with primary hypothyroidism were studied: clinical signs, neuropsychiatric disorders, organ damage. Statistical processing was performed using the *Statistica* package (*StatSoft, Inc*.) and nonparametric methods of evaluating the results. Patients were divided into two groups: the first group included people who were to take Dialipone and Vitaxon as hormone therapy, and the second group was prescribed *L*-thyroxine-randomly. Some patients from both groups got better, but a noticeable improvement was observed among people taking Dialipon and Vitaxon. In addition, the liver also showed positive dynamics, because it is responsible for controlling the body's metabolism. Therefore, given that metabolic changes can provoke nervous system damage, it is necessary to take additional medications [11, 12].

An interesting study of endocrinological disease was conducted on 48 mice. During the observation, the effect of stress and physical activity on the thyroid gland in the setting of hypothyroidism was analyzed. We directly studied how chronic stress and physical training can change the morphology of the gland by means of microscopy and statistical analysis based on the Student's *t*-test method. As a result, it was found that the effects of stress and exercise did not change the number of iodine-containing hormones and thyroid TSH in hypothyroidism.

Endocrinopathies also have an impact on the dental health of patients. Studies show that hypothyroidism and hyperthyroidism cause pathological processes in the periodontium, caries, and non-carious formations. These processes are caused by the fact that thyroid diseases disrupt metabolism, which provokes enamel and dentin erosion, enamel necrosis, and tooth abrasion. In patients with thyroid dysfunction, there is a correlation between the prevalence of periodontal disease (generalized periodontitis) and the time of disease development and the activity of the process [13–15].

Identification of previously unsolved parts of the overall problem. Purpose of the work, tasks

Currently, observations of known endocrinopathies are devoted to the study of the course, possible treatments and features of further complications, but the problem

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

of preventing one of the most common complications – the development of psychological disorders among people with hypothyroidism and hyperthyroidism – still remains relevant. The purpose of this study is to select the most optimal or optimal methods for predicting the deterioration of patients' psychological health, which will process (process) medical and psychological indicators of patients. And the next step will be to analyze the proposed ways to avoid these difficulties against the background of endocrinological disease [7].

Methods and materials

The multi-criteria task of choosing the most optimal forecasting method in the medical field is to determine the best one among all the proposed ones. Two types of methods are used for this purpose. The first set of methods aims to remove the number of evaluation criteria by making assumptions in the process of ranking the values of characteristics, and the second removes possible options before the comparison begins.

The most effective method for this observation is still the method from the first set, which includes the convolution method, boundary criteria, distance, and the main criterion.

The convolution method summarizes all the criteria. Such methods are divided into additive, multiplicative, and maximin convolution.

Additive is calculated by the formula

$$K(x)=\sum_{j=1}^{n} a_j K_j(x), \qquad (1)$$

where $K(x)$ –general criterion for the alternative $x \in X$; $\left(K_1(x),\dots K_j(x),\dots K_n(x)\right)$ – a set of initial criteria; n – number of initial criteria; a_j – a normalization factor indicating the weight of the alternative.

The best of all possible alternatives to the problem is calculated using the following formula:

$$x^* = \arg\max_{x\in x} K(x)\ . \qquad (2)$$

That is, the result is the largest value obtained by the convolution method.

Multiplicative convolution is calculated using the formula

$$K(x)=\Pi_{j=1}^{n} K_j^{a_j}(x). \qquad (3)$$

The maximin convolution is calculated by the formula

$$K(x)=\max_i \min_j a_{ij} K_j(x). \qquad (4)$$

The best results for the multiplicative and maximin convolutions are calculated using formula (2).

The method of threshold criteria is used in design and planning problems in which the threshold values of the criteria take on the values $k_j(x) \ge k_{j^0}$; $j=1\dots,n$. The calculation formula for this method is as follows:

$$K(x)=\min_j\left(K_j(x)/K_{j0}(x)\right). \qquad (5)$$

The best result is selected by formula (2).

The distance method uses distance as an additional metric. For example, the following information is enough to select the ideal solution $\left(K_0,\dots,K_{0n}\right)$. Let's calculate the distance to the maximum value $d(x)$ for each alternative. Then the best alternative will be determined by the formula

$$x^* = \arg\min_{x\in x} d(x). \qquad (6).$$

In working with the methods of the first group, methods from the second group are used, namely, the Pareto principle, when the best option is selected from the list of alternatives remaining after this method has eliminated the others by comparing their characteristics and identifying the worst options because their values had lower indicators.

The principle of equilibrium, or Nash's principle, aims to reduce the number of alternatives and calculates which one is inferior in terms of characteristics to the others; it is closely related to the Pareto principle.

However, there are cases when uncontrollable parameters complicate the solution of multi-criteria problems, which can arise for various reasons. In such cases, it is advisable to use the guaranteed outcome method, which allows you to determine the worst case response and a likely high and guaranteed value.

By considering the methods from the two groups, you can choose the method that is most effective for the study. It is not advisable not to analyze all the options, so we will use the method from the first group and convolution, because it is difficult to determine the thresholds of the criteria.

The most common and simplest option is linear additive convolution, so we will use it as a method for determining the usefulness of models to select the best one.

The first step is to identify all the criteria that will be involved in the analysis of alternatives, and then calculate their weighting values to describe the priority in choosing the best option.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

The values of each criterion can be quantitative or qualitative. This method operates with the former, so if the alternatives have values of the latter type, it is necessary to convert them to quantitative values.

When the quantitative values are known, alternatives can be eliminated from the list using the Pareto principle if the values of all criteria for a particular alternative are lower than those of other options. Next, the indicators are normalized if they are in different ranges or measures of measurement, which can lead to inaccurate and incorrect results, so it is better to normalize the values of all criteria of all alternatives with a range from 0 to 1.

In the case of maximization, you can normalize by dividing the value of the criterion by the maximum, while in the case of minimization, one is divided by it.

The next step is to rank the criteria for calculating the weighting coefficients. We have n criteria, of which the best one will have a value of n divided by n, the least important one will have a value of $n-1$ divided by n, and so on. Another way is to divide one by the sum of all the criteria scores.

The last step is to calculate the convolution value based on the alternatives: calculate the sum of the products of the criteria values and their weighting factors [16].

We have investigated forecasting methods for selecting the best model for observation purposes. To predict the development of mental disorders in patients with hypothyroidism and hyperthyroidism, it is necessary to apply methods that take into account the medical and psychological indicators of patients. Further research is aimed at analyzing the requirements for this task and selecting the most optimal model among all those described in the table.

The task is to solve a multi-criteria problem, namely, to determine which machine learning method will better predict the possible development of psychological disorders in people with hypothyroidism and hyperthyroidism to identify future ways to prevent them.

First, let's define a set of alternatives – these are models that are more commonly used in the medical field, among which we will choose an effective one for the study.

Let's assume that we have

– linear regression;
– polynomial regression;
– logistic regression;
– decision trees;
– multilayer perceptron;
– *k-nearest* neighbors model;
– *random forest*;
– gradient boosting;
– Bayesian classification;
– ensemble of models;
– SVM.

Medical and psychological indicators of a patient are determined by their unstable nature. They can have abnormal, incomplete, empty, nonlinear values in a rather significant amount, because psychological indicators will be accurately taken by questionnaires, interviews, and non-verbal tests. Therefore, the best option should not neglect the peculiarities of these indicators. This means that the model must process a large amount of input information, which may be nonlinear, given that the model must be able to handle missing indicators and respond to noise. Therefore, the selection criteria were as follows:

– model complexity;
– type of training;
– ability to process nonlinear information;
– whether the error is taken into account;
– tendency to overlearn;
– working with large amounts of information;
– working with missing information;
– work with noise.

Create and fill in a table with all the alternatives and the criteria that describe them (see Table 1).

Next, you need to convert the value of the criteria into a numerical value. Let's consider each of them.

The complexity of the model lies in the training method, the complexity and number of algorithms in one forecasting method, and the number of layers in the case of neural networks. Therefore, the values "simple", "medium", and "complex" are given accordingly.

The type of training presented in the table means "with a teacher" or "without a teacher". It is easier to create a model that does not require time to learn and search for information, but if learning is based on previous information, the result may be more likely, so if the value is "without a teacher", it is 1 point, and if it is the other way around, it is 2 points.

The criterion "ability to process nonlinear information" indicates whether the model is able to work with indicators that are scattered nonlinearly, as they may have unexpected values, which is likely to affect the overall result. Therefore, this feature should be taken into account in the study. If the model has this feature, the value is 1 point, if not, it is 0.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Table 1. *Experimental results*

Types	Characteristics							
	Simplicity of the model	**Type of training**	**Ability to process nonlinear information**	**Does they take into account an error?**	**Tendency to relearn**	**Working with large amounts of information**	**Working with missed information**	**Working with noise**
Linear regression	Simple	With a teacher	No	No	Less tend	Yes	Can not	Can not
Polynomial regression	Simple	With a teacher	Has	No	Tend to relearn	Works, but may be problems	Can not	Can not
Logistic regression	Simple	With a teacher	No	No	Less tend	Yes	Can not	Can not
Decision trees	Medium	With a teacher	Has	No	Tend to relearn	Works, but may be problems	Can	Can
Multilayer perceptron	Medium	With a teacher	Has	No	Tend to relearn	Yes	Can not	Can not
Model *k-nearest* neighbors	Simple	Without a teacher	Has	No	Tend to relearn	Small amount of information	Can not	Can not
Random Forest	Medium	With a teacher	Has	Yes	Tend to relearn	Yes	Can	Can
Gradient boosting	Complex	With a teacher	Has	Yes	Less tend	Works	Can not	Can
Bayesian classification	Simple	Without a teacher	Has	Yes	Less tend	Works	Can not	Can not
Ensemble of models	Complex	With a teacher	Has	Yes	Tend to relearn	Works	Can	Can
SVM	Medium	With a teacher	Has	No	Less tend	Works, but with limitations	Can not	Can not

Usually, a model cannot predict the exact percentage of probability, so it is advisable to use a method that takes into account the measurement error. If the model does, the value is 1 point, otherwise it is 0 points.

Each model may be prone to overfitting, which is possible under different conditions, or not at all, so if the method is likely to have this problem, which is the worst, the parameter value is 0, if it is likely to have it under special conditions, it is 1, and if the probability of this is low (which is the best), then it is 2 points.

One of the main characteristics is working with large amounts of information, so if this is natural for the model, it is 2 points, if there are some restrictions, then it is 1 point, otherwise - 0.

If the model can work with missing information or noise, then it gets 1 point, and otherwise 0 points.

Replace the values in the table with quantitative indicators. Also, at this stage, some alternatives can be eliminated according to the Pareto principle if they are inferior to other options by the criteria (see Table 2).

Table 2. *Modified table after changing the information with quantitative indicators and applying the Pareto principle*

Types	Characteristics							
	Simplicity of the model	**Type of training**	**Ability to process nonlinear information**	**Does they take into account an error?**	**Tendency to relearn**	**Working with large amounts of information**	**Working with missed information**	**Working with noise**
Linear regression	3	2	0	0	2	2	0	0
Logistic regression	3	2	0	0	2	2	0	0
Random Forest	2	2	1	1	1	2	1	1
Gradient boosting	1	2	1	1	2	2	0	1
Bayesian classification	3	1	1	1	2	2	0	0
SVM	2	2	1	0	2	1	0	0

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

The last step is actually to calculate the values of the linear additive convolution for each option, with the value of the normalization factor calculated first for each criterion (see Table 3).

Table 3. *Convolution results*

Model	Convolution value
Linear regression	0,75974026
Logistic regression	0,75974026
Random Forest	2,680735931
Gradient boosting	1,70021645
Bayesian classification	1,252164502

As we can see, according to the convolution results, the best model for this study is the random forest algorithm.

Research results and discussion

According to the results of the study, the best machine learning method to be used to predict the development of psychological disorders among people with hypothyroidism and hyperthyroidism is *Random Forest* [16].

Decision trees are known for their overfitting, which causes an increase in the variance of predictions. The *Random Forest* algorithm was developed to solve the above problem, allowing to build ensemble forecasts, but with a lower variance value, and it is similar to backpropagation. Random forest is a modified decision tree algorithm aimed at building not one but many trees, each of which produces a certain result, and the final one is the one that occurs most often. However, it differs in that it has a second level of randomness: in the process of optimizing node crushing, a random subset of features is analyzed for subsequent decoration of the estimators, and the random forest always determines the size of the bootstrapped data set, according to the size of the training sample [17].

Random forest significantly increases the accuracy and efficiency of forecasting and classification. The algorithm works as follows: first, during training, a tree based on random information is built, and in the process of dividing the nodes, a random subset of characteristics is selected and the result that occurs most often becomes the final one.

The advantages are:

– nonlinear information does not affect the efficiency of the algorithm;

– support for parallel processing;

– simplicity of application is that the only parameters of the method are the number of randomly selected features and the number of trees to be built on a randomly selected subset of the data sample;

– there is no need to reduce the tree;

– the algorithm estimates the importance of criteria and out-of-band accuracy in programs with large amounts of information, where estimates can be overestimated.

But like bagging, random forest is not defined by a lower bias. If a significant amount of information contains unequally distributed and mutually independent examples, overfitting is used – the process of selecting many identical decision trees by the random forest algorithm, each of which is overfitted, which is a known drawback.

In addition, it is prone to overfitting under certain conditions, namely, if there are too many trees in the forest, high correlation between them, small sample size, incorrect set of hyperparameters, and too complex data. The following methods are used to reduce the likelihood of overfitting:

– cross-validation;

– increasing the size of information;

– limiting the depth of trees;

– tree diversity.

Despite its drawbacks, the *Random Forest* algorithm is a more optimal option that processes a significant amount of information, is able to work with noise, and processes nonlinear, missing data, including measurement error. It has several implementation methods [18, 19].

The algorithm is used to evaluate the importance of the characteristics that need to be trained based on the average *out-of-bag* error for each subsample item. Next, before and after shuffling, it is necessary to determine the average value of the difference in *out-of-bag* errors on all trees, normalized by the standard deviation.

The main thing in building decision trees is the method of selecting the attribute by which the division will take place and the nodes will be built. There are the following methods:

– ID3 algorithm, which uses the Gini index or incremental method;

– C4.5 algorithm, which is a better version of ID3, which takes into account the normalized growth;

– CART algorithm;

– modifications of the CART algorithm – IndCART, DB-CART.

All trees are built in the following independent steps:

– generate a subset of size n from the training data set randomly;

– build a tree of m randomly selected features;

– continue the process without cutting off until the amount of data is complete.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

In general, the algorithm for constructing a decision tree is as follows: it is necessary to calculate the entropy of the input set s_0, if $s_0 = 0$, then:

1) all sample objects are of the same class;

2) store this class as a leaf of the tree.

If $s_0 != 0$, then:

1) determine the attribute that will divide the set in such a way as to reduce the average entropy value;

2) the found attribute becomes a node of the decision tree and is saved;

3) divide the sample into subsets depending on the values of the selected attribute;

4) recursively continue the process for each subset [20, 21].

Let's consider one of the tree building algorithms – the CART algorithm. In it, each node has two subnodes. At each step, the selected node attribute divides the set into two parts: the right part, in which the rule is executed, and the left part, in which the rule is not executed. To select the optimal rule, the partitioning quality evaluation function is applied. The evaluation function, which uses the CART algorithm, is based on the intuitive idea of reducing uncertainty in a node. This means a partitioning that will result in a node having as many examples of one class as possible and as few as possible of all other classes. This concept is close to entropy, but it uses a different measure of uncertainty, for which the term "dirty node" is appropriate. In the CART algorithm, the idea of a "dirty node" is formalized in the *Gini* index. If a data set T contains data from n classes, then the *Gini* index is defined as

$$Gini(T) = 1 - \sum_{i=1}^{n} p_i^2 , \qquad (7)$$

where the parameter p_i – is the probability of class i in T.

If the set T is split into two parts, T_1 and T_2 with the number of examples in each N_1 and N_2 respectively, then the quality index of the split is equal to

$$Gini_{split}(T) = \frac{N_1}{N} Gini(T_1) + \frac{N_2}{N} Gini(T_2). \qquad (8).$$

The best split is the one for which $Gini_{split}(T)$ is minimal. We denote the number of examples in a node as N, where L and R are the number of examples in the left and right descendants, respectively, l_i and r_i are the number of examples of the i-th class in the left/right descendant. Then the quality of the partitioning is estimated by the following formula:

$$Gini_{split}(T) = \frac{L}{N}\left(1 - \sum_{i=1}^{n}\left(\frac{l_i}{L}\right)^2\right) + \frac{R}{N}\left(1 - \sum_{i=1}^{n}\left(\frac{r_i}{R}\right)^2\right) \to \min. \quad (9)$$

The peculiarity of this index is the most optimal breakdown of data to build a better decision tree [22, 23].

Conclusions and prospects for further development

Modern medicine is increasingly using machine learning methods to diagnose and predict diseases, their course, and types of treatment. Such methods are becoming widespread in this field, as they increase the chances of a safer, more desirable, and accurate outcome. Specifically, endocrinopathies are known for being extremely difficult to detect in time, which leads to a significant number of unpredictable consequences.

The paper examined the psychological problems arising from hypothyroidism and hyperthyroidism, namely, analyzed machine learning methods that can calculate the likelihood of developing neurological complications in the setting of these diseases for timely action to eliminate the problem before it occurs.

An analysis of publications describing the ways in which machine learning methods are used has shown that the endocrinological issue is being studied, has prospects for research, and is accompanied by new theories, but these observations try to solve problems that have already arisen or find ways to alleviate the patient's condition without completely eliminating the problem. Therefore, the purpose of the article was to find analytical methods that would predict the likelihood of developing psychological disorders in order to avoid deterioration of patients diagnosed with hypothyroidism or hyperthyroidism.

The paper discusses the symptoms and consequences of hypothyroidism and hyperthyroidism, examples of the implementation of machine learning methods for prognostic purposes in the medical field, proposes the methods studied in this observation based on their common and distinctive characteristics, and analyzes their advantages and disadvantages.

Linear additive convolution was applied to select a more optimal model based on the requirements needed in the outlined task, according to the results of which the "random forest" algorithm is more effective [24, 25].

In the future, selected prediction methods based on medical and psychological indicators will be studied to predict the occurrence of psychological problems due to hypothyroidism and hyperthyroidism in order to introduce certain preventive measures aimed at avoiding neurological complications [26–28].

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

References

1. Pobihun, N.G. (2020), "Research on the impact of physical activity and stress on the thyroid gland in hypothyroidism". *Scientific and Practical Journal*, 3(№ 4 (12)), P. 97–101. available at: https://art-of-medicine.ifnmu.edu.ua/index.php/aom/article/view/402
2. Mubashir Alam, K., Tasnim Ahsan, Urooj Lal, R., Ruqshanda Jabeen, and Saad Farooq. (2017), "Subclinical hypothyroidism: frequency, clinical manifestations, and indications for treatment". *Pakistan Journal of Medical Sciences*, 33(4), P. 818–822. DOI: 10.12669/pjms.334.12921
3. Feldman, A.Z., Shrestha, R.T., & Geneslaw, J.V. (2013), "Neuropsychiatric manifestations of thyroid diseases". *Endocrinology and Metabolism Clinics of North America,* 42(3), P. 453–476. DOI: 10.1016/j.ecl.2013.05.005
4. Almeida, O.P., Alfonso, H., Flicker, L., Hankey, G., Chubb, S.A.P., & Yeap, B.B. (2011), "Thyroid hormones and depression". *The American Journal of Geriatric Psychiatry,* 19(9), P. 763–770. DOI: 10.1097/jgp.0b013e31820dcad5
5. Bunevicius, R., & Prange, A.J. (2010), "Thyroid diseases and mental disorders: cause and effect or only comorbidity?", *Current Opinion in Psychiatry*, 23(4), P. 363–368. DOI: 10.1097/yco.0b013e3283387b50
6. Yarach, D., Kukharska, A., Raevska-Rager, A., & Latska, K. (2012), "Cognitive functions and mood during chronic thyrotropin-suppressive L-thyroxine therapy in patients with differentiated thyroid carcinoma". *Journal of Endocrinological Research,* 35(8), P. 760–765.
7. Demartini, B., Ranieri, R., Masu, A., Selle, V., Scaroni, C., & Gambini, O. (2014), "Depressive symptoms and major depressive disorder in patients with subclinical hypothyroidism". *Journal of Nervous and Mental Disease*, 202(8), P. 603–607. DOI: 10.1097/nmd.0000000000000168
8. Kozhyna, N.M., Tovazhnyanska, O.L., Markova, M.V., Zelenska, K.O., & Kauka, O.I. (2020), "Features of primary hypothyroidism in forcibly displaced persons as a basis for the formation of cognitive and anxiety-depressive disorders". *Problems of Endocrine Pathology,* 73(3), P. 25–32. DOI: https://doi.org/10.21856/j-PEP.2020.3.03
9. Marian, G., Nica, E.A., Ionescu, B.E., & Guinea, D. (2009), "Hyperthyroidism – cause of depression and psychosis: clinical case". Journal of Medicine and Life, 2(4), P. 440–442.
10. Dabldey, A.R., & Sippel, R.S. (2020), "Hyperthyroidism". *Gland Surgery*, 9(1), P. 124-135. DOI: 10.21037/gs.2019.11.01
11. Soiri, I.N., & Reidpat, D.D. (2013), "Health forecasting review", *Environmental Health and Preventive Medicine,* №18, P. 1–9. https://doi.org/10.1007/s12199-012-0294-6
12. Armstrong, J.S. (2001), *"Principles of forecasting: A handbook for researchers and practitioners".* Norwell: Kluwer Academic Publishers. 458 p.
13. Savchuk, O. (2021), "Application of machine learning in clinical psychology". available at: https://ojs.tdmu.edu.ua/index.php/kl-stomat/article/download/6147/5624/21744
14. Hodovanyets, O.I. & Rozhko, M.M. (2015), "Features of the formation of the dental arch system in children with diffuse non-toxic goiter", *Bulletin of Biology and Medicine Issues*, Vol. 2, Issue 2(119), P. 37–39.
15. Zelinska, N.B., Tereshchenko, A.V., & Rudenko, N.G. (2013), "The state of providing specialized assistance to children with endocrine pathology in Ukraine in 2012 and prospects for its development". *Ukrainian Journal of Pediatric Endocrinology,* No 3, P. 31–39.
16. Lytvynenko, O. (2021), "Innovative approaches to processing psychological data using machine learning". available at: https://openarchive.nure.ua/server/api/core/bitstreams/86aa5c34-6f0f-44a4-ac51-76e7d191e085/content
17. Livingston, E.H. (2019), "Subclinical hypothyroidism". *JAMA,* 322(2), 180 p. DOI: 10.1001/jama.2019.9508
18. Kyslyi, O. (2021), "Using machine learning methods in psychological research". available at: https://ami-ejournal.cdu.edu.ua/article/view/4158/4438
19. Zbarazhskyi, M. (2021), "Analysis of psychological data using machine learning techniques". available at: https://ojs.tdmu.edu.ua/index.php/visnyk-nauk-dos/article/view/8460/7880
20. Ivanov, I. (2021), "Advanced machine learning techniques in psychology". available at: https://ela.kpi.ua/items/20f948bd-5b8a-420e-a86e-70d86be50866
21. Kyslyi, O. (2020), "Using machine learning methods in psychological research. Artificial Intelligence Methods". available at: https://ami-ejournal.cdu.edu.ua/article/view/4158/4438
22. Petrov, A. (2021), "Applications of machine learning in psychological studies". available at: https://ela.kpi.ua/server/api/core/bitstreams/17dfafd7-9874-4b51-b865-f20ea63e5076/content
23. Cutler, A., & Zhao, G. (2001), "*PERT – Perfect Random Tree Ensembles*". Computing Science and Statistics, № 33, P. 490–497.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

24. Babak, V.P., Biletskyi, A.Ya., Prystavka, O.P., & Prystavka, P.O. (2001), *"Statistical data processing"*. Kyiv: MIVVTS. 388 p.
25. Breiman, L. (2001), *"Random Forests"*. Machine Learning, P. 45.
26. Mochurad, L. & Ilkiv, A. (2022), "Advanced method of medical classification using parallelization algorithms". *Computer Systems and Information Technologies,* (1), P. 23–31. DOI: 10.31891/CSIT-2022-1-3
27. Ittermann, T., Fiolka, H., Baumeister, S.E., Appel, K., & Graabe, H.J. (2015), "Diagnosed thyroid diseases associated with depression and anxiety". *Social Psychiatry and Psychiatric Epidemiology,* 50(9), P. 1417–1425. DOI: 10.1007/s00127-015-1043-0
28. Martino, J.P. (1972), *"Forecasting the progress of technologies"*. New York, New York: Gordon and Breach Science Publishers. № 2. 15 p.

Надійшла (Received) 08.05.2024

Відомості про авторів / About the Authors

Гулієв Нурал Бахадур огли – Харківський національний університет радіоелектроніки, аспірант, Харків, Україна; e-mail: nural.huliiev@nure.ua; ORCID ID: https://orcid.org/0000-0003-2123-0377

Huliiev Nural Bahadur ohli – Kharkiv National University of Radio Electronics, PhD Student, Kharkiv, Ukraine.

ВИБІР МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ РОЗВИТКУ ПСИХОЛОГІЧНИХ РОЗЛАДІВ У ЛЮДЕЙ ІЗ ГІПОТИРЕОЗОМ ТА ГІПЕРТИРЕОЗОМ

Предметом дослідження в статті є ендокринологічні захворювання, а саме: аналіз ускладнень у людей з гіпотиреозом та гіпертиреозом. Відомо, що ці хвороби виникають безсимптомно або можуть бути наслідками інших захворювань, через що люди не підозрюють, на що саме хворіють. Пізніше хвороби зазвичай спричиняють ускладнення в організмі, найнебезпечнішими з яких є психологічні розлади: депресія, маніакальність, агресивність тощо. Тому **метою роботи** є розроблення методів прогнозування виникнення неврологічних погіршень організму в людей, у яких вже виявлено ендокринологічні захворювання. У статті розв'язувалися **завдання** вибору кращих моделей прогнозування виникнення психологічних розладів у пацієнтів з ендокринологічними проблемами. Аналізувалися методи машинного навчання, поширені в медичній галузі, та обирався один із них, який найбільш ефективно вирішує всі поставлені завдання. У виборі критеріїв узято до уваги потенційні проблеми з медичними та психологічними показниками. Упроваджувався **метод** лінійної адитивної згортки для вибору найкращих за результатами альтернатив, із принципом Парето, спрямованим на вилучення непідхожих альтернатив через те, що всі ознаки мають менші показники, ніж в інших варіантах. Для експерименту всі ознаки конвертувалися в кількісні для підрахунку значень згортки. Критерії оцінки наведені в роботі. Досягнуто таких **результатів**: моделлю прогнозування в подальшому дослідженні окресленого завдання буде випадковий ліс. **Висновки**: досліджено методи прогнозування та обрано більш оптимальну модель за допомогою лінійної адитивної згортки, а саме алгоритм "випадковий ліс", розглянуто переваги й недоліки зазначеної моделі. Більш детальний аналіз її розроблення буде запропоновано в наступних статтях. Надано математичний опис обраного методу прогнозування, що містить потенційні способи реалізації та кроки побудови алгоритму одного із цих способів.

Ключові слова: гіпотиреоз; гіпертиреоз; психологічні розлади; прогнозування; лінійна адитивна згортка; принцип Парето; алгоритм "випадковий ліс"; дерево рішень; індекс *Gini*.

Бібліографічні описи / Bibliographic descriptions

Гулієв Н. Б. Вибір моделей машинного навчання для прогнозування розвитку психологічних розладів у людей із гіпотиреозом та гіпертиреозом. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 76–85. DOI: https://doi.org/10.30837/2522-9818.2024.2.076

Huliiev, N. (2024), "Choice of machine learning models for predicting the development of psychological disorders in people with hypothireosis and hyperthireosis", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 76–85. DOI: https://doi.org/10.30837/2522-9818.2024.2.076

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

УДК 005.93:005.71-021.131 DOI: https://doi.org/10.30837/2522-9818.2024.2.005

А. Жук, Є. Павелко

ДОСЛІДЖЕННЯ ВПЛИВУ ГЛОБАЛЬНИХ КАТАСТРОФ НА ПОВЕДІНКУ ПОКУПЦЯ ІНТЕРНЕТ-МАГАЗИНІВ УКРАЇНИ

Предметом дослідження в статті є вплив глобальних катастроф, зокрема пандемії COVID-19 та російської збройної агресії проти України, на споживчу поведінку українців у інтернет-магазинах, зокрема на зміни в потребах споживачів і адаптація маркетингових стратегій підприємств. **Мета роботи** – аналіз змін споживчої поведінки в умовах глобальних катастроф і розроблення рекомендацій для бізнесу щодо ефективного реагування на нові виклики ринку. У статті виконуються такі **завдання**: досліджується вплив пандемії COVID-19 та російської збройної агресії проти України на споживчу поведінку в інтернет-магазинах; визначаються ключові чинники, що позначаються на рішеннях споживачів під час кризових ситуацій; аналізуються актуальні маркетингові стратегії та інструменти, що застосовуються компаніями в умовах кризи. Упроваджуються такі **методи**: математичне оброблення даних для аналізу результатів опитувань і статистичних досліджень; компаративний аналіз для порівняння поведінки споживачів до та під час пандемії; експертне оцінювання для визначення ефективності маркетингових стратегій; контент-аналіз для дослідження трендів у соціальних мережах та інших онлайн-платформах. **Досягнуті результати**. Сформульовано принципи адаптації маркетингових стратегій в умовах пандемії COVID-19 та воєнного стану. Визначено, що пріоритетами для споживачів стають здоров'я, доступність основних товарів і безпека, зокрема й кібербезпека. Виявлено зміни в споживчій поведінці: люди стали більш уважними до ціни, якості товарів та віддають перевагу продуктам місцевих виробників. Проведено маркетингове дослідження серед клієнтів компанії "Горгани", яке показало, що попит на товари для активного відпочинку залишається високим, навіть у період війни, і споживачі віддають перевагу якісним і доступним товарам вітчизняного виробництва. **Висновки**: застосування методу аналізу змін споживчої поведінки дало змогу визначити ключові фактори, що впливають на рішення про покупку в умовах глобальних криз, знання сприяє тому, що підприємства вчасно адаптують свої маркетингові стратегії та зберігають конкурентні переваги; оптимізація асортименту товарів та вдосконалення цифрової присутності є ключовими факторами успіху на сучасному ринку; підприємства, які швидко реагують на зміни споживчих пріоритетів і використовують новітні технології для комунікації з клієнтами, мають більше шансів на успіх.

Ключові слова: глобальні катаклізми; поведінка покупців; інтернет-магазини; пандемія; війна; економічна нестабільність.

Вступ

У сучасному світі, ставши свідками глобальних катастроф, таких як пандемія COVID-19 та російська збройна агресія проти України, людство стикається з несподіваними та надзвичайно складними викликами, що безпосередньо впливають на різні аспекти життя, зокрема споживчу поведінку. Особливо важливими в цьому контексті є зміни в споживчих звичках, що виявляються в інтернет-магазинах.

Пандемія COVID-19 та війна в Україні викликали серйозні зміни в житті людей і призвели до трансформації способів їхнього споживання. Обмежувальні заходи, упроваджені для стримування поширення вірусу, значно змінили звички клієнтів, змусивши їх перенести свої покупки з офлайн-магазинів до інтернет-платформ. З іншого боку, війна призвела до економічної нестабільності, загрози безпеки та невизначеності майбутнього, що вразливо позначилося на споживчих звичках, а також на діяльності електронної комерції [1].

Однак мало що відомо про те, як саме ці глобальні катастрофи впливають на поведінку покупців у інтернет-магазинах, які конкретні чинники визначають їхні рішення під час кризових ситуацій, та які стратегії можуть бути ефективними для підтримки й розвитку бізнесу в цих умовах. Тому виникає необхідність в глибокому аналізі окреслених питань, щоб зрозуміти вплив глобальних катастроф на споживчу поведінку в інтернет-магазинах та розробити стратегії, що дадуть змогу бізнесу адаптуватися до нових реалій та зберегти свою конкурентоспроможність.

Аналіз останніх публікацій

Автори розпочали розв'язання зазначеної проблеми та аналіз стандартних підходів до неї.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Вивчення впливу пандемії COVID-19 на ринок і загальний економічний стан описано в різних наукових працях вітчизняних і закордонних дослідників. Наприклад, Д. Долбнєва обґрунтовувала наслідки COVID-19 для світової економіки [2]; І. Вагнер та І. Демко досліджували вплив COVID-19 на розвиток малого та середнього бізнесу в Україні [3]; С. Кулицький аналізував перспективи української економіки в умовах пандемії COVID-19 [4]. Також чимало уваги приділялося вивченню змін поведінки споживачів. Наприклад, В. Комірна та О. Санжак розглядали реалізацію різних підходів до аналізу поведінки споживача [5]; Л. Василькевич провела аналітику структури поведінки споживачів і описала їх основне значення в межах економічних відносин [6]; О. Шаманська конкретизувала фактори й мотиви, що впливають на особливості споживчої поведінки домогосподарств, та чинники, що спонукають їх до дії [7]. Тому можна стверджувати про наявність широкого спектра досліджень поведінки споживачів і впливу COVID-19 на розвиток ринку та господарства.

Мета й завдання роботи

Метою статті є визначення впливу глобальних катастроф, зокрема пандемії COVID-19, на поведінку покупців інтернет-магазинів України та розроблення рекомендацій для адаптації маркетингових стратегій до нових умов.

Завдання дослідження передбачають аналіз наукових публікацій щодо впливу пандемії COVID-19 на поведінку споживача; оцінювання змін у поведінці споживачів під час пандемії та після її завершення; вивчення сучасних маркетингових стратегій, що впроваджуються підприємствами в умовах глобальних катастроф; розроблення рекомендацій для інтернет-магазинів щодо адаптації їх маркетингових стратегій до нових умов.

Матеріали та методи

У роботі було використано такі **матеріали та методи**: аналіз літературних джерел, оснований на ґрунтовному вивченні наукових статей, звітів і досліджень щодо впливу пандемії на поведінку споживачів і маркетингові стратегії; проведення маркетингового опитування серед 186 осіб під час пандемії для оцінювання змін у пріоритетах і поведінці споживачів; використання інформації дослідницьких агентств, зокрема *Buzzfactory Ukraine* та *Factum Group Ukraine*, для визначення глобальних маркетингових трендів під час пандемії.

Результати дослідження

Результати дослідження показали значні зміни в поведінці споживачів під час пандемії COVID-19. Переважна більшість опитаних (89%) зосереджували увагу на фізіологічних потребах, таких як забезпечення життєвих потреб та безпека. Змінилися маркетингові стратегії компаній, що спрямовані на задоволення нових потреб споживачів. Агенція *Buzzfactory Ukraine* визначила такі тренди, як зростання популярності челенджів, прямі трансляції, зміна напряму популярності серед інфлюенсерів.

Недостатньо вивченими є питання поведінки споживача в умовах COVID-19 з огляду на тренди маркетингу-2021, що зумовлює необхідність більш детального його вивчення. Вплив пандемії COVID-19 на споживчу поведінку вже розглядався в багатьох дослідженнях, де, зокрема, обговорювалися питання повернення до старих звичок або формування нових споживчих моделей [8].

У період пандемії сучасні люди звертають увагу на своє здоров'я, безпеку близьких, доступ до необхідних товарів і фінансовий стан. Це загальне хвилювання виявляється по-різному та позначається на споживачах. Для впливу на клієнтів компанії розробляють низку маркетингових стратегій та інструментів. Зважаючи на ідеї ієрархії потреб А. Маслоу (рис. 1), можна порівняти задоволення потреб споживача до періоду пандемії з айсбергом, де на поверхні можна спостерігати творчість та задоволення певних духовних потреб. Наприклад, це можуть бути екологічні аспекти, які не мають негативного впливу на довкілля, або соціальна активність, визнання з боку суспільства тощо. Під час COVID-19 сучасний споживач реорганізував свої пріоритети та цінності, переосмисливши їх у контексті власного добробуту та безпеки. Відповідно до теорії А. Маслоу про ієрархію потреб людей стали більш уважними до фізіологічних чинників – здоров'я та безпеки. У цей період самовираження та особистий розвиток відступають на другий план, зокрема для осіб, що перебувають у вразливому стані, які хворіють або ті люди, які більше турбуються про своє здоров'я.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

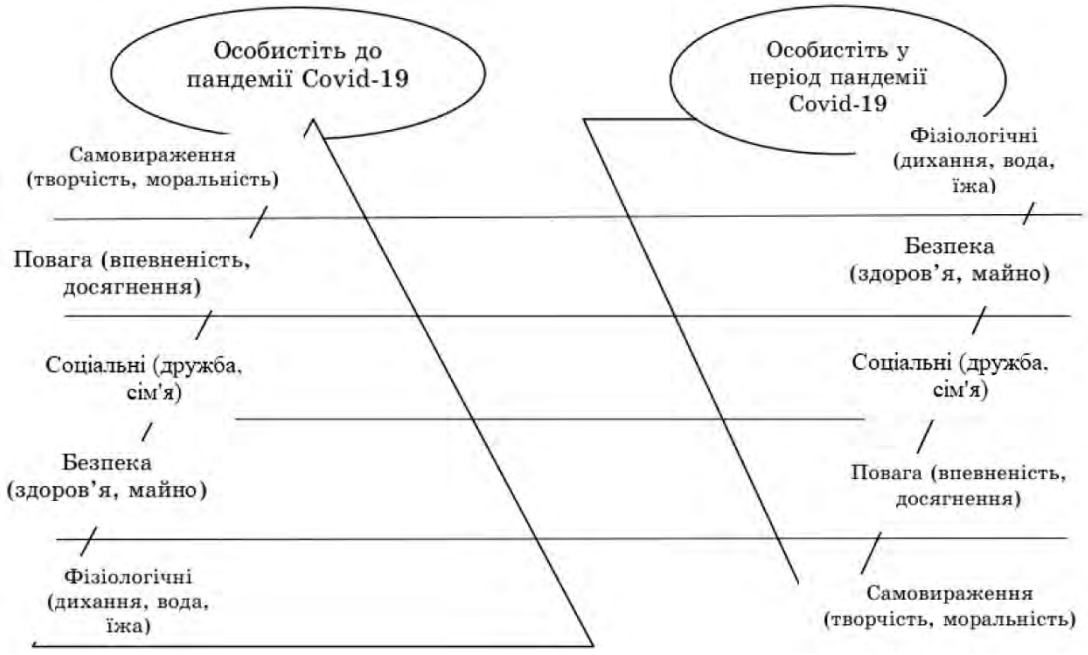


Рис. 1. Ієрархія потреб споживачів А. Маслоу до та під час пандемії COVID-1

Маркетингове дослідження [9], проведене серед 186 осіб під час пандемії, свідчить, що переважна більшість опитаних (близько 89%) уважні до фізіологічних чинників, таких як забезпечення життєвих потреб і безпека. Лише 11% віддають перевагу питанням безпеки, тоді як 7% звертаються до соціальних аспектів життя (рис. 2).

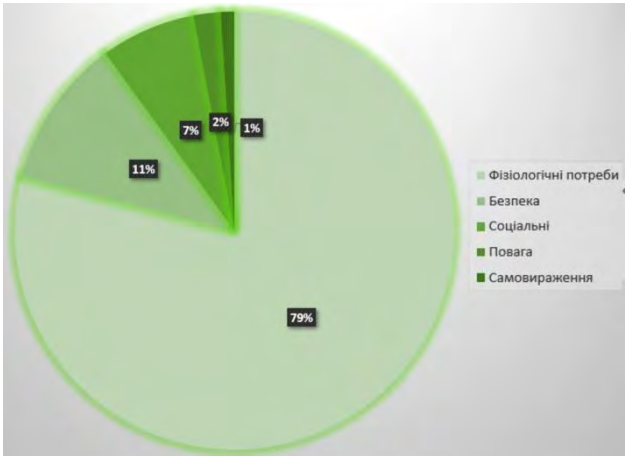


Рис. 2. Питома вага потреб споживачів за ієрархією потреб А. Маслоу в період пандемії COVID-19

Агенція *Buzzfactory Ukraine* ділиться дослідженням глобальних трендів у сфері маркетингу впливу під час кризи, викликаної епідемією COVID-19. Зокрема агенція зосереджувала увагу на кількох тенденціях.

1. *Челендж як спосіб боротьби з нудьгою*. У цей період спостерігаємо значну кількість челенджів різних видів, починаючи від соціальних, спрямованих на допомогу медичним працівникам і хворим, і завершуючи спортивними, що сприяють популяризації здорового способу життя серед населення.

2. *Прямі трансляції.* Відомі блогери розповідають про досвід, як вони проводять карантин, зірки організовують живі концерти, фітнес-тренери влаштовують прямі ефіри з тренуваннями, а психологи консультують, як залишатися в гармонії під час кризи.

3. *Зміна напряму популярності серед інфлюенсерів.* Набувають популярності категорії блогерів у сфері харчування, оскільки люди почали більше часу проводити вдома, рідше відвідувати ресторани й кав'ярні [9].

Factum Group Ukraine разом із Всеукраїнською рекламною коаліцією також проаналізували вплив COVID-19 на маркетинг в Україні. Унаслідок цього дослідження було виокремлено п'ять основних трендів.

1. Редукція та реструктуризація бюджетів: компанії зменшують фінансування маркетингу.

2. Примусова ізоляція, що приводить до повної цифрової трансформації. Цей процес передбачає перехід до інтернет-реклами, електронної торгівлі, онлайн-сервісів, досліджень у мережі, електронного обміну документами та віртуальних зустрічей.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

3. Фокус на споживача: у ситуації загострення конкуренції підприємства все більше уваги приділяють потребам клієнтів. Вони зазначають, що це охоплює як розроблення нових продуктів / послуг, так і їх рекламу (індивідуальний підхід до спілкування).

4. Підприємства, спрямовані не на вдосконалення наявних продуктів, а на створення нового типу товару.

5. Компанії, що впроваджують систематичні та постійні моніторинги показників, переходять до ситуативного маркетингу, коригують маркетингові стратегії та виготовляють нові відповідні продукти й комунікації. Це вимагає підвищеної креативності команд і швидкості в прийнятті управлінських рішень. Крім того, серед останніх тенденцій необхідно наголосити на зростанні корпоративної відповідальності бізнесу, на що звертають увагу сучасні споживачі.

Отже, з огляду на останні тенденції в маркетингу та пандемію COVID-19 люди змінюють свої потреби в товарах: вони уважніше обирають продукцію, що найбільше відповідає їхнім вимогам, і звертають увагу на ціну. Що вища цінність товару, то розсудливіше приймається рішення. Крім того, пандемія COVID-19 порушує проблеми кібербезпеки, оскільки кризові ситуації зазвичай стимулюють діяльність різних хакерських груп. Основними факторами, що потенційно сприяли зростанню деструктивної кіберактивності, є: збільшення кількості потенційно вразливих з'єднань, які можуть призвести до компрометації інформації або самої організації, або її працівників; інтенсифікація електронних платежів, що привертає більшу увагу кіберзловмисників до шахрайської діяльності; зростання кількості фішингових атак – збільшення фальшивих листів (із *malware*-вкладеннями) та фальшивих вебсайтів (для збору персональної та банківської інформації громадян); додаткове посилення паніки може бути однією з цілей операцій впливу з боку інших держав, що можуть використовувати ситуацію у власних інтересах [10].

Пандемія COVID-19 також суттєво вплинула на різні сектори економіки, зокрема на індустрію гостинності, яка зазнала значних змін і втрат [11]. Період карантину виявився найбільш сприятливим для підприємств у проведенні онлайн-взаємодії з наявними й потенційними клієнтами. Це підтверджується також статистикою змін у способах спілкування брендів. За інформацією дослідницької агенції *Sprout Social*, активність користувачів щодо публікацій сучасних компаній у соціальних мережах значно зросла під час карантину. Іншими словами, споживачі стали активніше реагувати на контент брендів, зокрема ставлять уподобайки, ретвітять і коментують, що сприяє швидшому поширенню інформації [12].

Отже, після аналізу власних досліджень та інших установ узагальнимо напрями розвитку маркетингу в умовах пандемії та сформулюємо рекомендації для виробників.

1. Підвищення рівня інвестицій у присутність в інтернеті.

2. Забезпечення максимальної зручності у зворотній комунікації від клієнта до виробника.

3. Зосередження уваги на підвищенні емоційного зв'язку з клієнтом.

4. Упровадження гнучких варіантів оплати.

5. Спрямування зусиль на збільшенні обсягу контенту за короткий період часу.

Було з'ясовано, що пандемія COVID-19 спричинила суттєві зміни в бізнес-процесах і методах проведення досліджень, зокрема у сфері управління ризиками [13]. В інтернеті спостерігається перенасиченість інформацією, тому споживачі зосереджуються на каналах, де доступний максимально потрібний, актуальний і корисний контент. Підприємства вкладають бюджет у виробництво відео для швидкого поширення цінної інформації серед своєї аудиторії. У такому середовищі споживачі потребують упевненості в бренді, привабливого контенту, особливо на фоні нестабільності.

Підприємці звертають увагу на онлайн-торгівлі, закриваючи офлайн-магазини для захисту персоналу та громадськості. Це важливий висновок, адже в такі періоди бренди ставлять спільні інтереси вище за прибуток і вживають заходів, щоб підтримати загальну боротьбу, з якою зіткнувся весь світ.

Пропозиції для споживачів:

1) сприяння розвитку місцевих бізнесів;

2) оптимізація та ефективне використання ресурсів;

3) зосередження уваги на збільшенні власних фінансових накопичень.

Пандемія також викликала різноманітні реакції серед споживачів, які адаптувалися до нових умов, змінюючи свої пріоритети та поведінку [14]. В умовах економічної та соціальної нестабільності перше місце посідає питання збереження власного здоров'я, що привело до зростання попиту на одноразові предмети, які зменшують ризик зараження вірусом. Серед товарів, популярність яких зросла в онлайн-торгівлі, найбільший попит

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

мають предмети першої необхідності, зокрема продукти харчування / бакалійні товари, побутова хімія та засоби особистої гігієни. Проте легка промисловість найбільше постраждала через зменшення обсягів продажу. Отже, маємо поляризовану картину: деякі групи товарів зазнають зростання попиту в наявних клієнтських базах, тоді як інші спостерігають зниження активності [15].

Популярність товарів для активного відпочинку формується під впливом різних факторів, основним з яких є зростання туризму всередині країни через обмеження виїзду за кордон у зв'язку з воєнним станом. Такий вид відпочинку стає не лише хобі або розвагою, але й можливістю для морального відновлення як для військових, так і для цивільного населення. Збільшення кількості туристів і громадян, що подорожують, приводить до зростання попиту на товари для активного відпочинку: туристична амуніція, кемпінгове обладнання, рюкзаки, намети, спальні мішки тощо [16]. Туристичне спорядження стає необхідним як для військових, так і для цивільного населення під час війни, в умовах без світла, води, або в разі перебування в укриттях, або для тих, хто змушений часто переїжджати та шукає відповідні товари. Магазини, що спеціалізуються на якісному, легкому, міцному, компактному та функціональному спорядженні, привертають увагу споживачів і стимулюють попит.

Як було зазначено вище, у сучасних умовах нестабільності та воєнного стану в країні споживачі стали обмежувати свої витрати, переважно заощаджуючи на розвагах, відпочинку та спорті. Однак, купуючи товари з цієї категорії, вони все ще мають певні очікування від брендів [17]:

- 55% клієнтів очікують, щоб бренди зважали на екологічність і вплив на довкілля;
- 54% хочуть відновлення асортименту товарів і брендів, що існував до війни;
- 53% прагнуть бачити менше розважальної реклами;
- 53% бажають мати доступніші пропозиції товарів.

Беручи до уваги актуальні тенденції на ринку та зміну споживчих пріоритетів, підприємства, що спеціалізуються на виробництві та продажу товарів для активного відпочинку, мають швидко адаптуватися до нових ринкових умов і виявляти інноваційний підхід у своїй маркетинговій стратегії. Ритейлери змушені були оперативно пристосовуватися до нових обставин під час пандемії COVID-19, що привело до значних змін у конкурентному середовищі [18]. Проведення маркетингових досліджень ринку товарів для активного відпочинку є ключовим інструментом для розуміння споживчих потреб і вимог, виявлення переваг і недоліків конкурентів, а також визначення оптимальних стратегій продажу та просування продукції. У цьому контексті проведено маркетингове дослідження серед клієнтів компанії "Горгани" – мережі магазинів, що спеціалізуються на продажу товарів для активного відпочинку [19]. Ця мережа працює на українському ринку з 2005 р. й налічує 12 фізичних магазинів, а також має інтернет-присутність. У дослідженні взяли участь 774 респонденти, серед яких 51% чоловіків і 49% жінок, з переважною кількістю опитаних (47%) у віці від 18 до 30 років.

За результатами опитування 82% респондентів відвідують магазини подібного типу, навіть у період воєнного стану (рис. 3).

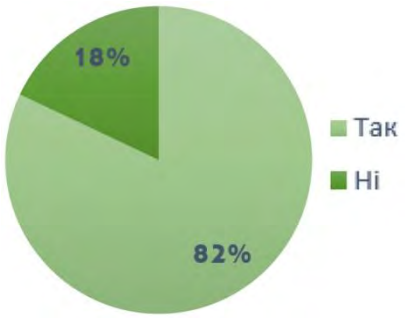


Рис. 3. Результати опитування щодо відвідування споживачів магазинів із товарами для активного відпочинку

Серед товарних категорій, що мають найбільший попит у клієнтів компанії "Горгани", переважають спорядження та продукти харчування, оскільки ці товари необхідні в зоні бойових дій, також ця група товарів має пріоритет і у волонтерів (рис. 4).

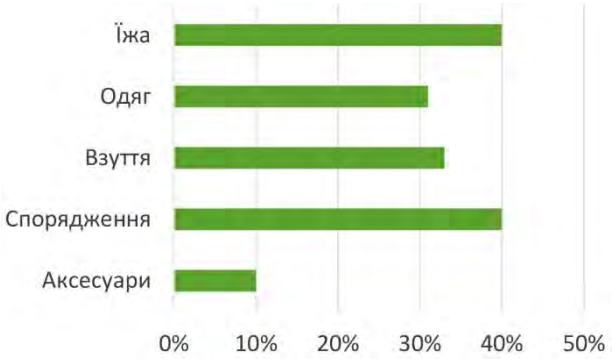


Рис. 4. Товарні категорії, що найчастіше купують

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Серед визначальних чинників, які найбільше впливають на рішення про здійснення покупки, якість посідає перше місце, а на другому місці – ціна. Приймаючи рішення про покупку товарів певної категорії, споживачі насамперед прагнуть придбати якісний продукт (рис. 5).

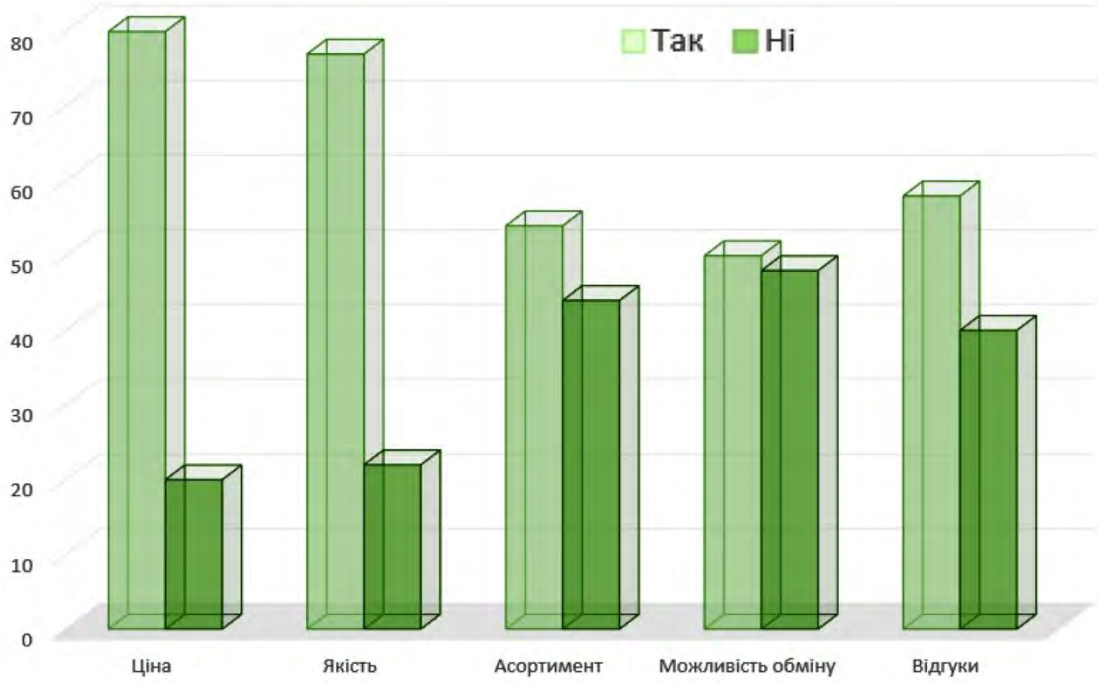


Рис. 5. Чинники, що впливають на прийняття рішення про покупку

Щодо підтримки українських виробників товарів для активного відпочинку спостерігається позитивна тенденція, оскільки 81% опитаних виявили схильність саме до вітчизняних товарів (рис. 6), 11% віддають перевагу закордонним виробникам через вищий рівень довіри та досвід попередньої покупки. Майже половина респондентів (47%) обирають онлайн-шопінг, тоді як третина опитаних прагне придбати товари у звичайних офлайн-магазинах (рис. 7).

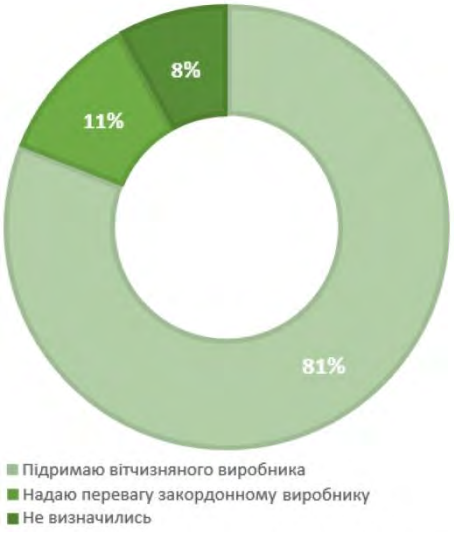


Рис. 6. Прихильність опитаних до країни-виробника продукції

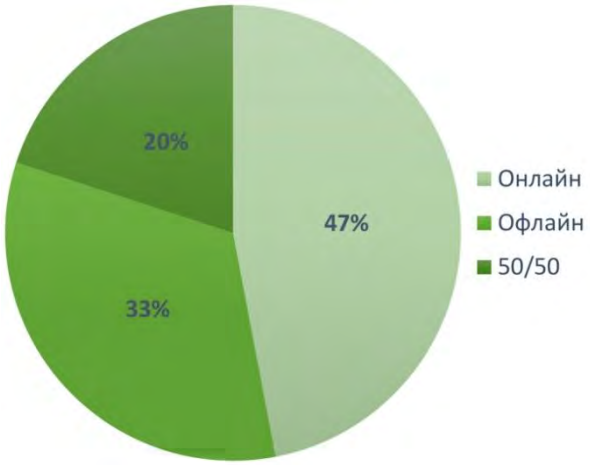


Рис. 7. Місця купівлі товарів для активного відпочинку

Споживачі найбільш зручними каналами комунікації з брендом назвали офіційний вебсайт (33%), telegram-канал (31%) та соціальні мережі (28%) (рис. 8). Саме ці канали компанії необхідно обрати для розроблення комунікаційної стратегії та підтримки зв'язку з клієнтами [20].

Як показують результати нашого дослідження, товари для активного відпочинку, зокрема туристичне спорядження, залишаються популярними навіть у період війни. Крім того, на групу цих товарів попит

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

продовжує зростати. До деяких категорій товарів підвищується інтерес як з боку цивільного населення, так і серед військових. Однак змінилися чинники, що впливають на прийняття рішення про покупку, особливо ціна, оскільки зниження купівельної спроможності людей стає очевидною тенденцією.

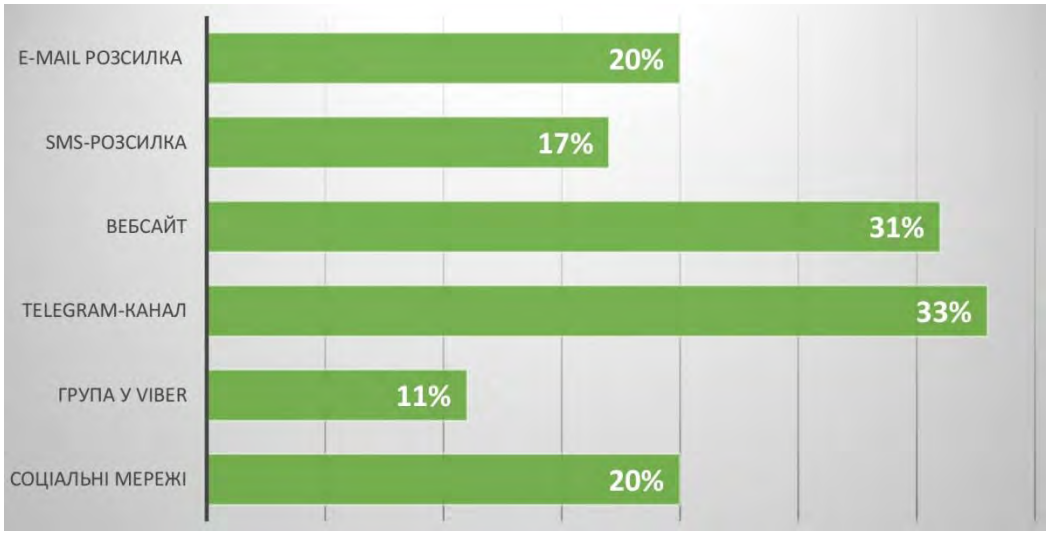


Рис. 8. Найбільш зручні для клієнтів канали комунікації з брендом

Магазинам, що спеціалізуються на товарах для активного відпочинку, варто переглянути свій асортимент, додавши до нього продукцію вітчизняних виробників за більш доступними цінами, але зі збереженням високої якості. Команда "Горгани" активно працює над новими ідеями для збереження своїх позицій на ринку, виявляючи інтерес до співпраці з новими брендами, які пропонують тактичне спорядження. Незважаючи на війну, розширюється мережа магазинів: відкрили новий заклад у Буковелі, а також у травні 2023 р. почав працювати перший магазин у м. Бухаресті, Румунія. Війна стала поштовхом до виходу українських фірм на міжнародний ринок.

Висновки

З моменту повномасштабного вторгнення українські споживачі швидко адаптувалися до умов воєнного стану, переглянувши свої пріоритети та фактори, що впливають на їхні рішення щодо покупок. Відповідно до цього український бізнес також реагував на ці зміни, адаптуючи свою діяльність для повноцінної функціональності та задоволення потреб клієнтів. Хоча споживання деяких категорій товарів, таких як продукти харчування та ліки, зазнали незначних змін, речі для активного відпочинку, зокрема туристичне спорядження, стали категорією товарів, на яку попит значно знизився на початку війни. Однак результати дослідження та енергетичні умови, що склалися взимку 2023 р., підтвердили, що без зазначених товарів було б набагато важче. Категорії "спорядження" та "їжа" мають попит із самого початку воєнного конфлікту як серед цивільного населення, так і серед військових. З огляду на підвищену прихильність до українських брендів магазинам товарів для активного відпочинку варто розширити свій асортимент вітчизняними товарами з установленням доступних цін.

Дослідження демонструє, що під час глобальних катастроф зростає попит на товари першої необхідності та предмети розкоші. Підприємства, які можуть вчасно реагувати на зміни споживчих пріоритетів і використовувати нові технології для комунікації з клієнтами, мають більше шансів на успіх. Також важливо зазначити, що COVID-19 став джерелом не лише проблем, але й нових можливостей для бізнесу, що супроводжували період потрясінь. Перспективами подальших досліджень є обґрунтування рівня втрат на ринку товарів першої необхідності через COVID-19.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Список літератури

1. Пандемія COVID-19 в Україні: соціальні наслідки / за наук. ред. В.П. Степаненка. Київ: ТОВ НВП «Інтерсервіс». 2021. 406 с.
2. Долбнєва Д.В. Вплив COVID-19 на економіку країн світу. Світова економіка та міжнародні відносини. URL: https://www.problecon.com/export_pdf/problems-of-economy2020-1_0-pages-20_26.pdf (дата звернення: 27.05.2024)
3. Вагнер І.М., Демко І.І. Вплив COVID-19 на економічний розвиток малого і середнього бізнесу в Україні. *Вісник Університету банківської справи*. 2020. № 1(37). С. 59–66.
4. Кулицький С. Оцінка перспектив розвитку української економіки в умовах пандемії COVID-19. URL: http://nbuviap.gov.ua/index.php?option=com_content&view=article&id=4890:otsinka-perspektiv-rozvitku-ukrajinskoji-ekonomiki-v-umovakhpandemiji-covid-376 (дата звернення: 27.05.2024).
5. Комірна В.В., Санжак О.Є. Особливості різних підходів до дослідження поведінки споживача. Прометей. 2013. № 1(40). URL: http://irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE_FILE_DOWNLOAD=1&Image_file_name=PDF/ Prom_2013_1_35.pdf (дата звернення: 28.05.2024).
6. Василькевич Л.О. Структура поведінки споживачів і характеристика її основних компонентів у системі економічних відносин. *Економіка і регіон.* 2011. № 4(31). С. 187–191.
7. Шаманська О.С. Споживча поведінка домогосподарств: основні чинники та мотиви. URL: http://www.rusnauka.com/9_NND_2012/Economics/6_104898.doc.htm (дата звернення: 29.05.2024).
8. Шет Дж. Вплив Covid-19 на поведінку споживачів: Чи повернуться старі звички або зникнуть? *Журнал бізнес-досліджень,* 2020. № 117. P. 280–283. DOI: 10.1016/j.jbusres.2020.05.059
9. Тренди маркетингу впливу в період кризи, викликаної епідемією COVID-19. URL: https://mmr.ua/show/marketingvplivu-v-period-krizi-viklikanoyi-epidemiyeyu-covid-19 (дата звернення: 29.05.2024).
10. Дубов Д. COVID-19: ключові кібербезпекові тренди. URL: https://niss.gov.ua/sites/default/files/2020-03/cybersecuritycovid-19.pdf (дата звернення: 29.05.2024).
11. Гюрсой Д., Чі С.Г. Вплив пандемії COVID-19 на індустрію гостинності: огляд поточної ситуації та дослідницький порядок денний. *Журнал маркетингу та управління гостинності.* 2020. № 29(5). P. 527–529. DOI: 10.1080/19368623.2020.1788231 (дата звернення: 30.05.2024).
12. Янішівська Г. Інтерв'ю з членом Європейської академії маркетингу Володимиром Мельником. Які маркетингові стратегії запустила пандемія коронавірусу. URL: https://lvbs.com.ua/news/yaki-marketyngovi-strategiyi-zapustyla-pandemiya-koronavirusu/ (дата звернення: 30.05.2024).
13. Донту Н., Густафссон А. Вплив COVID-19 на бізнес і дослідження. *Журнал бізнес-досліджень,* 2020. № 117. С. 284–289. DOI: 10.1016/j.jbusres.2020.06.008
14. Кірк К.П., Ріфкін Л.С. Я обміняю тобі діаманти на туалетний папір: Реакції, копінг і адаптаційна поведінка споживачів під час пандемії COVID-19. *Журнал бізнес-досліджень.* 2020. P. 124–131. DOI: 10.1016/j.jbusres.2020.05.028
15. Як COVID-19 змінює споживчі звички і впливає на тенденції в eCom? URL: https://platon.ua/ua/news/kak-covid19-menyaet-potrebitelskie-privychki-i-vliyaet-na-tendenczii-vecom.htm (дата звернення: 30.05.2024).
16. Гринкевич С.С. Регіональна маркетингова політика у галузі туризму: монографія. Львів, 2017. 234 с.
17. Закупи під час війни: на чому економлять українці та чому переходять в онлайн? URL: https://hmarochos.kiev.ua/2022/09/07/zakupy-pid-chas-vijny-na-chomu-ekonomlyat-ukrayinczi-ta-chomu-perehodyat-v-onlajn/ (дата звернення: 30.05.2024).
18. Пантано Е., Піцці Г., Скарпі Д., Деніс К. Конкуренція під час пандемії? Злети та падіння ритейлерів під час спалаху COVID-19. *Журнал бізнес-досліджень.* 2020. № 116. P. 209–213. DOI: 10.1016/j.jbusres.2020.05.036 (дата звернення: 30.05.2024).
19. Мороз О.І. Індустрія гостинності: стан, тенденції розвитку та перспективи галузі в умовах війни. *Збірник праць Міжнародного науково-практичного форуму.* Львів, 2023. 466 с.
20. Файвішенко Д.С. Образи сучасності в гуманітарному знанні: *матеріали II Міжнар. наук.-практ.* Київ, 2023. 302 с. DOI: 10.31617/k.knute.2023-10-23

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

References

1. Stepanenko, V.P. (2021), *"Pandemic of COVID-19 in Ukraine"*, social consequences / according to science. Kyiv: NVP Interservice LLC, 406 p.
2. Dolbneva, D.V. (2020), "The impact of COVID-19 on the world economy". *World economy and international relations.* available at: https://www.problecon.com/export_pdf/problems-of-economy2020-1_0-pages-20_26.pdf (last accessed 27.05.2024)
3. Wagner, I.M., Demko, I.I. (2020), "The impact of COVID-19 on the economic development of small and medium-sized businesses in Ukraine". *Bulletin of the University of Banking,* No. 1 (37). P. 59–66.
4. Kulytskyi, S. "Assessment of prospects for the development of the Ukrainian economy in the conditions of the COVID-19 pandemic". available at: http://nbuviap.gov.ua/index.php?option=com_content&view=article&id=4890:otsinka-perspektiv-rozvitku-ukrajinskoji-ekonomiki-v-umovakhpandemiji-covid-376 (last accessed 27.05.2024).
5. Komirna, V.V., Sanzhak, O.E. (2013), "Features of various approaches to the study of consumer behavior. Prometheus", No. 1 (40). available at: http://irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE_FILE_DOWNLOAD=1&Image_file_name=PDF/ Prom_2013_1_35.pdf (last accessed 28.05.2024).
6. Vasylkevich, L.O. (2011), "The structure of consumer behavior and the characteristics of its main components in the system of economic relations". *Economy and region*, No. 4 (31). P. 187–191.
7. Shamanska, O.S. "Consumer behavior of households: main factors and motives". available at: http://www.rusnauka.com/9_NND_2012/Economics/6_104898.doc.htm (last accessed 29.05.2024).
8. Sheth, J. (2020). "Impact of Covid-19 on consumer behavior: Will the old habits return or die?" *Journal of Business Research,* № 117, P. 280–283. DOI: 10.1016/j.jbusres.2020.05.059
9. "Influence marketing trends during the crisis caused by the COVID-19 epidemic". available at: https://mmr.ua/show/marketingvplivu-v-period-krizi-viklikanoyi-epidemiyeyu-covid-19 (last accessed 29.05.2024).
10. Dubov, D. "COVID-19: key cyber security trends". available at: https://niss.gov.ua/sites/default/files/2020-03/cybersecuritycovid-19.pdf (last accessed 29.05.2024)
11. Gursoy, D., Chi, C.G. (2020), "Effects of COVID-19 pandemic on hospitality industry: review of the current situations and a research agenda". *Journal of Hospitality Marketing & Management*, № 29(5), P. 527–529. DOI: 10.1080/19368623.2020.1788231
12. Yanishivska, G. "Interview with Volodymyr Melnyk, a member of the European Academy of Marketing. What marketing strategies have been launched by the coronavirus pandemic". available at: https://lvbs.com.ua/news/yaki-marketyngovi-strategiyi-zapustyla-pandemiya-koronavirusu/ (last accessed 30.05.2024).
13. Donthu, N., Gustafsson, A. (2020), "Effects of COVID-19 on business and research". *Journal of Business Research,* № 117, P. 284–289. DOI: 10.1016/j.jbusres.2020.06.008
14. Kirk, C.P., Rifkin, L.S. (2020), "I'll trade you diamonds for toilet paper: Consumer reacting, coping and adapting behaviors in the COVID-19 pandemic". *Journal of Business Research*, № 117, P. 124–131. DOI: 10.1016/j.jbusres.2020.05.028
15. "How is COVID-19 changing consumer habits and influencing trends in eCom?", available at: https://platon.ua/ua/news/kak-covid19-menyaet-potrebitelskie-privychki-i-vliyaet-na-tendenczii-vecom.htm (last accessed 30.05.2024).
16. Hrynkevich. S. S. (2017), "Regional marketing policy in the field of tourism: monograph", Lviv, 234 p.
17. "Shopping during the war: what do Ukrainians save on and why do they go online?", available at: https://hmarochos.kiev.ua/2022/09/07/zakupy-pid-chas-vijny-na-chomu-ekonomlyat-ukrayinczi-ta-chomu-perehodyat-v-onlajn/ (last accessed 30.05.2024).
18. Pantano, E., Pizzi, G., Scarpi, D., & Dennis, C. (2020), "Competing during a pandemic? Retailers' ups and downs during the COVID-19 outbreak". *Journal of Business Research*, №116, P. 209–213. DOI:10.1016/j.jbusres.2020.05.036
19. Moroz, O.I. (2023), "Hospitality industry: state, development trends and prospects of the industry in the conditions of war", *Collection of works of the International Scientific and Practical Forum,* "Kamula", Lviv, 466 p.
20. Faivishenko, D.S. (2023), "Images of modernity in humanitarian knowledge", *Materials of the II International Science and Practice Conference*, *State Trade and Economy University*, Kyiv, 302 p. DOI: 10.31617/k.knute.2023-10-23

Надійшла (Received) 05.06.2024

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Відомості про авторів / About the Authors

Жук Антон Вікторович – Запорізький національний університет, аспірант, Запоріжжя, Україна; e-mail: antonzhuk.ukraine@gmail.com; ORCID ID: https://orcid.org/0009-0001-2726-8862

Павелко Євген Вадимович – Класичний приватний університет, магістр, Запоріжжя, Україна; e-mail: evgen.pavelko.dev@gmail.com; ORCID ID: https://orcid.org/0009-0003-0683-7952

Zhuk Anton – Zaporizhzhia National University, PhD Student, Zaporizhzhia, Ukraine.

Pavelko Yevhen – Classic Private University, Master of Science, Zaporizhzhia, Ukraine.

IMPACT OF GLOBAL CATASTROPHES ON ONLINE SHOPPERS' BEHAVIOR

The article's **subject matter** is the impact of global catastrophes such as the COVID-19 pandemic and Russia's military aggression against Ukraine on consumer behavior in online stores, including changes in consumer purchasing habits and the adaptation of business marketing strategies. The work **aims** to develop methods for analyzing changes in consumer behavior in the face of global catastrophes and to develop recommendations for businesses to effectively respond to new market challenges. The following **tasks** were solved in the article: Investigating the impact of the COVID-19 pandemic and war in Ukraine on consumer behavior in online stores. Identifying key factors influencing consumer decisions during crises. Analyzing current marketing strategies and tools used by companies in crisis conditions. Developing recommendations for businesses to adapt to new realities and maintain competitiveness. The following **methods** are used: Mathematical data processing for analyzing survey results and statistical studies. Comparative analysis to compare consumer behavior before and during the pandemic. Expert assessments to determine the effectiveness of marketing strategies. Content analysis to study trends on social media and other online platforms. The following **results** were obtained - formulated principles for adapting marketing strategies in the context of the COVID-19 pandemic and wartime, identified that priorities for consumers include safety, health, availability of essential goods, and cybersecurity; changes in consumer behavior were identified, with consumers becoming more price-conscious and preferring products from local manufacturers; recommendations for businesses were developed regarding effective communication with customers, increased investments in online presence, ensuring convenience of feedback, enhancing emotional connection with customers, and offering flexible payment options; methods for supporting local businesses and optimizing resource utilization by consumers were proposed, emphasizing the importance of environmental responsibility and financial resource savings; a marketing study among the clients of the company "Gorgany" was conducted, which showed that the demand for outdoor recreation products remains high even during wartime, and that consumers prefer high-quality and affordable products from domestic manufacturers. **Conclusions:** the application of the method of analyzing changes in consumer behavior allowed to identify key factors influencing purchasing decisions in times of global crises, this gives businesses the opportunity to timely adapt their marketing strategies and maintain competitive advantages; the use of developed recommendations contributes to increasing business efficiency in times of pandemic and wartime, thanks to these recommendations, companies can better meet the needs of consumers, improve service quality, and increase customer loyalty; optimization of product range and improvement of digital presence are key success factors in the modern market, companies that quickly respond to changes in consumer priorities and use advanced technologies to communicate with customers have more chances for success.

Keywords: global cataclysms; consumer behavior; online shopping; pandemic; war; economic instability.

Бібліографічні описи / Bibliographic descriptions

Жук А. В., Павелко Є. В. Дослідження впливу глобальних катастроф на поведінку покупця інтернет-магазинів України. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 86–95. DOI: https://doi.org/10.30837/2522-9818.2024.2.086

Zhuk, A, Pavelko, Y (2024), "Impact of global catastrophes on online shoppers' behavior", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 86–95. DOI: https://doi.org/10.30837/2522-9818.2024.2.086

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

УДК 621.3.095.2 DOI: https://doi.org/10.30837/2522-9818.2024.2.096

І. Невлюдов, Р. Стрілець, Д. Близнюк

ЗАБЕЗПЕЧЕННЯ ЯКІСНИХ ПОКАЗНИКІВ ФОТОПОЛІМЕРНОГО 3D-ДРУКУ ЗА ДОПОМОГОЮ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ І ТЕСТОВИХ МОДЕЛЕЙ

Предметом дослідження в статті є аналіз впливу технологічних параметрів фотополімерного друку на появу дефектів у процесі друку із використанням тестових моделей, що містять елементи, на яких позначається зміна технологічних параметрів. **Мета роботи** – визначення залежності між технологічними параметрами фотополімерного друку та дефектами, що виникають унаслідок друку з використанням моделей для тестування. У статті виконуються такі **завдання**: аналіз наявних тестових моделей і визначення елементів моделі та впливу на них технологічних параметрів. **Методи, що впроваджуються**: математичний аналіз у вигляді однофакторної лінійної регресії та емпіричний метод, що полягає в порівнянні та вимірюванні різниці між окремими тестовими зразками для отримання значень, які в подальшому використовуватимуться в регресійному аналізі. **Досягнуті результати**: визначено залежність технологічних факторів та їх вплив на елементи тестової моделі, що полягає в зміні фізичних розмірів тестових моделей, де за умови недостатнього часу експонування розміри моделі зменшуються та утворюються дефекти. У процесі збільшення часу експонування розміри моделі лінійно зростають, виникають дефекти у вигляді зникнення отворів або зміни їх розмірів. **Висновки**. У дослідженні, що передбачало друк тестових моделей та їх аналіз за допомогою лінійної однофакторної регресії, визначено та підтверджено залежність між часом експонування та фізичними розмірами моделі. Описано метод дослідження відповідності розмірів залежно від часу експонування з використанням тестових моделей і математичного моделювання у вигляді регресійного однофакторного аналізу. Надалі запропоновано визначити вплив висоти шару на час експонування та значущість окремих технологічних факторів, а також установити їх вплив на дефекти, що виникають унаслідок друку. Побудовано модель регресійного аналізу, визначено кореляції технологічних параметрів та їх вплив на показники якості. Установлено коефіцієнт детермінації побудованої моделі.

Ключові слова: 3D-принтер; адитивне виробництво; фотополімерний друк; регресійний аналіз; тестові моделі; математичний аналіз; дослідження.

Вступ

Нині 3D-друк стає все більш поширеним і в деяких сферах замінює інші технології. Це пояснюється низкою переваг, які має ця технологія перед класичними. Більш низькі вимоги до кваліфікації персоналу й дешеве та доступне обладнання забезпечили 3D-друку популярність у світі. Безумовно, технології 3D-друку існують дуже давно, і вони відрізняються одна від одної, але водночас це різноманіття дає змогу використовувати їх у широкому спектрі напрямів – від виробництва машин до ливарної промисловості.

Однією з багатьох технологій 3D-друку є фотополімерний друк [1]. Він дає змогу друкувати рідким фотополімером, що під дією ультрафіолетового випромінювання полімеризується з рідкого стану у твердий. Особливістю фотополімерного друку є висока роздільна здатність, якщо порівнювати з більшістю інших технологій, зокрема *FDM, LOM, 3DP, EBF*. Якщо порівнювати фотополімерний друк з екструзійним, як найбільш поширеним, то по осі *Z* фотополімерний друк мінімально має 30 мкм, тоді як технології *FDM* – 50 мкм. Роздільна здатність по осях *XY* у фотополімерному друці становить мінімально 25 мкм, що відповідає роздільній здатності трафаретного екрана, крізь який проходить ультрафіолетове випромінювання. У технології *FDM* роздільна здатність становить мінімально 100 мкм по осях *XY*, що відповідає розміру сопла, крізь яке проходить розігрітий філамент. Тобто, якщо порівнювати з *FDM*-технологією, фотополімерна технологія забезпечує роздільну здатність більш ніж удвічі, маючи 50 мкм роздільну здатність по осях *XY*, що робить цю технологією однією із найбільш високоточних технологій 3D-друку.

Однак для отримання бажаної якості виробів, надрукованих за допомогою фотополімерної технології, необхідно встановити технологічні параметри, за яких буде досягнута максимальна якість. Основними параметрами технології фотополімерного друку, від яких залежить якість надрукованої деталі,

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

є висота шару й час полімерізації. За умови збільшення висоти шару деталь набуває шорсткості відповідно до товщини шару, водночас час експонування стає тривалішим. Збільшуючи або зменшуючи час експонування, деталь змінює свої геометричні розміри, згладжуючи цим тонкі елементи в процесі збільшення часу експонування або втрачаючи їх у разі зменшення. Іншими параметрами, що впливають на вихідну якість, є час експонування базових шарів, висота перемішування та швидкість. Отже, постає проблема у визначенні залежності між технологічними параметрами та встановлення їх впливу на кінцеву якість надрукованих деталей.

Аналіз останніх досліджень і публікацій

Фотополімерний друк поширений у різних сферах промисловості – від машинобудування до виробництва зубних протезів у медицині. Розвиток технологій виробництва електронних засобів приводить до появи нових *LCD*-екранів, що змінює параметри фотополімерного 3D-друку. Екрани мають більшу роздільну здатність, перехід від кольорових пікселів до монохромних дав змогу скоротити час друку та подовжити термін експлуатації самих екранів. Завдяки цьому збільшується якість отримуваних виробів і зменшується час на їх друк. Велике різноманіття фотополімерів, що використовуються у друці, відкриває нові перспективи для їх застосування. Основним напрямом досліджень у фотополімерному 3D-друці є розширення сфер упровадження та вдосконалення наявних методів і засобів друку фотополімерними смолами. Найбільш відомим для визначення технологічних параметрів є емпіричний метод із вимірюванням, що дозволяє встановити необхідні технологічні параметри для досягнення якісних показників. Недоліком цього методу є необхідність високої кількості друку тестових моделей, значні витрати матеріалу й часу.

З аналізу статей можна зробити висновок, що тема фотополімерного друку не є поширеною, наявні дослідження дають лише загальний огляд 3D-друку, не пропонуючи методів і засобів удосконалення цієї технології.

Постановка завдань і мета дослідження

Метою дослідження є аналіз наявних тестових моделей та елементів контролю, які вони містять, для визначення показників якості та впливу на них технологічних параметрів фотополімерного 3D-друку. Для виконання окресленої мети необхідно:

– визначити основні технологічні параметри фотополімерного друку;

– дослідити вплив технологічних параметрів на появу дефектів друку;

– побудувати регресійну модель впливу технологічних параметрів на якість друкованих виробів.

Кінцевим результатом є досягнення емпіричних показників, на основі яких будується кореляційно-регресійна модель впливу технологічних параметрів друку (часу експонування) на показники якості, такі як відповідність геометрії та відхилення розмірів.

Дослідження та аналіз досягнутих результатів

Технологія фотополімерного друку основана на пошаровому експонуванні світлочутливого фотополімеру, що під дією ультрафіолетового випромінювання полімеризується з рідкого стану у твердий.

Фотополімерна технологія має три основні підтипи, що відрізняються конструктивними особливостями принтерів. Перший підтип *SLA* [7] джерелом ультрафіолетового (УФ) випромінювання використовує лазер, що одночасно полімеризує фотополімер і формує горизонтальний переріз моделі за допомогою системи відхилення лазерного променя – сканатора (рис. 1).

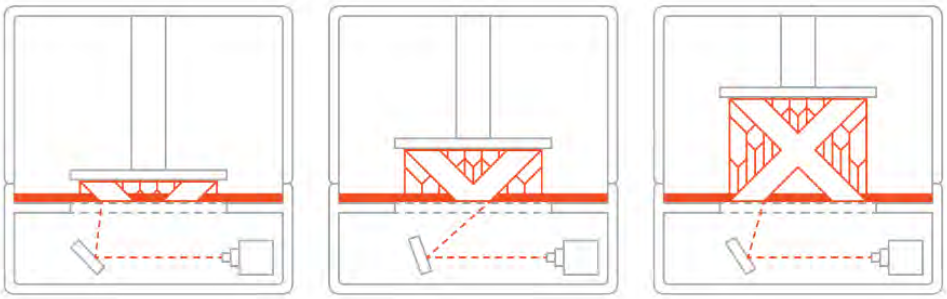


Рис. 1. Технологія фотополімерного друку *SLA*

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Другий підтип фотополімерної технології – *DLP* [8] – для фотополімеризації застосовує проєктор з *DLP*-матрицею, що затемнює окремі елементи, крізь які не проходить УФ-випромінювання, формує та полімеризує шар одночасно по всьому перерізу моделі (рис. 2).

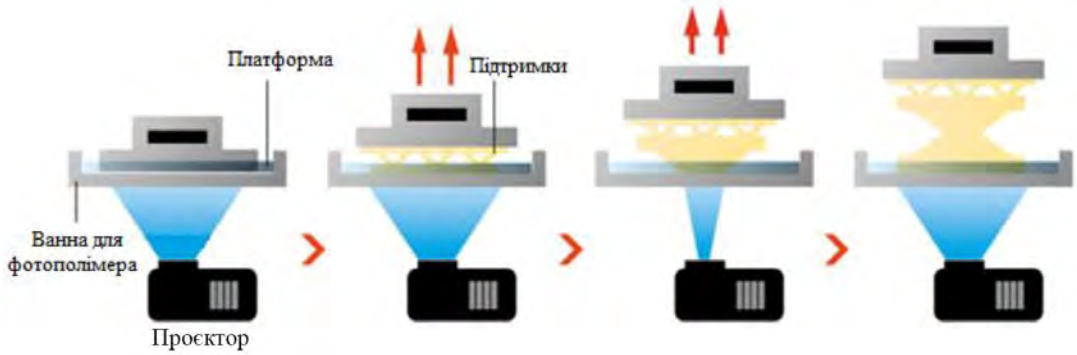


Рис. 2. Технологія фотополімерного друку *DLP*

Останнім підтипом фотополімерної технології є *LCD* [5], що містить УФ світлодіодну матрицю та маскувальний екран, який створює горизонтальний переріз моделі. Цей переріз, як і в технології *DLP*, має прозорі та непрозорі елементи. Прозорі пікселі здатні пропускати крізь себе УФ-випромінювання, що утворено світлодіодною матрицею. Непрозорі пікселі не пропускають це випромінювання, завдяки чому чорно-білий горизонтальний переріз здатний формувати шари моделі (рис. 3).

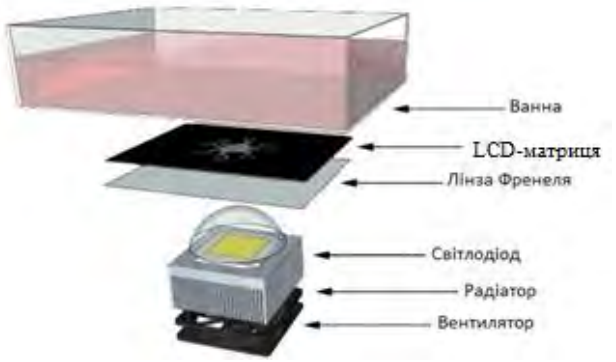


Рис. 3. Технологія фотополімерного друку *LCD*

Фотополімерний *LCD* 3D-принтер містить джерело УФ-випромінювання – світлодіодну матрицю, екран, ванну з прозорим і гнучким дном і платформу, що рухається по осі *Z*.

За вертикальну роздільну здатність принтера відповідає вісь *Z*, на яку кріпиться платформа. Загалом роздільна здатність по вертикалі становить від 30 мкм. За роздільну здатність по осях *XY* відповідає екран. Мінімально можливе значення роздільної здатності відповідає розміру пікселя екрана. За умови роздільної здатності екрана 1620x2560 розмір пікселя становить 40 мкм.

Отже, будова принтера визначає технологічні параметри, від яких залежить надрукована модель.

Першим технологічним параметром є час експонування. Цей параметр поділяється на час експонування базових шарів, що забезпечують надійне з'єднання перших шарів моделі з платформою, та час експонування звичайних шарів. Зазначені параметри залежать від трьох факторів: фотополімер, система експонування принтера та екран. Фотополімер – рідина, в основі якої містяться різні хімічні речовини, які мають різну чутливість до УФ-випромінювання, а також може містити різні домішки, що додають різних властивостей фотополімеру та змінюють вплив на нього УФ-випромінювання. Серед таких домішків можуть бути речовини, що додають еластичності надрукованій моделі, або домішки, що зроблять її міцнішою. Барвники також здатні впливати на час експонування, що потребує визначення цього часу для кожного фотополімеру окремо.

Системи засвітлення в принтерах відрізняються між собою за типом використаної матриці та системою випрямлення, яка необхідна для того, щоб УФ-випромінювання потрапляло на екран рівномірно. Матриці можуть містити або один світлодіод, що має лінзу Фрінеля для рівномірного засвітлення всієї площі екрана, або світлодіодну матрицю з використанням світлодіодів із малим кутом розсіювання світла, або і те, і те – світлодіодну матрицю та лінзу Фрінеля, забезпечуючи потужне й рівномірне засвітлення (рис. 4).

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Рис. 4. Світлодіодна матриця з лінзами Френеля

Останнім чинником, що впливає на час експонування, є екран. Екрани у фотополімерній технології розрізняють двох типів. Перший тип має кольорові пікселі, унаслідок чого УФ-випромінювання затримується на поверхні екрана й не проходить далі. Другий тип екрана монохромної структури, і піксель має лише два кольори – білий і чорний. Також існує можливість створення градієнта сірого. Зазначений тип екрана менше затримує УФ-випромінювання, завдяки чому час експонування може зменшуватися вчетверо-уп’ятеро. Для прикладу можна навести час експонування для звичайних екранів, що становить 12 с, тоді як принтер, що має аналогічну систему засвітлення та використовує той самий фотополімер із монохромним екраном, має час експонування 3 с.

Наступним параметром є висота шару, що регулює отриману шорсткість моделі (рис. 5). Цей параметр також поділяється на висоту шару базових шарів і на висоту шару звичайних шарів. Що більший розмір шару, то більшою буде шорсткість. У цьому разі час експонування збільшується.

Рис. 5. Моделі із шарами 25 мкм, 50 мкм та 100 мкм

Іншими параметрами, що забезпечують процес переходу між шарами, є висота та швидкість перемішування. Цей процес забезпечує рух фотополімеру на дні ванни, що не дає осаджуватися пігментам у фотополімері на дно. За низьких рівнів фотополімеру у ванній забезпечується покриття всієї площі дна ємності фотополімером. Іншим призначенням перемішування є відрив моделі від плівки дна ванни. Через високу еластичність плівки та неповну відсутність адгезії до неї між моделлю та плівкою виникає вакуум, що ускладнює відрив моделі. За відсутності перемішування перехід на наступний шар може не здійснюватися, оскільки модель залишиться у зчепленні з плівкою.

Останнім параметром є швидкість друку, що зі свого боку поділяється на швидкість друку базових шарів і звичайних. Цим параметром регулюємо швидкість перемішування, підйом після шару та його повернення на наступний шар. Швидкий відрив моделі від плівки може спричинити розшарування, коли частина моделі або тонкі елементи моделі залишаються на плівці.

Керування принтером відбувається з використанням спеціальної програми – *G-code*, для створення якої застосовується *CAM*-система – слайсер [9]. Це програма перетворює модель на сукупність графічних зображень горизонтальних перерізів моделі (рис. 6), перехід між якими виконується за допомогою команд *G-code*.

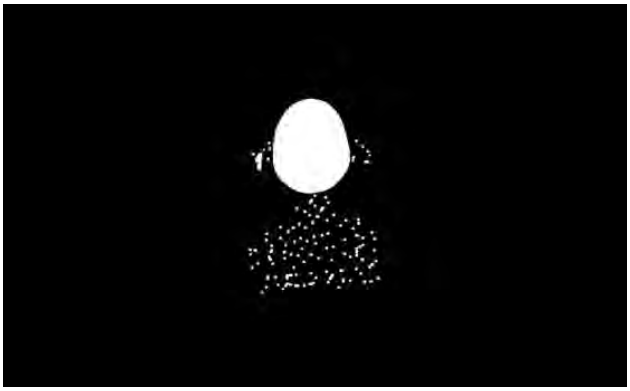


Рис. 6. Горизонтальний переріз моделі, що виводиться на екран

Ця програма налаштовує всі технологічні параметри, зазначені вище (рис. 7).

Крім того, дає змогу маніпулювати моделлю, її розмірами та орієнтацією, а також застосовувати підтримки – друковані елементи, що встановлюються під навислі частини моделі (рис. 8).

Для визначення впливу технологічних параметрів на вихідні показники якості друкованої моделі у фотополімерному 3D-друці використовуються тестові моделі, що дають змогу побудувати модель керування якісними показниками. Унаслідок аналізу

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

надрукованих тестових моделей можна встановити залежність між технологічними параметрами та їх вплив на показники якості, що визначаються виникненням дефектів.

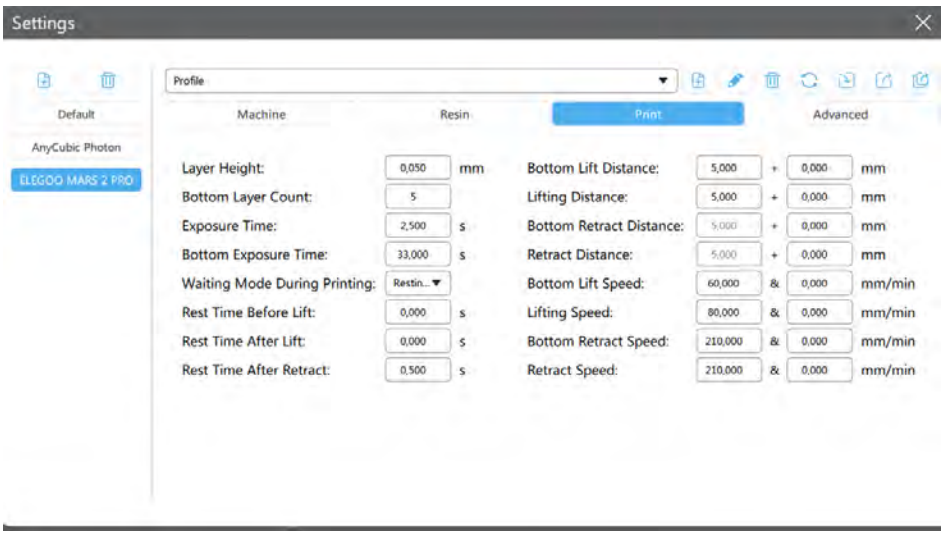


Рис. 7. Вікно параметрів слайсера

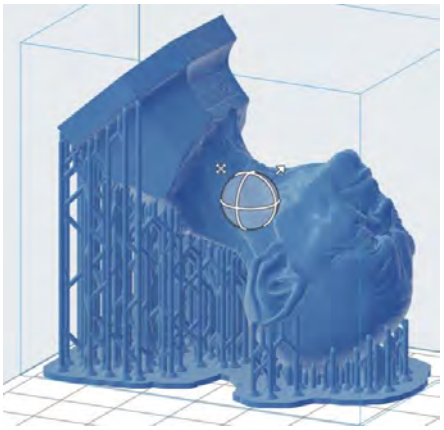


Рис. 8. Налаштована модель із використанням підтримок

Нині існує значна кількість тестових моделей, призначених для визначення основних технологічних параметрів: часу експонування, висоти шару, часу експонування базових шарів, висоти підйому та швидкості.

Розробники тестових моделей намагаються брати до уваги можливість виникнення різноманітних дефектів, щоб створити модель, яка буде очікувано змінюватися в разі зміни технологічних параметрів друку. В одній моделі зазвичай містяться декілька окремих тестів, що показують окремі дефекти та відхилення технологічних параметрів.

Тестові моделі зображені на рис. 9.

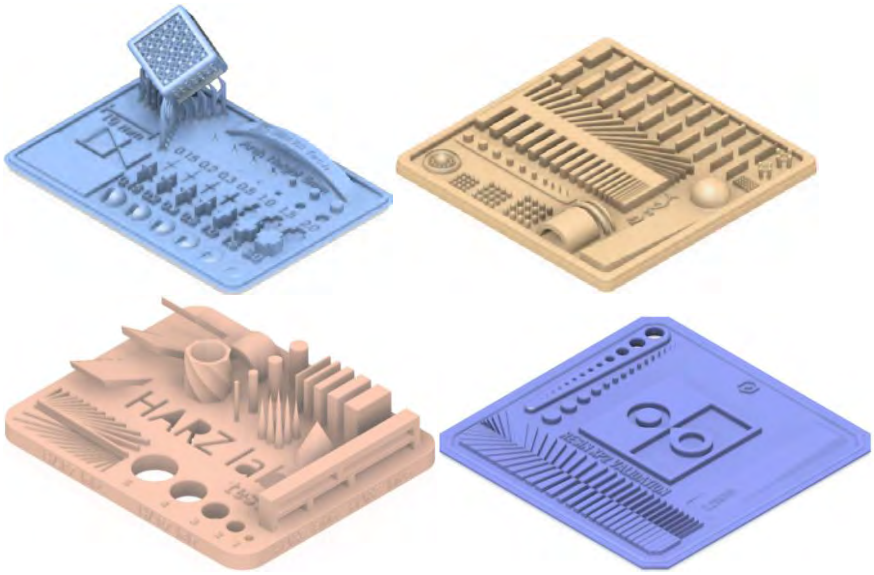


Рис. 9. Тестові моделі, за якими визначаються технологічні параметри фотополімерного друку

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Для встановлення впливу технологічних параметрів на результати друку обрано одну тестову модель, що може очікувано змінити якісні показники, на основі чого буде побудовано модель регресійного аналізу.

Тестовою була модель від *Syraya Tech*, що містить п'ять тестових елементів, які змінюються унаслідок заданої відстані (рис. 10).

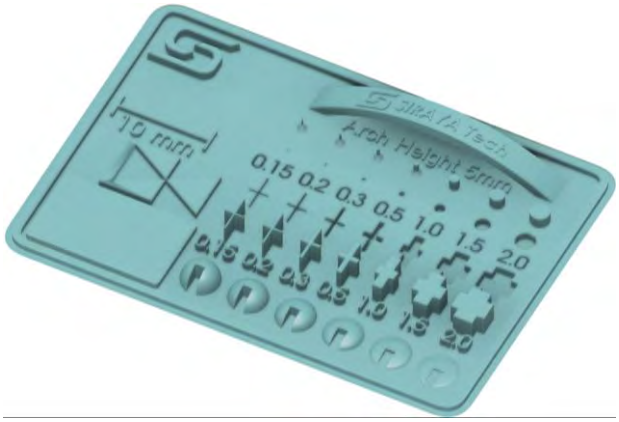


Рис. 10. Тестова модель *Syraya Tech*

Тестова модель з розмірами 48,72x34,41x5,24 має сім циліндрів і сім отворів, що змінюють свої розміри від 0,15 мм до 2 мм. Наступним елементом є набір семи хрестоподібних елементів зі стінками завтовшки від 0,15 мм до 2 мм та набір сфер заввишки від 0,15 мм до 2 мм. Додатковими елементами є арка заввишки 5 мм.

Застосовуючи цю модель, можна точно визначити такі технологічні параметри:

– час експонування базових шарів;
– час експонування звичайних шарів;
– товщину базових шарів;
– товщину звичайних шарів;
– висоту перемішування;
– швидкість переміщення осі *Z*.

Для побудови математичної моделі надруковано сім тестових моделей із різним часом експонування (рис. 11).

Рис. 11. Надрукована тестова модель

Час експонування шарів змінюється з кроком 0,25 с, від 1,5 с до 3 с. Час базових шарів сталий і становить 33 с. Підйом перемішування – 5 мм, швидкість перемішування – 80 мм/хв (рис. 12).

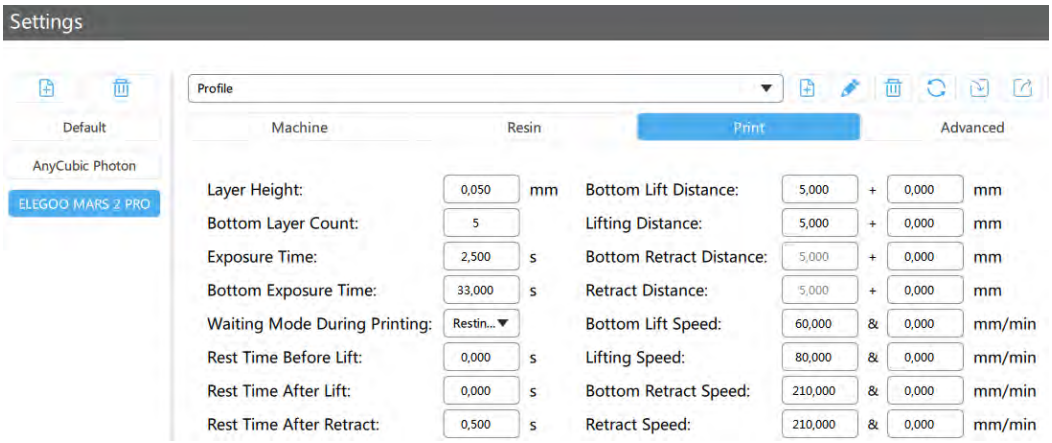


Рис. 12. Параметри друку тестових моделей

Тестові зразки друкувалися на модифікованому принтері *Anycubic Photon*, на який встановили монохромний екран і який має систему засвітлення з однією світлодіодною матрицею та розтрубом (рис. 13).

Унаслідок друку отримано сім тестових моделей з різним часом експонування. Для визначення результатів використовується штангенциркуль із точністю ± 30 мкм. Результати подані в табл. 1. На наступному етапі дослідження встановлено,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

що визначити точний розмір менших елементів неможливо, тому для уточнення даних із семи тестових елементів (циліндрів, отворів) було взято два найбільших за розміром тестових елементи розміром 2 мм та 1,5 мм.

З огляду на показники таблиці можна встановити, що в разі зміни часу експонування змінюються геометричні розміри моделі. У тестових моделях із найкоротшим часом експонування візуально можна побачити дефекти на моделі, що мають вигляд відсутності найменших елементів (рис. 14).

Рис. 13. Фотополімерний принтер *Anycubic Photon*

Таблиця 1. *Результати експериментальних досліджень*

Тестовий зразок	Час експонування шарів, секунди	Тестовий елемент: циліндри, мм		Тестовий елемент: хрестоподібні елементи, мм		Тестовий елемент: циліндричні отвори, мм		Тестовий елемент: хрестоподібні отвори, мм	
		2 мм	1.5 мм	2 мм	1.5 мм	2 мм	1.5 мм	2 мм	1.5 мм
1	1,50	1,87	1,40	1,84	1,41	1,73	1,55	2,12	1,67
2	1,75	1,95	1,45	1,90	1,47	1,80	1,53	1,95	1,54
3	2,00	2,00	1,48	2,07	1,55	1,81	1,54	1,98	1,55
4	2,25	2,10	1,61	2,23	1,63	1,70	1,31	1,93	1,44
5	2,50	2,19	1,63	2,20	1,78	1,26	0,50	1,85	1,30
6	2,75	2,11	1,66	2,18	1,80	1,47	0,69	1,87	1,41
7	3,00	2,30	1,73	2,39	1,77	1,61	1,16	1,72	1,35

Рис. 14. Тестова модель із дефектом

Для підтвердження візуальних спостережень необхідно провести математичний аналіз у вигляді лінійного однофакторного регресійного аналізу [10] з метою визначення залежності між часом експонування та розмірами елементів. Рівняння лінійної регресії має такий вигляд:

$$Y = b1 \cdot X + c. \quad (1)$$

Розрахунок моделі регресійного аналізу проведено в *EXCEL*, унаслідок чого досягнуто певних результатів для тестових циліндрів із діаметрами 2 мм та 1,5 мм (див. табл. 2).

Таблиця 2. *Результат регресійного аналізу залежності розміру тестових циліндрів від часу експонування*

SUMMARY OUTPUT		RESIDUAL OUTPUT						
Regression Statistics		*Observation*	*Predicted Y*	*Residuals*				
Multiple R	0,98186815	1	1,530753	-0,03075				
R Square	0,96406506	2	1,751107	-0,00111				
Adjusted R Square	0,94609758	3	1,883683	0,116317				
Standard Error	0,12538556	4	2,43696	-0,18696				
Observations	7	5	2,535651	-0,03565				
		6	2,644582	0,105418				
ANOVA		7	2,967265	0,032735				
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	1,687113849	0,843557	53,65613	0,00129132			
Residual	4	0,062886151	0,015722					
Total	6	1,75						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	-4,5718651	0,741585605	-6,16499	0,003514	-6,630836798	-2,512893	-6,6308368	-2,5128934
X Variable 1(тестовий циліндр 2.0 мм)	0,18188817	1,261614613	0,144171	0,892337	-3,320915544	3,6846919	-3,3209155	3,68469189
X Variable 2(тестовий циліндр 1.5 мм)	4,11606198	1,508825415	2,727991	0,052552	-0,07310896	8,3052329	-0,073109	8,30523291

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

За результатами регресійного аналізу коефіцієнт детермінації становить 96%, що підтверджує залежність отриманих розмірів від часу експонування. Коефіцієнт *X Variable 1* відповідає тестовому циліндру з діаметром 2 мм, *X Variable 2* відповідає тестовому циліндру з діаметром 1,5 мм. Також отримано графіки підбору та графіки залишків (рис. 15).

Далі проаналізовано залежність між часом експонування та хрестоподібними елементами зі стінками завтовшки 2 мм та 1,5 мм (див. табл. 3).

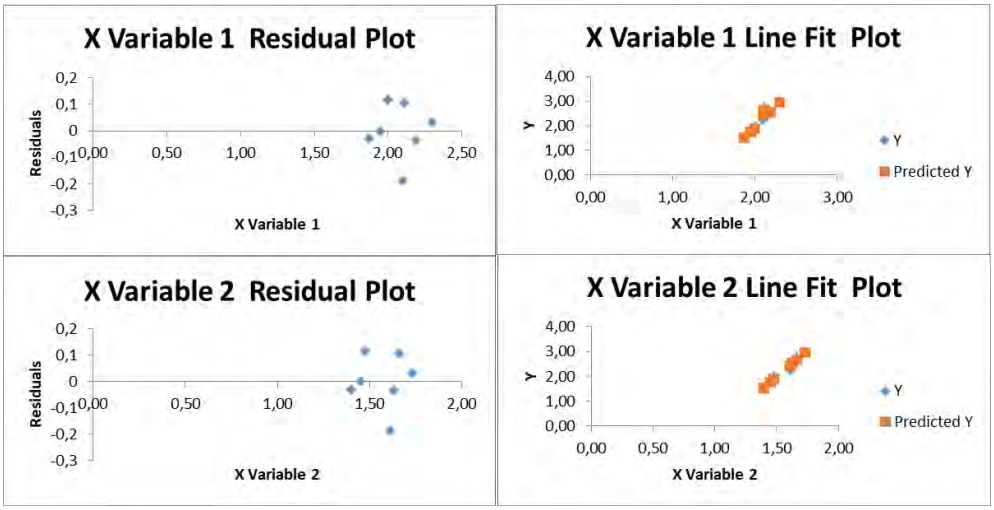


Рис. 15. Графіки підбору та графіки залишків

Таблиця 3. *Результат регресійного аналізу залежності розміру хрестоподібних елементів із стінками завтовшки 2 мм та 1,5 мм від часу експонування*

SUMMARY OUTPUT		RESIDUAL OUTPUT						
Regression Statistics		*Observation*	*Predicted Y*	*Residuals*				
Multiple R	0,9780182	1	1,485568	0,0144317				
R Square	0,9565196	2	1,676426	0,0735741				
Adjusted R Square	0,9347795	3	2,03529	-0,03529				
Standard Error	0,1379226	4	2,382555	-0,132555				
Observations	7	5	2,650925	-0,150925				
		6	2,66815	0,0818499				
ANOVA		7	2,851086	0,1489138				
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	1,673909384	0,836955	43,997788	0,001890541			
Residual	4	0,076090616	0,019023					
Total	6	1,75						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	-3,4983227	0,623878319	-5,60738	0,0049684	-5,23048665	-1,766159	-5,23048665	-1,7661588
X Variable 1(хрестоподібний елемент 2.0 мм)	1,1598539	0,596917767	1,943072	0,1239393	-0,49745547	2,817163	-0,49745547	2,8171634
X Variable 2(хрестоподібний елемент 1.5 мм)	2,0211062	0,726342904	2,782579	0,0496873	0,004455049	4,037757	0,00445505	4,0377574

Унаслідок аналізу досягнуто показників, що також підтверджують залежність розмірів від часу експонування, коефіцієнт детермінації становить 95,6%, що повторює результат попереднього регресійного аналізу. У цьому моделюванні коефіцієнт *X Variable 1* відповідає хрестоподібному елементу зі стінкою завтовшки 2 мм, *X Variable 2* відповідає хрестоподібному елементу зі стінкою завтовшки 1,5 мм. Крім того, отримані графіки залишків і підбору (рис. 16).

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

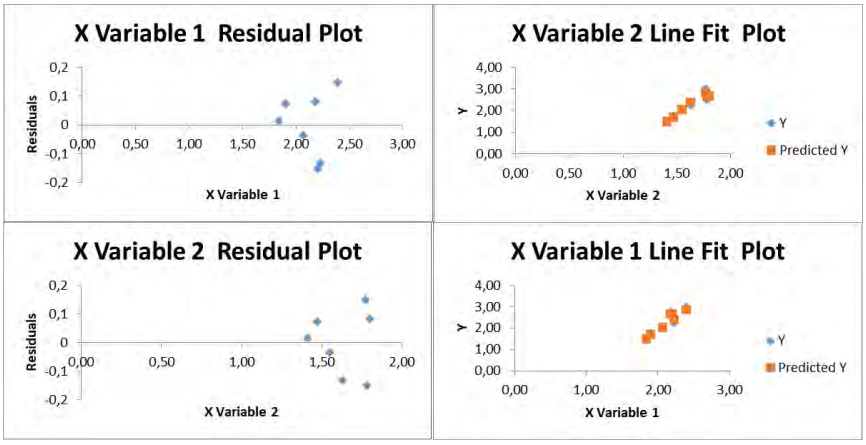


Рис. 16. Графіки підбору та графіки залишків для хрестоподібних елементів

Висновки

Під час досліджень встановлено та підтверджено, що час експонування у фотополімерному друці є ключовим технологічним параметром, що впливає на основні показники якості отриманої моделі, тобто на геометричні розміри. За умови скорочення часу, необхідного на експонування шару, геометричні розміри зменшуються, спричиняються дефекти, за яких тонкі елементи відсутні. У разі зростання часу експонування елементи збільшуються, що також може призводити до дефектів, коли отвори, що є в моделі, запливають фотополімерною смолою та зникають.

Унаслідок друку семи тестових елементів із часом від 1,5 с до 3 с визначено час експонування, що забезпечує найбільше збереження встановлених геометричних розмірів на 2 с, де відхилення від розмірів становить до 50 мкм. Це є кращим результатом, якщо порівнювати з іншими тестовими зразками.

Отже, із застосуванням математичного моделювання в поєднанні з друком тестових моделей на фотополімерному 3D-принтері можна отримати прогнозовані геометричні розміри з мінімальним відхиленням від заданих, що дасть змогу в кінці друку досягти найбільш точного результату та мінімізувати виникнення різноманітних дефектів, на які впливає встановлений час експонування шарів.

Надалі існує необхідність проведення подібного математичного аналізу з використанням тестових моделей для встановлення залежності висоти шару, часу експонування та показників якості. Проаналізувавши залежність між технологічними параметрами та показниками якості, можна визначити метод оптимізації технологічних параметрів зі скороченням витрат матеріалів, часу та досягнення кращих показників якості.

Список літератури

1. Технології фотополімерного 3D-друку: опис, переваги та недоліки. URL: https://monofilament.com.ua/ua/blog-novini-3d-druku-ta-additivnih-tehnologij/sla-dlp-lcd-tehnologiji-fotopolimernogo-3d-druku-opis-perevagi-ta-nedoliki. (дата звернення: 28.05.2024).
2. Адитивні технології: перспективи і проблеми 3D-друку. URL: https://nti.ukrintei.ua/wp-content/uploads/2018/05/2017-1_stat9_UA_povn.pdf (дата звернення: 04.06.2024).
3. Rajat Chaudhary, Paride Fabbri, Enrico Leoni, Francesca Mazzanti, Raziyeh Akbari, Carlo Antonini. Additive manufacturing by digital light processing: a review. *Progress in Additive Manufacturing* (2023) 8. P. 331–351. DOI: https://doi.org/10.1007/s40964-022-00336-0
4. Ishak Ertugrul, Osman Ulkir, Sezgin Ersoy, Minvydas Ragulskis. Additive Manufactured Strain Sensor Using Stereolithography Method with Photopolymer Material. *Polymers* 15(4). 991p. DOI: 10.3390/polym15040991
5. Ian Gibson, David Rosen, Brent Stucker. Additive Manufacturing Technologies 3D Printing, Rapid Prototyping, and Direct Digital Manufacturing. Second Edition. *Springer*. 2015. 510 p. DOI 10.1007/978-1-4939-2113-3
6. Ben Redwood, Filemon Schöffer, Brian Garret. The 3D Printing Handbook. Technologies, design and applications. *3D Hubs*. 2017. 347 p.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

7. Стереолітографія (Laser Stereolithography, SLA). URL: https://pro3d.com.ua/a367313-stereolitografiya-laser-stereolithography.html (дата звернення: 28.05.2024).
8. Що таке DLP 3D друк. URL: https://pro3d.com.ua/a393155-scho-take-dlp.html (дата звернення: 28.05.2024).
9. Що таке LCD 3D-принтер та де він використовується? URL: https://www.0332.ua/list/471643 (дата звернення: 28.05.2024).
10. Слайсери для 3d-друку. URL: https://makerhub.org/slicers-3d-print/ (дата звернення: 28.05.2024).
11. Anycubic Photon LCD 3D принтер. URL: https://3dreams.com.ua/ua/product/anycubic-photon-lcd-3d-принтер/ (дата звернення: 28.05.2024).
12. Регресійний аналіз. URL: https://ukrayinska.libretexts.org/Статистика/Прикладна_статистика/Книга%3A_Статистика_бізнесу_(OpenStax)/13%3A_Лінійна_регресія_та_кореляція/13.04%3A_Рівняння_регресії (дата звернення: 28.05.2024).
13. Основи кореляційного та регресійного аналізу. URL: https://pchilka-litsei.in.ua/excel-book/basis_analysis.html (дата звернення: 29.05.2024).
14. Метод регресійного аналізу в MS Excel. URL: https://modeling.at.ua/publ/10-1-0-58 (дата звернення: 29.05.2024).
15. Нікітін Д.О., Стрілець Р.Є., Близнюк Д.С. Сравнительный анализ технологий 3D прототипирования SLA, DLP и LCD. Разработка автоматизированной станции для 3D печати. *VII Міжнародна науково-технічна Internet-конференція «Сучасні методи, інформаційне, програмне та технічне забезпечення систем керування організаційно-технічними та технологічними комплексами» (НУХТ)*. 2020 С. 55–56. URL: https://dspace.nuft.edu.ua/bitstreams/934a9612-ecc5-4a3a-8377-320f864dac10/download
16. Igor Nevlyudov, Ievgenii Razumov-Fryziuk, Dmytro Nikitin, Danylo Blyzniuk, Roman Strelets. Cost Estimation of Photopolymer Resin for 3D Exposure of Circuit Boards. *Technology Audit and Production Reserves* № 2/2(64), 2022. P.43-49. DOI: 10.15587/2706-5448.2022.256538
17. I. Nevliudov, E. Razumov-Fryzyuk, D. Nikitin, D. Bliznyuk, R. Strelets. Technology for creating the topology of printed circuit boards using polymer 3D masks. *Сучасний стан наукових досліджень та технологій в промисловості* № 1 (15). 2021 С.120–131. DOI: https://doi.org/10.30837/ITSSI.2021.15.120
18. Redwood B. 2020. 3D Printing. Practical guide. Book Workshop. 2022 [Redvud B. 2020. 3D-druk. Praktychnyi posibnyk. Knyzhkova Maisternia] 2022
19. Стереолітографія – що потрібно знати про технологію. URL: https://www.3dprinter.ua/stereolitografiya-shho-potribno-znaty-pro-tehnologiyu/ (дата звернення: 29.05.2024).
20. Методи і техніка досліджень. URL: https://elib.tsatu.edu.ua/dep/mtf/ophv_10/page3.html (дата звернення: 29.05.2024).
21. Методологія і організація наукових досліджень URL: https://shron1.chtyvo.org.ua/Burhu_Yurii/Metodolohiia_i_orhanizatsiia_naukovykh_doslidzhen.pdf (дата звернення: 29.05.2024).
22. Основи наукових досліджень. URL: https://core.ac.uk/download/pdf/162019668.pdf (дата звернення: 29.05.2024).

References

1. Technologies of photopolymer 3D printing: description, advantages, and disadvantages [Tekhnolohii fotopolimernoho 3D-druku: opys, perevahy ta nedoliky]: available at: https://monofilament.com.ua/ua/blog-novini-3d-druku-ta-additivnih-tehnologij/sla-dlp-lcd-tehnologiji-fotopolimernogo-3d-druku-opis-perevagi-ta-nedoliki (last accessed: 28.05.2024).
2. Additive technology: prospects and challenges 3D-print [Adytyvni tekhnolohii: perspektyvy i problemy 3D-druku] available at: https://nti.ukrintei.ua/wp-content/uploads/2018/05/2017-1_stat9_UA_povn.pdf (last accessed: 04.06.2024).
3. Rajat Chaudhary, Paride Fabbri, Enrico Leoni, Francesca Mazzanti, Raziyeh Akbari, Carlo Antonini, "Additive manufacturing by digital light processing: a review". Progress in Additive Manufacturing (2023) 8. P. 331–351. DOI: https://doi.org/10.1007/s40964-022-00336-0
4. Ishak Ertugrul, Osman Ulkir, Sezgin Ersoy, Minvydas Ragulskis, "Additive Manufactured Strain Sensor Using Stereolithography Method with Photopolymer Material" Polymers 15(4). 991p. DOI: 10.3390/polym15040991
5. Ian Gibson, David Rosen, Brent Stucker. Additive Manufacturing Technologies 3D Printing, Rapid Prototyping, and Direct Digital Manufacturing. Second Edition. Springer. 2015. 510 p. DOI: 10.1007/978-1-4939-2113-3
6. Ben Redwood, Filemon Schöffer, Brian Garret. The 3D Printing Handbook. Technologies, design and applications. 3D Hubs. 2017. 347 p.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

7. Stereolithography (Laser Stereolithography, SLA) [Stereolitohrafiia (Laser Stereolithography, SLA)]: available at: https://pro3d.com.ua/a367313-stereolitografiya-laser-stereolithography.html (last accessed: 28.05.2024).
8. What is DLP 3D printing? [Shcho take DLP 3D druk]: available at: https://pro3d.com.ua/a393155-scho-take-dlp.html (last accessed: 28.05.2024).
9. What is an LCD 3D printer and where is it used? [Shcho take LCD 3D-prynter ta de vin vykorystovuietsia?]: available at: https://www.0332.ua/list/471643 (last accessed: 28.05.2024).
10. Slicers for 3d printing [Slaisery dlia 3d-druku]: available at: https://makerhub.org/slicers-3d-print/ (last accessed 28.05.2024).
11. Anycubic Photon LCD 3D printer [Anycubic Photon LCD 3D prynter]: available at: https://3dreams.com.ua/ua/product/anycubic-photon-lcd-3d-принтер/ (last accessed: 28.05.2024).
12. Regression analysis [Rehresiinyi analiz]: available at: https://ukrayinska.libretexts.org/Статистика/Прикладна_статистика/Книга%3A_Статистика_бізнесу_(OpenStax)/13%3A_Лінійна_регресія_та_кореляція/13.04%3A_Рівняння_регресії (last accessed: 28.05.2024).
13. Basics of correlation and regression analysis [Osnovy koreliatsiinoho ta rehresiinoho analizu]: available at: https://pchilka-litsei.in.ua/excel-book/basis_analysis.html (last accessed: 29.05.2024).
14. The method of regression analysis in MS Excel [Metod rehresiinoho analizu v MS Excel]: available at: https://modeling.at.ua/publ/10-1-0-58 (last accessed: 29.05.2024).
15. Nikitin D.O., Strilets R.E., Blyzniuk D.S. "Porivnyalnil analyz tekhnolohyi 3D prototypyrovanyia SLA, DLP y LCD. Razrabotka avtomatyzyrovannoi stantsyy dlia 3D pechaty" VII Mizhnarodna naukovo-tekhnichna Internet-konferentsiia "Suchasni metody, informatsiine, prohramne ta tekhnichne zabezpechennia system keruvannia orhanizatsiino-tekhnichnymy ta tekhnolohichnymy kompleksamy" (NUKhT).2020 P. 55–56. available at: https://dspace.nuft.edu.ua/bitstreams/934a9612-ecc5-4a3a-8377-320f864dac10/download
16. Igor Nevlyudov, Ievgenii Razumov-Fryziuk, Dmytro Nikitin, Danylo Blyzniuk, Roman Strelets "Cost Estimation of Photopolymer Resin for 3D Exposure of Circuit Boards" Technology Audit and Production Reserves — № 2/2(64), 2022. P.43–49. DOI: 10.15587/2706-5448.2022.256538
17. I. Nevliudov, E. Razumov-Fryzyuk, D. Nikitin, D. Bliznyuk, R. Strelets "Technology for creating the topology of printed circuit boards using polymer 3d masks" Suchasnyi stan naukovykh doslidzhen ta tekhnolohii v promyslovosti № 1 (15). 2021 С. 120–131. DOI: https://doi.org/10.30837/ITSSI.2021.15.120
18. Redwood B. 2020. 3D Printing. Practical guide. Book Workshop. 2022 [Redvud B. 2020. 3D-druk. Praktychnyi posibnyk. Knyzhkova Maisternia] 2022.
19. Stereolithography – what you need to know about the technology [Stereolitohrafiia – shcho potribno znaty pro tekhnolohiiu]: available at: https://www.3dprinter.ua/stereolitografiya-shho-potribno-znaty-pro-tehnologiyu/ (last accessed: 29.05.2024).
20. Research methods and techniques [Metody i tekhnika doslidzhen]: available at: https://elib.tsatu.edu.ua/dep/mtf/ophv_10/page3.html (last accessed: 29.05.2024).
21. Methodology and organization of scientific research [Metodolohiya i orhanizatsiya naukovyh doslidzhen]: available at: https://shron1.chtyvo.org.ua/Burhu_Yurii/Metodolohiia_i_orhanizatsiia_naukovykh_doslidzhen.pdf (last accessed: 29.05.2024).
22. Basics of the scientific research [Osnovy naukovyh doslidzen]: available at: https://core.ac.uk/download/pdf/162019668.pdf (last accessed: 29.05.2024).

Надійшла 31.05.2024

Відомості про авторів / About the Authors

Невлюдов Ігор Шакирович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, завідувач кафедри комп'ютерно-інтегрованих технологій, автоматизації та мехатроніки, Харків, Україна; e-mail: igor.nevliudov@nure.ua; ORCID ID: https://orcid.org/0000-0002-9837-2309

Стрілець Роман Євгенійович – Харківський національний університет радіоелектроніки, аспірант, Харків, Україна; e-mail: roman.strilets@nure.ua; ORCID ID: https://orcid.org/0000-0001-5123-8703

Близнюк Данило Сергійович – Харківський національний університет радіоелектроніки, аспірант, Харків, Україна; e-mail: danylo.blyzniuk@nure.ua; ORCID ID: https://orcid.org/0000-0002-3041-1885

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Nevliudov Igor – Doctor of Sciences (Engineering), Professor, Kharkiv National University of Radioelectronics, Head at the Department of Computer-integrated Technologies, Automation and Mechatronics, Kharkiv, Ukraine.

Strilets Roman – Kharkiv National University of Radioelectronics, PhD Student, Kharkiv, Ukraine.

Blyzniuk Danylo – Kharkiv National University of Radioelectronics, PhD Student, Kharkiv, Ukraine.

ENSURING QUALITY INDICATORS OF PHOTOPOLYMER 3D PRINTING BY USING MATHEMATICAL MODELING AND TEST MODELS

The subject of the research in the article is the analysis of the influence of technological parameters of photopolymer printing on the appearance of defects in the printing process. Test models for analysis are used, which contain elements that are affected by changes in technological parameters. **The purpose** of the work is to determine the dependence between the technological parameters of photopolymer printing and the defects that arise because of printing using models for testing. The article addresses the following **tasks**: analysis of existing test models and determination of model elements and the influence of technological parameters on them. The methods used for the research are mathematical analysis in the form of univariate linear regression and the empirical method. This method consists in comparing and measuring the difference between individual test samples to obtain values that will later be used in regression analysis. The following **results** were obtained: the dependence of technological factors and their influence on the elements of the test model was determined, which consists in changing the physical dimensions of the test models. With insufficient exposure time, the dimensions of the model are reduced, and defects are created. When the exposure time increases, the dimensions of the model increase linearly, defects appear in the form of holes disappearing or their sizes changing. **Conclusions**: because of an experimental study, which consists in printing test models and their analysis with the help of linear one-factor regression, the dependence between the exposure time and the physical dimensions of the model was determined and confirmed. Then a method of studying the correspondence of dimensions depending on the exposure time using test models was proposed and mathematical modeling in the form of regression univariate analysis. In the following, it is proposed to determine the influence of the layer height on the exposure time. Further, it is proposed to determine the significance of individual technological factors among themselves and their influence on the defects obtained because of printing. Building a regression analysis model, determining the correlation of technological parameters and their influence on quality indicators. Determination of the coefficient of determination of the constructed model.

Keywords: 3D printer; additive manufacturing; photopolymer printing; regression analysis; test models; mathematical analysis; research.

Бібліографічні описи / Bibliographic descriptions

Невлюдов І. Ш., Стрілець Р. Є., Близнюк Д. С. Забезпечення якісних показників фотополімерного 3D-друку за допомогою математичного моделювання і тестових моделей. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 96–107. DOI: https://doi.org/10.30837/2522-9818.2024.2.096

Nevliudov, I., Strilets, R., Blyzniuk, D. (2024), "Ensuring quality indicators of photopolymer 3D printing by using mathematical modeling and test models", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 96–107. DOI: https://doi.org/10.30837/2522-9818.2024.2.096

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

УДК 004.89 DOI: https://doi.org/10.30837/2522-9818.2024.2.108

А. Новаковський, І. Яловега

УПРОВАДЖЕННЯ ТЕХНОЛОГІЙ ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ В ТВОРЧУ ДІЯЛЬНІСТЬ: РОЗРОБЛЕННЯ СТРУКТУРНОЇ МОДЕЛІ ДИЗАЙН-МИСЛЕННЯ

Предметом дослідження є системні зміни в методології дизайн-мислення, що відбуваються під впливом розвитку та поширення технологій генеративного штучного інтелекту (ШІ) в дизайні та інших креативних індустріях. **Метою роботи** є: аналіз сучасних досліджень щодо впливу технологій генеративного ШІ на креативні індустрії, дизайн, зокрема, на дизайн-мислення; розроблення структурної моделі дизайн-мислення для подальшого дослідження еволюції методології. У статті визначені такі **завдання**: проаналізувати сучасні наукові публікації щодо сутності, структури та змістовного наповнення дизайн-мислення; розглянути дослідження щодо переваг та викликів застосування генеративного ШІ у процесах дизайну; розробити модель, що дасть змогу ідентифікувати та описати зміни в ключових компонентах методології дизайн-мислення, які виникають під впливом широкого впровадження технологій генеративного ШІ. Під час дослідження використані такі **методи**: аналіз і синтез змісту технічних, економічних, філософських, лінгвістичних, історичних та методичних досліджень щодо проблем формування понятійного апарату методології дизайн-мислення та застосування генеративного ШІ у процесах дизайну; порівняльно-історичний, ретроспективний методи; структурно-логічний аналіз. Досягнуто таких **результатів**: актуалізована потреба в комплексному дослідницькому підході для аналізу багатогранного впливу технологій ШІ на дизайн; визначено ключові переваги та виклики, пов'язані з інтеграцією ШІ в креативні процеси; розроблено структурну модель подання методології дизайн-мислення у вигляді чотирьох взаємопов'язаних структурних шарів із подальшою декомпозицією кожного з них на складники. У **висновках** наголошується на глибині та багатогранності змін, що відбуваються в дизайні та інших креативних індустріях під впливом генеративного ШІ та потребують подальших ґрунтовних досліджень. Розроблена структурна модель методології дизайн-мислення дає змогу до певної міри декомпозувати складний творчий процес, закладаючи основу для всебічного аналізу еволюції методології та системного впровадження технологій генеративного штучного інтелекту в процеси дизайну.

Ключові слова: методологія дизайн-мислення; генеративний штучний інтелект; інновації в дизайні; структурна модель; творча діяльність.

Вступ

Дизайн-мислення – ітеративна евристична методологія розроблення рішень складних проблем в умовах високої невизначеності з орієнтацією на потреби людини. Методологія набула значного поширення та визнання як у комерційній сфері, так і в науковому середовищі, зокрема як ефективний інструмент для створення інновацій. Евристичний підхід до розроблення рішень, властивий дизайн-мисленню, робить його потужним інструментом для розв'язання проблем, що виникають в складних організаційних, економічних і соціальних системах. Незважаючи на значну кількість досліджень [1–5], методологія все ще перебуває в процесі розвитку, осмислення та формалізації.

Гнучка, ітеративна та відкрита до нового природа дизайн-мислення знаходить, зокрема, вираження в постійній еволюції методології під впливом низки трендів, таких як технологічний прогрес, зростання значення ідей сталого розвитку та соціальної відповідальності, поширення нових форматів дистанційної та гібридної праці. Аналітичний звіт 2023 р. від *McKinsey & Company* – міжнародної консалтингової компанії, що спеціалізується на розв'язанні завдань, пов'язаних із стратегічним управлінням, – наголошує на значному трансформаційному впливі на креативні індустрії та дизайн, що викликаний бурхливим розвитком технологій генеративного штучного інтелекту [6].

Хоча сучасна концепція дизайн-мислення сформувалася відносно недавно, вона є природним наслідком тривалого філософського осмислення сутності творчості та її результатів (артефактів), яке започаткували мислителі Платон, Арістотель і Демокріт [7]. І. Яловега та С. Зуб наочно продемонстрували, що спроби зрозуміти природу творчості спостерігалися ще в глибоку давнину в працях відомих філософів, учених та інженерів різних галузей [1, 2].

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Аналіз досліджень і публікацій

Як показано в роботах У. Йоханссон-Шельдберг [3] та Дж. Ліедтка [4], можна виокремити два сучасних дискурси, в яких обговорюється дизайн-мислення: дизайн і менеджмент. Концепція почала формуватися в межах сучасної дискусії про теорію дизайну, початок якої традиційно припадає на кінець 1960-х рр., коли була опублікована робота Герберта А. Саймона "Наука про штучне". Упродовж 2000-х рр. дизайн-мислення почало поширюватися в літературі, присвяченій менеджменту та інноваціям, завдяки впливу таких мислителів, як Т. Браун і Р. Мартін. Отже, сучасна концепція дизайн-мислення пройшла еволюцію від підходу, орієнтованого на дизайн, до методології вирішення бізнес- та організаційних проблем [3–5].

Щоб уникнути потенційної плутанини, пов'язаної з назвою, необхідно наголосити на різниці між методологією "дизайн-мислення" та "мислення дизайнера". Хоча особливий спосіб мислення, притаманний дизайнерам, є важливим для практики "дизайн-мислення", це лише один з елементів методології. Крім того, нині дизайн-мислення широко впроваджується за межами професійного дизайну та зарекомендувало себе як ефективний метод пошуку рішень складних бізнес-проблем з орієнтацією на потреби людини.

Комп'ютерні технології значно трансформують дизайн протягом останніх десятиліть. У 1960-х рр. починається розвиток та застосування систем автоматизованого проєктування (САПР) [10]. Протягом 1980-х рр. з'являється програмне та апаратне забезпечення для комп'ютерної верстки, що дало змогу суттєво оптимізувати ресурсомісткі видавничі процеси [11]. Сучасні інструменти генеративного дизайну, наприклад *Autodesk Fusion 360*, використовують комп'ютерні алгоритми для швидкого створення множини рішень, що відповідають наперед заданому дизайнером набору параметрів та обмежень [14]. Протягом останнього року з'явилася низка публікацій [10–13], в яких досліджуються переваги та недоліки використання в процесі дизайну нової групи комп'ютерних технологій – генеративного штучного інтелекту, зокрема засобів на кшталт *ChatGPT* чи *DALL·E*. Дослідники наголошують на трансформаційному потенціалі генеративного штучного інтелекту, а також на значних супутніх викликах [10–13].

Мета роботи

Метою роботи є:

– аналіз сучасних досліджень щодо впливу технологій генеративного ШІ на креативні індустрії, дизайн та, зокрема, на дизайн-мислення;

– розроблення структурної моделі дизайн-мислення для подальшого дослідження еволюції методології.

Результати досліджень та їх обговорення

Огляд сучасних підходів до впровадження LLM у виробничі процеси

Ключова особливість технологій генеративного штучного інтелекту полягає в здатності створювати новий контент, такий як текст, зображення чи аудіо. Великі мовні моделі (*Large Language Model, LLM*), такі як *GPT, LLAMA, Gemini,* демонструють вражаючі можливості в обробленні та генерації тексту. Крім того, вони мають такі емерджентні властивості та здатності, як:

– ведення тривалого осмисленого діалогу з огляду на контекст;

– декомпозиція складних тверджень та задач;

– планування дій та узгодження їх між собою;

– генерування пропозиції;

– оцінювання опції, обрання та обґрунтування.

Хоча перелічені можливості перебувають у процесі розвитку й наразі мають певні обмеження, вони дають змогу *LLM* виконувати складні комунікативні, креативні та інтелектуальні функції, що до появи технологій генеративного ШІ була здатна виконувати тільки людина. Великі мовні моделі вже зараз допомагають виконувати широкий спектр робочих завдань – від простих, таких як пошук синонімів чи класифікація коментарів, до більш складних дій: написання історій, редагування текстів, генерація та організація ідей [8].

В основі технологій генеративного ШІ лежать підходи та методи глибокого машинного навчання (*deep learning*) багатошарових нейронних мереж безпосередньо із значної кількості інформації. Навчання не потребує попередньої розмітки даних чи зворотного зв'язку з боку людини. Аналізуючи значні обсяги інформації, нейронна мережа самостійно виокремлює специфічні патерни, закономірності та характеристики, притаманні певному типу інформації (наприклад, текст, зображення чи музика).

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Після завершення навчання нейронна мережа здатна ідентифікувати вивчені патерни, закономірності, характеристики та/або створювати нові масиви даних на їх основі. Важливо наголосити, що новий контент, згенерований нейронною мережею, не завжди може бути унікальним або новаторським унаслідок того, що він згенерований на основі закономірностей, вивчених з уже наявних даних. Водночас здатність технології створювати реалістичні зображення, текст, музику та генерувати нові ідеї та дизайн робить її потужним інструментом у творчих і дизайнерських процесах.

Хоча розвиток і впровадження генеративного штучного інтелекту значно вплине на більшість галузей та бізнес-функції, найбільші зміни очікуються в інтелектуальних роботах у таких сферах, як прийняття рішень, координація та комунікація, творчість та дизайн, що раніше вважалися "виключно людськими" [6].

Взаємодія між людиною та IT-продуктами, побудованими на базі великих мовних моделей, відбувається зазвичай у форматі діалогу "природною" мовою, наприклад англійською чи українською. Користувачі комунікують з *LLM* за допомогою промптів (*prompts*) – інструкцій або запитань, на які великі мовні моделі генерують відповіді (*completions*). Промпти можуть бути як простими (одне слово чи фраза), так і складними, з особливою структурою та форматуванням. Користувачі можуть отримувати різні формати відповідей (*completions*) залежно від призначення програми та специфіки запиту, наприклад факти, пояснення, ідеї, резюме, таблиці, картинки.

Для розв'язання складних завдань за допомогою *LLM* сучасні інженери та дослідники використовують техніки ланцюгових викликів (*chain of prompts*), у яких складне завдання розбивається на послідовність більш простих завдань та відповідних інструкцій для *LLM*. До того ж результат виконання інструкції на певному кроці впливає на вхідну інформацію інструкцій на наступних кроках [12]. Ще одним перспективним підходом до впровадження великих мовних моделей у виробничі процеси є розроблення "агентів" (*agents*) – автономних (частково) інтелектуальних систем, здатних самостійно (відносно) обирати стратегії виконання завдання, робити декомпозиції складних завдань, створювати та контролювати план виконання [8, 9]. Зазначені сучасні підходи взаємодії з великими мовними моделями відкривають широкі потенційні можливості для виконання складних завдань дизайну за допомогою *LLM*.

Схоже, що людство вступає в нову еру впровадження технологій в інтелектуальну та креативну працю. Зараз ще важко уявити, який вигляд матимуть "інтелектуальні фабрики" майбутнього, на яких людина буде співпрацювати з машиною. Але вже зараз бачимо, що великі мовні моделі здатні виконувати певну частину базових інтелектуальних операцій і технологія продовжує стрімко розвиватися.

Трансформація дизайну під впливом комп'ютерних технологій

Розвиток технологій генеративного штучного інтелекту не тільки впливає на процес дизайну, але й вимагає переосмислення багатьох концептів у сфері креативності, зокрема ролі ШІ у творчому процесі. У практичній площині машина стає здатною виконувати інтелектуальні завдання, які до цього вважалися "виключно людськими". З погляду етики це ставить нові питання щодо майбутнього співіснування та співтворчості людини й машини. Виникають та розвиваються такі концепти, як постантропоцентричний дизайн, співтворчість (людини та машини), агентність машини тощо. Ставляться питання пошуку синергії між сильними аспектами людського та машинного інтелекту для створення майбутнього дизайну та інновацій [10].

Дж. Толандер та М. Джонсон [10] досліджують особливості процесу генерації ідей та ескізів із застосуванням *ChatGPT* та *DALL·E*. Спостереження та висновки, наведені в публікації, добре відповідають особистому практичному досвіду використання інструментів генеративного ШІ в процесі дизайну.

1. Швидкість генерації, різноманітність ідей і артефактів виокремлюються серед переваг використання інструментів на кшталт *ChatGPT* та *DALL·E*. Однак глибина та новизна матеріалів, створених штучним інтелектом, може бути нестабільною та за якістю часто поступатися людській праці.

2. Взаємодія дизайнерів з інструментами генеративного ШІ на поточному етапі їх розвитку вимагає спеціалізованих навичок і додаткових зусиль, що може ускладнювати та порушувати плин творчого процесу.

3. Автори додатково звертають увагу на обмеження щодо передачі моделям ШІ контексту. Ці обмеження мають подвійну природу: з одного боку, вони випливають з технічних обмежень обсягу "вікна контексту" сучасних моделей ШІ; з іншого – у дизайнерів виникають труднощі з точним

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

визначенням та описом ключових аспектів контексту, які треба передати моделі.

4. Хоча генерація ідей та ескізів за допомогою ШІ є швидким та ефективним рішенням, це може призводити до зменшення рівня занурення дизайнера в проблему. Зникає або суттєво зменшується когнітивний і тактильний складник у процесі створення ескізів, що дизайнери іноді метафорично описують як "мислення руками". Крім того, автори висувають гіпотезу, що надмірна детальність ідей та ескізів, згенерованих за допомогою ШІ, звужує бачення дизайнера, що може бути небажаним на ранішніх етапах.

Важливо зазначити, що хоча Дж. Толандер та М. Джонсон проводили дослідження із застосуванням попередніх версій *ChatGPT* та *DALL·E*, наведені спостереження є актуальними й для більш нових версій інструментів.

С. Ванг та колеги [12] вивчають здатність *ChatGPT* виконувати різні типи завдань, пов'язаних з оперуванням знаннями, у виробничих процесах та, зокрема, віндустріальному дизайні, а саме:

– коректно розрізняти концепти відповідно до контексту та надавати їм пояснення;

– використовувати загальні знання для розв'язання конкретних завдань;

– аналізувати інформацію, виокремлювати її компоненти та змістовні характеристики;

– синтезувати нові знання;

– переносити знання, набуті в одній предметній галузі, в інші.

Дослідники формулюють чотири основних висновки.

1. *ChatGPT здатний розрізняти більшість загальновідомих концептів і надавати пов'язану з ними інформацію відповідно до запиту користувача, хоча не завжди робить це точно та акуратно.* Одна із зазначених проблем полягає в тому, що відповіді *ChatGPT* не завжди є релевантними до специфічного контексту. "Наприклад, коли ми запитали про процес розроблення специфікацій продукту, його відповідь була більш актуальною для розроблення програмного забезпечення, ніж для промислових продуктів", – зазначають автори. Також трапляються випадки, коли *ChatGPT* надавав явно неправдиву інформацію.

2. *Директивне формулювання інструкцій може додатково спонукати ChatGPT фабрикувати факти для того, щоб виконати поставлене завдання.* За своєю сутністю *ChatGPT* є ймовірнісною мовною моделлю, яка генерує зв'язний текст способом підбору кожного наступного слова відповідно до змісту попереднього тексту та на основі статистичних зв'язків між словами, що вона визначила, обробляючи навчальну інформацію. Так, модель генерує найкращий можливий (найімовірніший) варіант тексту, що відповідає запиту користувача. Дослідники зазначають, що незначні зміни в тональності формулювання інструкції суттєво впливають на результат. З огляду на результати експериментів автори роблять висновок, що в деяких випадках для того, щоб уникнути фабрикації даних, більш доцільно формулювати інструкції у формі запитань (наприклад, "Чи можуть спостереження за зміями надихнути до розроблення функціонального дизайну кавомашини?"), ніж у вигляді команд (наприклад, "Розроби функціональний дизайн кавомашини, надихаючись спостереженнями за зміями"). Такий підхід дає змогу знизити кількість випадків, коли модель фабрикує неправдиві або нерелевантні дані. Крім того, у дослідженні наголошено на важливості чіткості та детальності інструкцій для досягнення якісного результату.

3. *ChatGPT здатний виконувати складні креативні завдання, але досягнуті результати не перевищують рівень роботи досвідчених інженерів.* Дослідники провели тестування за допомогою відкритих запитань. Для відповідей необхідно було спочатку декомпозувати завдання на складники й надалі генерувати ідеї рішення для кожного зі складників. Результати показують, що *ChatGPT* може творчо синтезувати концепції різноманітними нетривіальними способами. Крім того, важливою особливістю технології є здатність генерувати значну кількість опцій "без втоми", яка притаманна людині. З іншого боку, дослідники роблять висновки, що, зважаючи на обмеження технології, рішення, розроблені *ChatGPT*, навряд чи перевищують середню якість рішень досвідчених інженерів.

4. *На момент проведення дослідження здатність ChatGPT розв'язувати завдання, що потребують складних аналітичних здібностей, була розвинута недостатньо.* З 26 запитань, що потребували критичного аналізу, *ChatGPT* коректно відповів лише на 46%. Дослідники зазначають, що складнощі виникали як з якісним, так і з кількісним аналізом, наголошуючи на трьох ключових обмеженнях:

1) технологія демонструє схильність до недостатньо обґрунтованих індуктивних узагальнень на основі декількох спостережень;

2) пріоритети, на основі яких *ChatGPT* робить якісні висновки, не завжди є очевидними;

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

3) значна кількість формул, обрана технологією для розв'язання задач, були неправильними.

Підсумовуючи, зазначимо, що розглянуті публікації приділяють увагу трансформаційному потенціалу генеративного штучного інтелекту, а також значним супутнім викликам. Для того, щоб більш повно осягнути багатогранний вплив ШІ на дизайн, необхідний комплексний дослідницький підхід.

Концепт-підхід для системного розуміння еволюції методології дизайн-мислення під впливом генеративного штучного інтелекту

Хоча процес дизайну є складним і нелінійним, дизайн-мислення пропонує структуровану методологію для його розбиття на серію ітеративних кроків, що спрямовують роботу дизайнера від формування дизайнерського завдання до створення та перевірки рішення. Виходячи із загальної траєкторії, що задається методологією, дизайнер на кожному кроці підбирає техніки відповідно до потреб. У процесі виконання технік створюються та збагачуються артефакти, у яких накопичуються набуті знання та згенеровані ідеї.

Як видно з аналізу публікацій, різні автори дизайн-мислення розглядають як поняття "дисципліна", "підхід", "відношення / принципи", "спосіб мислення", "процес", "застосування методів" та "методологія" [5]. Вважаємо, що погляд на дизайн-мислення саме як на методологію, тобто систему методів, основану на певних принципах, цінностях і теоретичних концептах, дає змогу осягнути його багатогранну сутність та дає підставу для комплексного розуміння його трансформації під впливом генеративного штучного інтелекту. Для подальшого аналізу вважаємо, що доцільно розглядати методологію дизайн-мислення у вигляді чотирьох взаємопов'язаних структурних шарів (рис. 1).

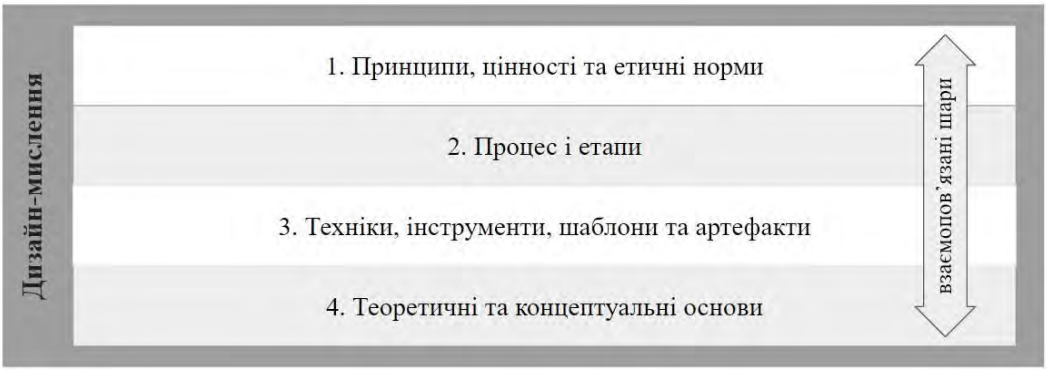


Рис. 1. Структурна модель методології дизайн-мислення (розробка авторів статті)

Детальне дослідження складників кожного із шарів дасть змогу закласти основу для подальшого всебічного аналізу впливу технологій генеративного ШІ на дизайн-мислення.

1. Принципи, цінності та етичні норми дизайн-мислення

Принципи дизайн-мислення – це фундаментальні ідеї, що визначають сутність методології, спрямовують діяльність команд і надалі розкриваються більш конкретно в процесах, техніках, методах та інструментах.

Принцип людиноцентричності визначає постійне зосередження уваги з боку дизайн-команди на глибоку емпатію людей, для яких команда працює. Один із найважливіших артефактів дизайн-мислення – Персона (*Persona*) – збірний образ, що уособлює риси, звички, проблеми та потреби реальної групи людей, для яких команда розробляє рішення. Глибоке емпатичне занурення в контекст, у якому живе та діє Персона, породжує інсайти – глибокі, нетривіальні, емоційно-забарвленні розуміння, що є каталізаторами для генерації креативних рішень. Дизайн-мислення пропонує багату палітру різноманітних технік і методів емпатії для занурення в контекст Персони.

Принцип ітеративності стверджує, що розроблення рішення має просуватися вперед (та іноді, якщо необхідно, назад) невеличкими швидкими ітеративними кроками. На кожному з кроків команда отримує певний інкремент та перевіряє, чи робить він позитивний внесок у загальне рішення за трьома критеріями:

1) бажаність (*desirability*) – чи відповідає отриманий інкремент потребам і бажанням користувачів;

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

2) життєздатність (*viability*) – чи є він економічно та організаційно життєздатним;

3) реалізованість (*feasibility*) – чи це технічно та матеріально досяжно.

Отже, реалізується ще один важливий для дизайн-мислення принцип – відповідність дизайну трьом критеріям.

Принцип системного підходу задає орієнтацію на постійне заглиблення в деталі контексту проблеми, над якою працює команда, з проясненням складних взаємозв'язків потреб і проблем зацікавлених сторін, а також інших важливих факторів.

Принцип евристичного підходу визначає орієнтацію команди на пошук дієвого практично-достатнього рішення для визначеної проблеми, навіть якщо воно не є ідеальним чи теоретично обґрунтованим.

Цінності дизайн-мислення виражають основні переконання про те, якою має бути атмосфера роботи команди та очікування щодо відносин і поведінки учасників.

Хоча можливе й індивідуальне застосування дизайн-мислення, більшість авторів звертають увагу на такі цінності методології, як командна робота, різноманітність досвіду, креативна синергія та відкритість до різних поглядів та ідей. Дизайн-мислення заохочує залучення користувачів у процес створення рішень, що уможливлює активну взаємодію між дизайнерами та користувачами.

Неосудливе ставлення, оптимізм, відкритість, нестандартне мислення та здатність ставити під сумнів наявні норми є важливими передумовами для створення атмосфери, в якій можуть бути згенеровані нетривіальні інноваційні рішення. Скетчинг та швидке прототипування дають змогу дизайнерам заощадити час і зусилля, спростовуючи гіпотези на ранніх етапах, та допомагають навчатися на практиці.

Етичні норми дизайн-мислення визначають засади взаємодії дизайнера із суспільством, створюючи моральну компоненту, що має доповнювати та врівноважувати інші мотивації та інтереси в прийнятті рішень у процесі дизайну.

Прагнучи до швидких і дешевих експериментів, дизайнери водночас не мають іти на компроміси в етичних питаннях. Важливо чесно та відверто інформувати людей про цілі досліджень, у повному обсязі відповідати на додаткові запитання, забезпечувати прозорість щодо дизайн-процесу.

Хоча перші прототипи можуть містити низку обмежень і недоліків, важливо не йти на компроміси в питаннях безпеки та вжити всіх необхідних заходів, щоб не завдати шкоди. Дизайнери мають усвідомлювати власну відповідальність за потенційні наслідки використання їх рішень, зокрема зважаючи на потенційний вплив на соціальні, екологічні, економічні, політичні системи.

Значна частина сучасної критики дизайн-мислення пов'язана з тим, що методологія фокусує команду на потребах людини та приділяє недостатньо уваги питанням сталого розвитку. Тоді як державні регулятори та міжнародні інституції стимулюють виробників і проєктувальників бути більш екологічними за допомогою економічних заходів, дизайнери мають взяти на себе особисті зобов'язання щодо розроблення рішень, які відповідають принципам сталого розвитку. Дизайн, орієнтований на людство (*humanity-centered design*), Д. Нормана є чудовим прикладом розвитку "зелених" ідей у дизайнерській спільноті [15].

У табл. 1 наведено перелік основних принципів, цінностей та етичних норм дизайн-мислення.

Таблиця 1. *Принципи, цінності та етичні норми дизайн-мислення*

Принципи	Цінності	Етичні норми
1) людиноцентричність; 2) ітеративність; 3) принцип відповідності дизайну трьом критеріям; 4) системний підхід; 5) евристичний підхід	1) командна робота; 2) відкритість до різних поглядів, ідей та досвіду; 3) креативна синергія; 4) залучення користувачів та активна взаємодія; 5) неосудливе ставлення; 6) оптимізм; 7) відкритість; 8) нестандартне мислення; 9) здатність ставити під сумнів чинні норми; 10) скетчинг і швидке прототипування; 11) навчання на практиці	1) чесність; 2) прозорість; 3) безпека; 4) відповідальність; 5) відповідність принципам сталого розвитку

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

2. Процес та етапи дизайн-мислення

У табл. 2 подано переклад таблиці з публікації Н. Рьош та інших [5]. Автори описали кілька типів структур процесу, що містять від трьох до шести етапів. У цьому разі трьохетапний процес, запропонований Дж. Лієдтка (2015), може бути "найменшим спільним знаменником".

Використовуючи різні підходи до розбиття на послідовні етапи процесу дизайн-мислення у власній професійній та освітній діяльностях та адаптуючи їх під потреби конкретних проєктів, дослідники визнали трьохтапну модель як найбільш оптимальною. На рис. 2 запропоновано трьохетапний процес дизайн-мислення (розробка авторів статті) з визначеними відповідними ключовимим діяльностями на кожному з етапів. Важливо зауважити, що послідовність діяльностей за дизайн-мисленням є ітеративною та нелінійною – отже, розмежування між етапами може бути нечітким.

Таблиця 2. *Етапи дизайн-мислення, запропоновані різними дослідниками* (*переклад з публікації* [5])

Автор(и)	Етапи процесу					
Лієдтка (2015)	Збір даних про потреби користувачів			Генерація ідей	Тестування	
Бекман і Барри (2007)	Спостерігати й помічати		Формувати і покращувати	Уявляти і дизайнити	Робити та експериментувати	
Беверленд та інші (2015)	Дестабілізація		Визначати і розробляти		Трансформація	
Браун (2008)	Натхнення			Ідеація	Реалізація	
Глен та інші (2015)	Знаходження проблеми	Спостереження	Візуалізація / формування значення	Ідеація	Прототипування і тестування	Перевірка на життєздатність
Да Сілва та інші (2020)	Розуміння	Спостереження	Визначення	Ідеація	Прототипування	Тест
Шапіра та інші (2017)	Дослідження		Інтерпретація	Ідеація	Експеримент	Еволюція

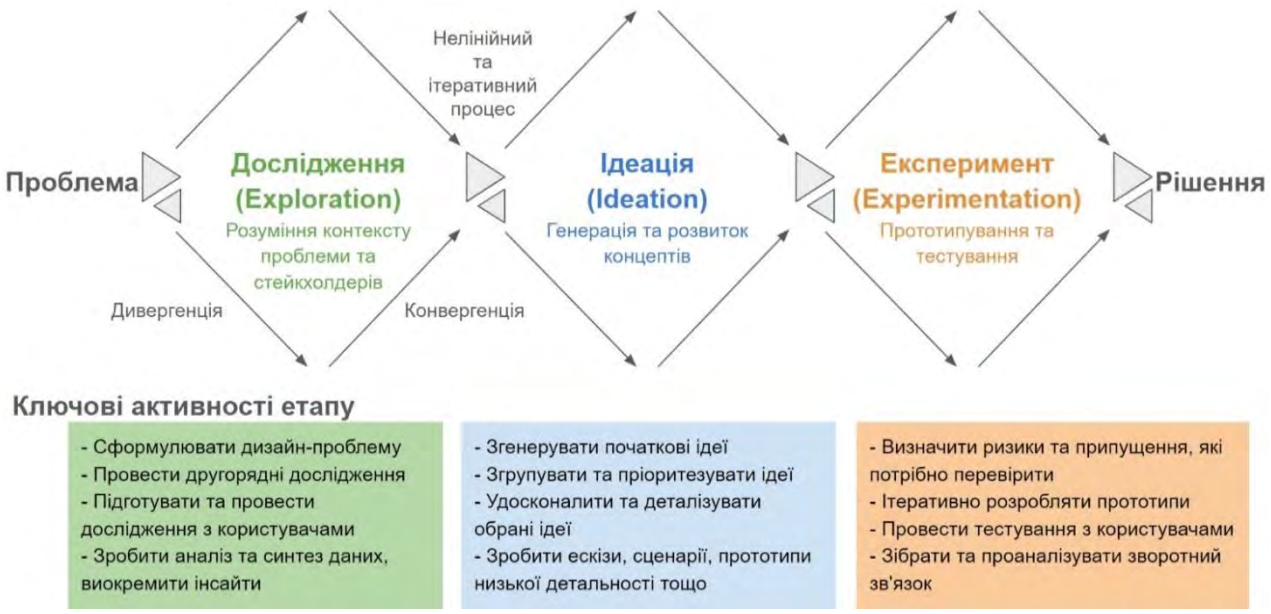


Рис. 2. Трьохетапний процес дизайн-мислення (розробка авторів статті)

Розглянемо докладніше етапи запропонованої послідовності дій процесу дизайн-мислення.

Етап 1. Дослідження (Exploration)

Цей етап присвячений збору та аналізу інформації про контекст проблеми, зацікавлені сторони, їхні потреби та виклики; виявленню зв'язків та впливових факторів. Мета – ретельно проаналізувати наявні знання, зібрати нові дані (переважно через емпатію) та виокремити практичні інсайти.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Етап 2. Ідеація (Ideation)

На другому етапі дизайн-мислителі генерують, формулюють та деталізують потенційні рішення для визначених проблем. Мета – спочатку згенерувати широкий спектр ідей, а потім зосередитися на найбільш перспективних, покращити та уточнити їх за необхідності.

Етап 3. Експеримент (Experiment)

На цьому етапі ретельно тестуються ідеї, ідентифікуються ризики, створюються прототипи та проводяться експерименти для валідації рішень в умовах, максимально наближених до реальності.

Хоча це не завжди явно зазначено в літературі, гарною практикою для команд, що працюють за дизайн-мисленням, буде спроба сформулювати дизайнерську проблему перед початком першого етапу. Якщо визначення проблеми не чітке, достатньо написати його у вигляді чернетки, яку можна ітеративно уточнювати під час роботи, коли команда збере більше інформації.

Увага команди, що працює за методологією дизайн-мислення, коливається між двома режимами – *конвергентним* і *дивергентним*. Дивергентне мислення полягає в розширенні уваги для пошуку та долучення якомога більшої кількості інформації або ідей. Цей режим вимагає неосудливого ставлення та наслідування принципу "що більше, то краще". Конвергентне мислення, навпаки, полягає в зосередженні уваги для визначення та відбору найбільш відповідної інформації та ідей, їх удосконалення. У цьому режимі заохочується конструктивна критика, орієнтації на деталі та принцип "менше краще". Для дизайн-мислителів критично важливо розуміти особливості обох режимів, впізнавати моменти для переходу між ними та вміти ефективно працювати в кожному з них.

3. Техніки, інструменти, шаблони й артефакти дизайн-мислення

Дизайн-мислення пропонує широку палітру технік, інструментів і шаблонів (*canvas*) для роботи на кожному з етапів процесу – від визначення проблеми до прототипування й тестування рішення. Відображаючи відкриту до нового природу методології, набір інструментів дизайн-мислення є динамічним і постійно інтегрує нові досягнення та інсайти з різноманітних галузей, наприклад, з таких як дизайн, бізнес, інженерія та наука.

Техніки дизайн-мислення спрямовують роботу команди та надають практичні вказівки щодо виконання конкретних завдань (як то зібрати відомості про користувача, придумати ідеї, протестувати прототипи) в процесі роботи над проблемою. Приклади технік передбачають інтерв'ю з користувачами, мозковий штурм, A/B тест.

Інструменти – це фізичні предмети або цифрові застосунки, що використовуються для виконання технік. Приклади фізичних інструментів містять стікери, маркери, дошки для письма тощо. Складніші цифрові інструменти можна проілюструвати такими прикладами, як *Miro, Figma, Adobe Creative Cloud, Sketch* та інші онлайн-платформи для співпраці.

Шаблони (*canvas*) дизайн-мислення допомагають команді візуально структурувати та організувати інформацію, ідеї або процеси. Вони спрямовують роботу для виконання конкретних технік під час дизайн-мислення, структуруючи спільну роботу та рефлексію. Декілька прикладів популярних шаблонів: канвас бізнес-моделі (*business model canvas*), канвас ціннісної пропозиції (*value proposition canvas*), карти емпатії (*empathy map*). Під час роботи команда заповнює ці та інші шаблони, зберігаючи в структурованому вигляді такі надбання, як знання, ідеї, карти процесів, моделі тощо.

Артефакти – це матеріальні та цифрові результати роботи команди, створені під час дизайн-мислення, зокрема записи інтерв'ю, описи Персон (*Persona*), ескізи, прототипи, дизайни, продукти та їх фрагменти. В артефактах зберігаються знання, ідеї та інші надбання команди. Також артефакти виконують комунікаційну функцію, передаючи інформацію між членами команди та іншими стейкхолдерами в скомпресованому та структурованому вигляді.

Ці чотири компоненти з інструментарію дизайн-мислення щільно взаємопов'язані. Шаблони структурують та спрямовують роботу під час виконання певних технік за допомогою фізичних або цифрових інструментів. Досягнуті результати втілюються, зберігаються та передаються у вигляді артефактів.

Детальний аналіз інструментарію методології виходить за межі цієї статті, але вважаємо необхідним навести декілька прикладів технік і артефактів, які часто використовуються в дизайн-мисленні (табл. 3).

Процес розв'язання кожної проблеми є творчим та нелінійним, водночас приклади технік і артефактів з різних етапів дизайн-мислення (табл. 3) ілюструють підхід до того, як можна декомпозувати складний творчий процес на нелінійну ітеративну послідовність

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

кроків, на кожному з яких дизайнер виконує специфічні техніки. Результати кожного з кроків зберігаються у вигляді відповідних артефактів, збагачуючи обсяг і якість накопичених знань щодо проблеми та ідей її розв'язання. На кожному кроці дизайнер визначає наступний, залежно від обсягу та детальності накопичених артефактів. У такий спосіб методологія спрямовує роботу дизайнерів на шляху від визначення дизайнерської проблеми до створення і тестування рішення.

Таблиця 3. *Приклади технік і артефактів дизайн-мислення*

Етап процесу	**Дослідження** (Exploration)	**Ідеація** (Ideation)	**Експеримент** (Experimentation)
Приклади технік	Питання: "Як би ми могли…" П'ять запитань "Чому?" Огляд літератури / звітів / кейс-стаді Аналіз ринку / трендів / конкурентів Мапи користувачів / зацікавлених сторін Інтерв'ю, етнографія, спостереження "Один день у взутті користувача" Спільні воркшопи з користувачами	Брейнстормінг Брейнрайтинг SCAMPER Ментальні карти Швидке голосування Уявні інвестиції Скетчинг Прототипування низької детальності	Швидке прототипування А/Б тестування Тестування на зручність користування "Чарівник з країни Оз" / "Фальшиві двері" / "Консьєрж" тест Розроблення мінімально життєздатного продукту (MVP)
Приклади артефактів	Формулювання дизайн-проблеми Персона користувача / клієнта Мапа зацікавлених сторін Мапа емпатії Мапа шляху користувача / клієнта Точка зору (PoV) Формулювання інсайтів	Карта / дошка ідей Матриця пріоритетів Ескізи, сценарії, пітч-презентації Прототипи низької детальності	Прототипи (паперові, цифрові, фізичні) Плани тестування Звіти про експерименти та зворотний зв'язок Мінімально життєздатний продукт (MVP)

4. Теоретичні та концептуальні основи дизайн-мислення

Дизайн-мислення є невід'ємним складником тривалої філософської рефлексії щодо природи творчості. Методологія містить концепції та теоретичні основи широкого спектра дисциплін, таких як евристика, системний аналіз, теорія прийняття рішень, інженерні науки, філософія, психологія, дизайн, менеджмент, економіка, соціологія, антропологія. Це відтворює складну багатогранну та міждисциплінарну природу дизайн-мислення. Методологія постійно еволюціонує та адаптується у відповідь на нові виклики, технологічні досягнення та суспільні зміни, що підкреслює її динамічну природу, гнучкість, актуальність та зумовлює ефективність у розв'язанні непростих проблем.

Стисло розглянемо концептуальні та теоретичні витоки методології з теорій дизайну та менеджменту.

Вплив теорії дизайну на концептуальні та теоретичні основи дизайн-мислення

Дж. Лієдтка [4] досліджує концептуальні витоки дизайн-мислення, аналізуючи літературу з теорії дизайну, датовану із середини XX ст., коли під впливом розвитку математики та науки про системи починається формування сучасної теорії дизайну. Відбувається швидка еволюція наукової думки від бачення дизайну як лінійного процесу до визнання його нелінійної ітеративної природи, що є необхідною для вирішення "заплутаних" проблем в умовах високої невизначеності, з якими має справу дизайн. Дж. Лієдтка виокремлює ще один важливий аспект дизайну (що робить його спорідненим з класичним науковим методом) – гіпотезо-орієнтований підхід з вагомим пізнавальним складником. Формулюючи та тестуючи гіпотези, дизайнери набувають нових знань про проблему, її контекст та можливі рішення. Крім того, ключовою розбіжністю між діяльністю дизайнерів та науковців є те, що дизайнери "шукають спосіб втілити те, чого ще не існує", тоді як "науковці шукають пояснення того, що вже є".

Дж. Лієдтка визначає такі спільні характеристики дизайну та дизайн-мислення: дуальність зосередження уваги одночасно на проблемі та на потенційному рішенні; гіпотезо-орієнтований ітеративний та експериментальний підхід; пошук можливостей у межах наявних обмежень; ефективність у розв'язанні комплексних проблем в умовах високої невизначеності; розроблення рішення конкретної проблеми (без необхідності узагальнення). Водночас Дж. Лієдтка виокремлює три основні розбіжності між дизайн-мисленням і традиційним дизайном: використання дизайн-мислення непрофесійними дизайнерами; вирішальна роль емпатії та ключова роль швидкого прототипування в дизайн-мисленні.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Розглядаючи концептуальні та теоретичні витоки дизайн-мислення, У. Йоханссон-Шельдберг [3] розрізняє пять субдискурсів у теорії дизайну, у кожному з яких можна визначити фундаментальні роботи, академічних послідовників та кожен з яких концептуалізує дизайн під певним кутом:

1) створення артефактів (Саймон, 1969);

2) рефлексивна практика (Шен, 1983);

3) діяльність, спрямована на вирішення проблем (Б'юкенен, 1992);

4) спосіб міркування / осмислення речей (Лоусон, 2006; Крос, 2006);

5) створення сенсів (Кріппендорфф, 2006).

Дизайн-мислення в контексті теорії менеджменту

У контексті теорії менеджменту У. Йоханссон-Шельдберг [3] виокремлює три різні дискурси, в яких дизайн-мислення розглядається як:

1) спосіб роботи в галузі дизайну та інновацій (Келлі, 2001, 2005; Браун, 2008, 2009);

2) підхід до організаційних питань із високою долею невизначеності та необхідна навичка для менеджерів-практиків (Данн і Мартін, 2006);

3) компонент теорії менеджменту (Боланд і Коллопі, 2004).

Дж. Лієдтка [4] простежує розвиток бачення дизайну як складника бізнес-практики та менеджменту від 1969 р., коли Герберт А. Саймон висловив ідею про те, що дизайн має бути елементом професійних тренінгів, адже бізнес принципово зацікавлений "не тим, якими речі є, а тим, якими вони можуть бути". Значний резонанс у бізнес-середовищі викликали роботи Р. Мартіна (2007, 2009), у яких він пропонує концепції інтегративного мислення – здатність синтезувати протилежні ідеї, знаходячи рішення вищого порядку, та дизайн-мислення – можливість перемикатися між інтуїтивним і аналітичним режимами мислення. Спираючись на роботи Р. Мартіна, Дж. Лієдтка наголошує на стратегічній, інтегративній та інноваційній ролі дизайн-мислення в практиці менеджменту. Вона визнає важливість гіпотезо-орієнтованого підходу до вирішення бізнес-проблем, який збалансовує інтуїтивні та аналітичні способи міркування із неабияким пріоритетом абдуктивного мислення, співпраці та співтворчості з користувачами. Таке "дизайнерське" мислення є альтернативним підходом щодо "традиційного" аналітичного бізнес-мислення, з огляду на унікальну здатність дизайн-мислення розв'язувати складні проблеми зі значним рівнем невизначеності.

Висновки та напрями майбутніх досліджень

Застосування генеративного ШІ у процесах дизайну для виконання креативних завдань має як численні переваги, так і суттєві обмеження, що потребують подальших досліджень. Такі інструменти, як *ChatGPT* та *DALL·E*, здатні швидко генерувати значну кількість альтернативних ідей і артефактів "без втоми", яка притаманна людині. Великі мовні моделі вже на поточному етапі розвитку розрізняють більшість загальновідомих концептів та вільно оперують ними, а також здатні виконувати складні креативні завдання. Водночас спостерігається і низка обмежень: результати роботи моделей генеративного ШІ не завжди є точними та акуратними, в окремих випадках ідеться про надання недостовірних фактів; добуті артефакти не перевищують за якістю роботи досвідчених дизайнерів та інженерів; здатність технології розв'язувати завдання, що потребують складних аналітичних здібностей, на цей момент розвинута недостатньо; використання технології вимагає спеціалізованих навичок і додаткових зусиль, що може ускладнювати та порушувати плин творчого процесу; у дизайнерів можуть виникати труднощі з передачею моделям ШІ контексту; існують ризики зменшення рівня занурення дизайнера в проблему. Крім того, розвиток технологій генеративного штучного інтелекту не тільки впливає на процес дизайну, але й вимагає переосмислення багатьох концептів у сфері креативності, зокрема ролі ШІ у творчій діяльності.

Синтезуючи зміст сучасних наукових публікацій щодо сутності, структури та змістовного наповнення дизайн-мислення, автори пропонують концепт-підхід, що розглядає методологію у вигляді чотирьох взаємопов'язаних структурних шарів:

1) принципи, цінності та етичні норми;

2) процес і етапи;

3) техніки, інструменти, шаблони та артефакти;

4) теоретичні та концептуальні основи.

У роботі також подано стислий опис елементів кожного із цих шарів. Така структурна модель методології дизайн-мислення дає змогу до певної міри декомпозувати складний творчий процес, закладаючи основу для подальшого всебічного аналізу впливу генеративного ШІ на дизайн-

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

мислення, а також для системного впровадження технології в дизайнерську діяльність.

Ґрунтуючись на результатах роботи, можна сформулювати декілька проблемних питань для подальших досліджень. Як має еволюціонувати методологія дизайн-мислення, що містить виникнення творчої агентності машини? Де проходитиме межа в співтворчості людини та машини, зумовлена власними обмеженнями технології генеративного ШІ? Чи мають бути проведені додаткові регуляторні межі, зумовлені етичними та іншими суспільними факторами?

Список літератури

1. Yaloveha I. Sources of design thinking: heuristic in the first and second stages of the history of philosophy and science. *Physical and Mathematical Education.* 2019. No. 4. P. 150–156. DOI: 10.31110/2413-1571-2019-022-4-023
2. Zub S., Yaloveha I. Development of heuristic methods at the beginning of the third stage of the history of philosophy and science. *Physical and Mathematical Education.* 2020. No. 2. P. 58–65. DOI: 10.31110/2413-1571-2020-024-2-008
3. Johansson-Sköldberg U., Woodilla J., Çetinkaya M. Design thinking: Past, present, and possible futures. *Creativity and innovation management.* 2013. No. 2. P. 121–146. DOI: 10.1111/caim.12023
4. Liedtka J. Why design thinking works. *Harvard Business Review.* 2018. No. 5. P. 72–79. URL: https://hbr.org/2018/09/why-design-thinking-works.
5. Rösch N., Tiberius V., Kraus S. Design thinking for innovation: context factors, process, and outcomes. *European Journal of Innovation Management.* 2023. No. 7. P. 160–176. DOI: 10.1108/EJIM-03-2022-0164
6. Chui M. et al. The economic potential of generative AI. 2023. URL: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier.
7. Franssen, Maarten, Gert-Jan Lokhorst, and Ibo van de Poel. Philosophy of Technology. *The Stanford Encyclopedia of Philosophy.* Spring 2023 Edition. URL: https://plato.stanford.edu/archives/spr2023/entries/technology/.
8. Grunde-McLaughlin M. et al. Designing LLM Chains by Adapting Techniques from Crowdsourcing Workflows. *arXiv preprint arXiv:2312.11681*. 2023. DOI: 10.48550/arXiv.2312.11681
9. Autonomous AI design architect. *Microsoft Learn.* URL: https://learn.microsoft.com/en-us/training/paths/autonomous-ai-design-architect/.
10. Tholander J., Jonsson M. Design ideation with ai-sketching, thinking, and talking with Generative Machine Learning Models. *Proceedings of the 2023 ACM Designing Interactive Systems Conference.* 2023. P. 1930–1940. DOI: 10.1145/3563657.3596014
11. Meron, Y., Araci, Y. T. Artificial intelligence in design education: evaluating ChatGPT as a virtual colleague for post-graduate course development. *Design Science.* 2023. No. 9. 30 p. DOI: 10.1017/dsj.2023.28
12. Wang X. et al. ChatGPT for design, manufacturing, and education. *Procedia CIRP.* 2023. No. 119. P. 7–14. DOI: 10.1016/j.procir.2023.04.001
13. Filippi S. Measuring the impact of ChatGPT on fostering concept generation in innovative product design. *Electronics.* 2023. No. 16. 3535 p. DOI: 10.3390/electronics12163535
14. Saadi J. I., Yang M. C. Generative Design: Reframing the Role of the Designer in Early-Stage Design Process. *Journal of Mechanical Design.* 2023. No. 145. 41411 p. DOI: 10.1115/1.4056799
15. Norman D. A. Design for a better world: Meaningful, sustainable, humanity centered. *MIT Press.* 2023.

References

1. Yaloveha, I. (2019), "Sources of design thinking: heuristic in the first and second stages of the history of philosophy and science", *Physical and Mathematical Education,* No. 4, P. 150–156. DOI: 10.31110/2413-1571-2019-022-4-023
2. Zub, S., Yaloveha, I. (2020), "Development of heuristic methods at the beginning of the third stage of the history of philosophy and science", *Physical and Mathematical Education,* No. 2, P. 58–65. DOI: 10.31110/2413-1571-2020-024-2-008
3. Johansson-Sköldberg, U., Woodilla, J., Çetinkaya, M. (2013), "Design thinking: Past, present, and possible futures", *Creativity and innovation management,* No. 2. P. 121–146. DOI: 10.1111/caim.12023
4. Liedtka, J. (2018), "Why design thinking works", *Harvard Business Review,* No. 5, P. 72–79, available at: https://hbr.org/2018/09/why-design-thinking-works.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) *ISSN 2524-2296 (online)*

5. Rösch, N., Tiberius, V., Kraus, S. (2023), "Design thinking for innovation: context factors, process, and outcomes", *European Journal of Innovation Management,* No. 7, P. 160–176. DOI: 10.1108/EJIM-03-2022-0164
6. Chui, M. et al. (2023), "The economic potential of generative AI", available at: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier.
7. Franssen, Maarten, Gert-Jan, Lokhorst, and Ibo van de Poel (2023), "Philosophy of Technology", *The Stanford Encyclopedia of Philosophy,* Spring 2023 Edition, available at: https://plato.stanford.edu/archives/spr2023/entries/technology/.
8. Grunde-McLaughlin, M. et al. (2023), "Designing LLM Chains by Adapting Techniques from Crowdsourcing Workflows", *arXiv preprint arXiv:2312.11681*. DOI: 10.48550/arXiv.2312.11681
9. Autonomous AI design architect. *Microsoft Learn,* available at: https://learn.microsoft.com/en-us/training/paths/autonomous-ai-design-architect/.
10. Tholander, J., Jonsson, M. (2023), "Design ideation with ai-sketching, thinking, and talking with Generative Machine Learning Models", *Proceedings of the 2023 ACM Designing Interactive Systems Conference,* P. 1930–1940. DOI: 10.1145/3563657.3596014
11. Meron, Y., Araci, Y. T. (2023), "Artificial intelligence in design education: evaluating ChatGPT as a virtual colleague for post-graduate course development", *Design Science,* No. 9, 30 p. DOI: 10.1017/dsj.2023.28
12. Wang, X. et al. (2023), "ChatGPT for design, manufacturing, and education", *Procedia CIRP,* No. 119, P. 7–14. DOI: 10.1016/j.procir.2023.04.001
13. Filippi, S. (2023), "Measuring the impact of ChatGPT on fostering concept generation in innovative product design", *Electronics,* No. 16, 3535 p. DOI: 10.3390/electronics12163535
14. Saadi, J. I., Yang, M. C. (2023), "Generative Design: Reframing the Role of the Designer in Early-Stage Design Process", *Journal of Mechanical Design*, No. 145, 41411 p. DOI: 10.1115/1.4056799
15. Norman, D. A. (2023), "Design for a better world: Meaningful, sustainable, humanity centered", *MIT Press.*

Надійшла (Received) 05.06.2024

Відомості про авторів / About the Authors

Новаковський Антон Валерійович – Харківський національний університет радіоелектроніки, аспірант кафедри прикладної математики, Харків, Україна; e-mail: anton.novakovskyi@nure.ua; ORCID ID: https://orcid.org/0009-0007-6129-5374

Яловега Ірина Георгіївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри прикладної математики; Харківський національний економічний університет імені Семена Кузнеця, доцент кафедри вищої математики та економіко-математичних методів, Харків, Україна; e-mail: iryna.ialoveha@nure.ua; ORCID ID: https://orcid.org/0000-0002-2486-1812

Novakovskyi Anton – Kharkiv National University of Radio Electronics, Postgraduate at the Department of Applied Mathematics, Kharkiv, Ukraine.

Yaloveha Iryna – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Applied Mathematics; Simon Kuznets Kharkiv National University of Economics, Associate Professor at the Department of Higher Mathematics and Economic and Mathematical Methods, Kharkiv, Ukraine.

IMPLEMENTATION OF GENERATIVE ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN CREATIVE ACTIVITIES: DEVELOPMENT OF A STRUCTURAL MODEL OF DESIGN THINKING

The **subject** of the study is systemic changes in the methodology of design thinking, taking place under the influence of the development and spread of generative artificial intelligence (AI) technologies in design and other creative industries. The **purpose** of the work is: analysis of modern research on the impact of generative AI technologies on creative industries,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

design and on design thinking; development of a structural model of design thinking to further explore the evolution of the methodology. The following **tasks** are set in the article: to analyze modern scientific publications regarding the essence, structure and content of design thinking; review research on the benefits and challenges of using generative AI in design processes; to develop a model that allows identifying and describing changes in key components of the design thinking methodology arising under the influence of widespread adoption of generative AI technologies. During the research, the following **methods** were used: analysis and synthesis of the content of technical, economic, philosophical, linguistic, historical scientific and methodical research on the problems of forming the conceptual apparatus of the design-thinking methodology and the use of generative AI in design processes; comparative-historical, retrospective methods; structural and logical analysis. The following **results** were achieved: the actualized need for a comprehensive research approach to analyze the multifaceted impact of AI technologies on design; the key advantages and challenges associated with the integration of AI into creative processes are identified; a structural model of presentation of the design-thinking methodology was developed in the form of four interconnected structural layers with subsequent decomposition of each of the layers into constituent elements. The **conclusions** highlight the depth and multifaceted nature of the changes taking place in design and other creative industries under the influence of generative AI and need further in-depth research. The developed structural model of the design-thinking methodology allows to decompose the complex creative process to a certain extent, laying the foundation for a comprehensive analysis of the evolution of the methodology and the systematic introduction of generative artificial intelligence technologies into design processes.

Keywords: design-thinking methodology; generative artificial intelligence; innovations in design; structural model; creative process.

Бібліографічні описи / Bibliographic descriptions

Новаковський А. В., Яловега І. Г. Упровадження технологій генеративного штучного інтелекту в творчу діяльність: розроблення структурної моделі дизайн-мислення. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 108–120. DOI: https://doi.org/10.30837/2522-9818.2024.2.108

Novakovskyi, A., Yaloveha, I. (2024), "Implementation of generative artificial intelligence technologies in creative activities: development of a structural model of design thinking", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 108–120. DOI: https://doi.org/10.30837/2522-9818.2024.2.108

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 004.75
DOI: https://doi.org/10.30837/2522-9818.2024.2.121

М. Перетяга

МЕТОДИ ВИЯВЛЕННЯ АНОМАЛІЙ У МІКРОСЕРВІСАХ ІЗ ВИКОРИСТАННЯМ СТАТИСТИЧНОГО АНАЛІЗУ

Предметом дослідження є методи виявлення аномалій у мікросервісах із використанням статистичного аналізу. Мікросервіси є популярною архітектурою для розроблення програмного забезпечення, що дає змогу створювати гнучкі та масштабовані системи. Однак через свою складність такі системи можуть бути вразливими до різного роду аномалій, що здатні впливати на їх продуктивність і надійність. **Мета роботи** полягає в аналітичному огляді сучасних методів виявлення аномалій у мікросервісних системах із впровадженням методів статистичного аналізу. Виявлення аномалій є критично важливим для стабільної роботи системи та швидкого реагування на можливі проблеми. Для досягнення мети сформульовано такі **завдання**: огляд методів виявлення аномалій у мікросервісах; опис принципів регресійного аналізу, кластерного аналізу та методу головних компонент; порівняння методів за критеріями ефективності, обчислювальної складності, стійкості до шуму та адаптивності; рекомендації щодо вибору методу та можливість їх комбінування; підбиття висновків та визначення напрямів для майбутніх досліджень. Розглянуто **метод** для виявлення аномалій у мікросервісах, що передбачає регресійний аналіз, кластерний аналіз та метод головних компонент (PCA). **Результати дослідження** підтвердили, що кожен метод має переваги та обмеження. Регресійний аналіз ефективний у системах з явними трендами, але менш ефективний у складних і динамічних системах. Кластерний аналіз стійкий до шуму та здатний виявляти як окремі аномалії, так і групи аномальних подій, але вимагає значних обчислювальних ресурсів. Метод головних компонент (PCA) є потужним інструментом для аналізу високорозмірних даних, але має обмеження у високій складності обчислень та інтерпретації результатів. Кожен із розглянутих методів має свої переваги й недоліки, тому в дослідженні запропоновано новий метод, що полягатиме в їх комбінуванні. **Висновки** наголошують на важливості статистичного аналізу для моніторингу мікросервісних систем. Правильно підібрані методи аналізу інформації полегшують виявлення аномалій у складних середовищах, таких як мікросервіси. Використання регресійного аналізу, кластерного аналізу та методу головних компонент дає змогу отримати глибокий інсайт щодо роботи системи. Проте для найкращих результатів рекомендується комбінувати різні методи та аналізувати їх застосування в контексті конкретної системи. Такий підхід забезпечує більшу стійкість до аномалій та швидше реагування на них у мікросервісних архітектурах.

Ключові слова: виявлення аномалій; мікросервіси; статистичний аналіз; регресійний аналіз; кластеризація.

Вступ

У сучасному світі інформаційні технології швидко розвиваються і мікросервісна архітектура стає все більш популярною завдяки своїй спроможності забезпечувати гнучкість, масштабованість та незалежність компонентів програмного забезпечення. Мікросервіси дають змогу розробникам створювати автономні сервіси, що можуть легко взаємодіяти один з одним за допомогою стандартизованих інтерфейсів. Однак ця архітектура також висуває нові виклики, особливо у сфері управління та моніторингу системи, де важливим завданням є виявлення та реагування на аномалії, що можуть виникати в роботі системи. Аномалії в мікросервісах бувають різноманітні, починаючи від збоїв у роботі окремих компонентів до проблем у взаємодії між сервісами [1]. Вчасне виявлення таких аномалій є ключовим для забезпечення стабільної та безперервної роботи системи. Традиційні методи моніторингу часто не можуть впоратися з цим завданням через складність і динамічність мікросервісних середовищ. Отже, виникає потреба у використанні більш складних і адаптивних методів аналізу, таких як статистичний аналіз [2]. Цей метод пропонує широкий набір інструментів для аналізу інформації, виявлення закономірностей та відхилень від нормального функціонування. Зазначені методи можуть бути дуже корисними для моніторингу мікросервісів та виявлення аномалій у їх роботі. Серед методів статистичного аналізу, що можуть застосовуватися для цієї мети, варто виокремити регресійний аналіз, контрольні карти та методи кластеризації [3]. Крім того, сучасні методи на основі машинного навчання дають змогу створювати точніші та більш адаптивні моделі для виявлення аномалій [4]. Важливим аспектом цього дослідження є інтеграція статистичних методів у реальні мікросервісні системи. Це передбачає не лише

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

розроблення методів, але й створення інструментів, що дають змогу автоматизувати процес моніторингу та виявлення аномалій. Така інтеграція може значно покращити здатність систем реагувати на потенційні проблеми в режимі реального часу, знижуючи ризики та підвищуючи загальну ефективність роботи [5].

Це дослідження присвячене розробленню та оцінюванню ефективності методів статистичного аналізу для виявлення аномалій у мікросервісних системах. У статті будуть проаналізовані наявні підходи, їх адаптація до особливих потреб мікросервісної архітектури, а також виконана експериментальна перевірка ефективності запропонованих методів на основі реальних даних. Очікується, що результати цього дослідження сприятимуть підвищенню надійності та продуктивності мікросервісних систем. Виявлені аномалії можуть бути вчасно виправлені, що дасть змогу уникнути збоїв і забезпечити стабільну роботу системи. Запропоновані методи можна впроваджувати в практику, що сприятиме покращенню ефективності роботи мікросервісів та зниженню ризиків, пов'язаних з імовірними збоями.

Аналіз останніх досліджень і публікацій

У сучасних умовах розвитку інформаційних технологій мікросервісна архітектура набуває все більшої популярності, що зумовлює зростання інтересу до питань виявлення аномалій у таких системах. Останні дослідження та публікації в окресленій галузі зосереджуються на розробленні ефективних методів і підходів для виявлення аномалій, що беруть до уваги особливості мікросервісних архітектур, їх складність та динамічність.

Одним із найактуальніших напрямів досліджень є застосування методів машинного навчання для виявлення аномалій у мікросервісах. Так, низка досліджень присвячена використанню нейронних мереж для аналізу поведінки мікросервісів. Нейронні мережі можуть навчатися на великих обсягах інформації та виявляти складні патерни, що вказують на аномалії. Наприклад, у роботі колективу авторів *X. Jiang, C. Chen, Y. Zhou, J. Zhang, X. Liu* (2021) запропоновано застосування рекурентних нейронних мереж для аналізу часових рядів даних, отриманих від мікросервісів [6]. Цей підхід дає змогу виявляти аномалії, що виникають унаслідок тимчасових залежностей між сервісами.

Інший підхід до виявлення аномалій оснований на алгоритмах класифікації та кластеризації. Дослідники *Z. Li, W. Zhang, Z. Li, M. Zhao* (2020) розглядають використання методу *k*-середніх для кластеризації даних мікросервісів та виявлення аномальних кластерів [7]. Цей метод допомагає ідентифікувати групи подій, які відрізняються від нормальної поведінки, що може свідчити про наявність аномалій.

Адаптивні методи виявлення аномалій також привертають значну увагу вчених. Вони орієнтовані на розроблення алгоритмів, що можуть автоматично підлаштовуватися до змін у середовищі мікросервісів. У дослідженні *S. He, H. Jin, W. Dai, X. Hu* (2018) запропоновано адаптивний метод, що використовує еволюційні алгоритми для налаштування параметрів моделі в режимі реального часу [8]. Це дає змогу системі адаптуватися до змін у навантаженні або конфігурації мікросервісів, забезпечуючи точне виявлення аномалій.

Окремий напрям досліджень пов'язаний з інтеграцією статистичних методів у системи моніторингу мікросервісів. Традиційні статистичні методи, такі як контрольні карти та регресійний аналіз, можуть бути використані для виявлення відхилень у поведінці системи. У роботі *F. Liu, K. Ting, Z. Zhou* (2012) описано застосування контрольних карт Шухарта для моніторингу показників продуктивності мікросервісів [9]. Цей підхід дає змогу розрізнити значні відхилення та нормальну поведінку і вчасно реагувати на потенційні проблеми.

Сучасні дослідження також наголошують на важливості візуалізації даних та взаємодії з ними. Інтерактивні інструменти для візуалізації можуть значно полегшити процес аналізу та виявлення аномалій. У роботі авторів *J. Luo, J. Wu, Y. Zhang, L. Sun, Y. Liu* (2022) запропоновано інструмент для візуалізації даних мікросервісів, який дає змогу операторам системи швидко ідентифікувати аномалії на основі візуальних індикаторів [10].

Значну увагу дослідників привертає проблема масштабованості методів виявлення аномалій. Мікросервісні архітектури часто містять чималу кількість компонентів, що генерують великі обсяги інформації. Це вимагає розроблення методів, що можуть ефективно працювати з великими даними. У праці *P. He, R. Ranjan, J. Nogueira, L. Veiga, W. Zhao* (2021) розглянуто використання розподілених обчислювальних методів для аналізу даних мікросервісів [11]. Упровадження таких підходів

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

дає змогу обробляти значні обсяги інформації в режимі реального часу, забезпечуючи вчасне виявлення аномалій.

У багатьох публікаціях наголошується на важливості зменшення кількості хибних спрацювань унаслідок виявлення аномалій. Хибні спрацювання можуть призводити до марного використання ресурсів та зниження ефективності роботи системи. У дослідженні *A. Pimentel, L. Clifton, L. Clifton, Y. Lee, A. Kang* (2014) запропоновано метод комбінованого аналізу, що одночасно застосовує декілька статистичних і машинних методів для підвищення точності виявлення аномалій та зменшення кількості хибних спрацювань [12].

Останні дослідження також висвітлюють важливість використання історичних даних для прогнозування можливих аномалій у майбутньому. Прогнозування на основі аналізу історичних даних дає змогу активно керувати мікросервісами та запобігати можливим проблемам. У роботі *Y. Jiang, K. Tan, S. Lam, P. Chen* (2017) розглянуто застосування методів прогнозування на основі часових рядів для виявлення аномалій у мікросервісах [13]. Цей підхід допомагає виявляти потенційні проблеми до їх виникнення, що значно підвищує надійність та ефективність системи.

Отже, аналіз останніх досліджень і публікацій з виявлення аномалій у мікросервісах із використанням статистичного аналізу свідчить про значний прогрес у цій сфері. Сучасні підходи, основані на застосуванні машинного навчання, адаптивних методів, інтеграції статистичних інструментів, візуалізації даних та аналізу великих обсягів інформації, дають змогу значно підвищити точність та ефективність виявлення аномалій. Подальші дослідження в цьому напрямі мають потенціал для розроблення ще більш точних та ефективних методів, що забезпечать стабільну й надійну роботу мікросервісних систем у довгостроковій перспективі.

Визначення не розв'язаних раніше частин загальної проблеми. Мета роботи, завдання

Незважаючи на значний прогрес у розробленні методів виявлення аномалій у мікросервісах, чимало аспектів цієї проблеми залишаються недостатньо вивченими та вирішеними. Одним із ключових питань, що потребує розв'язання, є забезпечення високої точності виявлення аномалій за одночасного зниження кількості хибних спрацювань. Хибні спрацювання спричиняють неправильне реагування на ситуації, що зі свого боку може викликати нераціональне використання ресурсів і зниження загальної ефективності системи. Більшість сучасних методів стикаються з викликами, пов'язаними з балансом між чутливістю та специфічністю моделей.

Іншим важливим аспектом є адаптивність методів виявлення аномалій. Мікросервісні архітектури часто визначаються високою динамічністю – компоненти можуть змінюватися, додаватися або вилучатися, навантаження може значно варіюватися в часі. У таких умовах методи, не здатні адаптуватися до змін, втрачають свою ефективність [14]. Існує потреба в розробленні алгоритмів, які можуть автоматично підлаштовувати свої параметри відповідно до змін у середовищі.

Ще одним нерозв'язаним питанням є оброблення значних обсягів інформації в реальному часі. Мікросервісні системи генерують величезні обсяги даних, що створює значні виклики для аналізу та виявлення аномалій у реальному часі. Багато сучасних методів потребують чималих обчислювальних ресурсів і не завжди здатні забезпечити оперативність оброблення інформації, що необхідно для вчасного виявлення аномалій [15].

Проблема інтеграції різних методів також залишається актуальною. Використання комбінації статистичних методів і машинного навчання може значно підвищити точність і надійність виявлення аномалій. Однак ефективна інтеграція таких методів вимагає додаткових досліджень для виявлення оптимальних стратегій об'єднання даних із різних джерел і алгоритмів.

Нарешті, проблема прогнозування аномалій на основі історичних даних також потребує подальшого вивчення. Хоча існують деякі успішні приклади використання часових рядів для прогнозування аномалій, більшість методів все ще потребує поліпшення в частині точності та надійності прогнозів. Застосування глибоких нейронних мереж та інших передових методів машинного навчання для аналізу історичних даних може відкрити нові можливості у цій сфері.

Метою роботи є аналітичний огляд сучасних методів виявлення аномалій у мікросервісних системах із використанням статистичних підходів для розроблення нового, комбінованого. Виявлення аномалій є ключовим для забезпечення стабільної

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

роботи системи та вчасного реагування на можливі проблеми.

Завданнями роботи є дослідження наявних методів статистичного аналізу, що застосовуються для виявлення аномалій, а також аналіз адаптивних методів виявлення аномалій, які можуть підлаштовуватися до змін у мікросервісних архітектурах, беручи до уваги змінні навантаження та динамічні зміни компонентів. Важливим є також оцінювання сучасних підходів до оброблення великих обсягів інформації в режимі реального часу та визначення їх ефективності в контексті виявлення аномалій у мікросервісах. На основі проведеного аналізу робота спрямована на пропозицію методу зниження кількості хибних спрацювань або підвищення адаптивності наявних методів на основі комбінування вище від зазначених методів.

Матеріали й методи

Регресійний аналіз

У контексті виявлення аномалій у мікросервісних системах регресійний аналіз є одним із ключових статистичних методів, що дає змогу моделювати взаємозв'язки між різними змінними. Це важливо для розуміння нормальної поведінки системи та ідентифікації відхилень, які можуть вказувати на аномалії [17].

Основи регресійного аналізу

Регресійний аналіз – це набір статистичних процесів для оцінювання взаємозв'язків між змінними. Основна мета регресійного аналізу – визначити, як незалежні змінні впливають на залежну змінну. У контексті мікросервісних архітектур регресійний аналіз допомагає моделювати нормальну поведінку сервісів, зважаючи на різні фактори, зокрема обсяг трафіку, час відповіді, завантаження ресурсів тощо.

Типи регресійного аналізу

1. Лінійна регресія

– Однофакторна лінійна регресія: моделює взаємозв'язок між однією незалежною змінною та однією залежною змінною. Основна модель однофакторної лінійної регресії має такий вигляд:

$$y = \beta_0 + \beta_1 x + \varepsilon \ , \qquad (1)$$

де y – залежна змінна;

x – незалежна змінна;

β_0 – вільний член (інтерцептор);

β_1 – коефіцієнт регресії (нахил лінії);

ε – випадкова похибка.

– Множинна лінійна регресія: бере до уваги кілька незалежних змінних. Основна модель множинної лінійної регресії має такий вигляд:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \ , \qquad (2)$$

де y – залежна змінна;

$x_1, x_2, \ldots, x_n$ – незалежні змінні;

β_0 – вільний член;

$\beta_1, \beta_2, \ldots, \beta_n$ – коефіцієнти регресії;

ε – випадкова похибка.

2. Поліноміальна регресія

Використовується, коли взаємозв'язок між змінними не є лінійним. Основна модель поліноміальної регресії другого порядку має такий вигляд:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \ , \qquad (3)$$

де y – залежна змінна;

x – незалежна змінна;

β_0 – вільний член;

β_1 – коефіцієнт при лінійному члені;

β_2 – коефіцієнт при квадратному члені;

ε – випадкова похибка.

3. Логістична регресія

Використовується для моделювання бінарних залежностей, тобто коли залежна змінна може приймати тільки два значення (наприклад, аномалія / немає аномалії). Основна модель логістичної регресії має такий вигляд:

$$\log it\left(p\right) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n , \qquad (4)$$

де p – ймовірність настання події (аномалії);

β_0 – вільний член;

$\beta_1, \beta_2, \ldots, \beta_n$ – коефіцієнти регресії;

$x_1, x_2, \ldots, x_n$ – незалежні змінні.

Процес виконання регресійного аналізу розпочинається зі збору відповідних даних з мікросервісних систем, таких як журнали запитів, метрики продуктивності та відомості про навантаження. Після збору даних вони попередньо обробляються: очищаються, заповнюються пропуски, проводиться нормалізація та масштабування змінних.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Потім визначається тип регресійного аналізу, що найкраще підходить для вивчення природи даних та досліджуваних взаємозв'язків. Дані розподіляються на навчальні та тестові вибірки для оцінювання точності моделі.

Далі модель оцінюється. Використання статистичних критеріїв для оцінювання точності моделі, таких як коефіцієнт детермінації $\left(R^2\right)$ (див. формулу 5), середньоквадратична помилка (MSE), середня абсолютна помилка (MAE).

$$R^2 = 1 - \frac{\sum_i^n = 1\left(y_i - \gamma_i\right)^2}{\sum_i^n = 1\left(y_i - \mu_i\right)^2}\ , \qquad (5)$$

де y_i – фактичні значення залежної змінної;

μ_i – передбачені значення;

γ_i – середнє значення залежної змінної;

n – кількість спостережень.

Наступним кроком є інтерпретація результатів. Аналіз коефіцієнтів регресії для встановлення впливу кожної незалежної змінної на залежну змінну.

Визначення порогових значень для ідентифікації аномалій на основі відхилень від передбачених значень.

Використання регресійного аналізу для виявлення аномалій

Регресійний аналіз дає змогу створити базову лінію нормальної поведінки мікросервісів. Виявлення аномалій здійснюється способом порівняння фактичних значень змінних із передбаченими моделлю значеннями. Значні відхилення від моделі можуть указувати на наявність аномалій.

Наприклад, якщо регресійна модель передбачає, що час відповіді сервісу має бути в межах 100–200 мс за певного обсягу трафіку, а фактичний час відповіді перевищує 500 мс, це може вказувати на аномалію. Важливим аспектом є налаштування порогів виявлення аномалій для мінімізації хибних спрацювань та підвищення точності виявлення.

Кластерний аналіз

Кластерний аналіз – це метод машинного навчання, що застосовується для групування подій або об'єктів у наборі даних у різні кластери чи групи в такий спосіб, щоб об'єкти всередині одного кластера були максимально схожими між собою, а об'єкти з різних кластерів якомога більше відрізнялися [18].

Нехай маємо набір даних:

$$X = \left\{x_1, x_2, \ldots, x_n\right\}\ , \qquad (6)$$

де x_i – це вектор, що є об'єктом або подією. Метою кластерного аналізу є розділення цих об'єктів на k кластерів таким чином, щоб об'єкти всередині одного кластера були максимально схожими між собою, а об'єкти з різних кластерів якомога більше відрізнялися.

Перше, що потрібно зробити, – визначити функцію відстані між об'єктами. Це може бути евклідова відстань, манхеттенська відстань, косинусна схожість тощо. Позначимо $d\left(x_i, x_j\right)$ як відстань між об'єктами x_i та x_j.

Функція вибору центроїдів визначає, як обрати початкові центроїди кластерів. Це може бути випадкове обрання з набору даних або будь-яка інша евристика.

Процес кластеризації розпочинається з вибору початкових центроїдів та призначення кожного об'єкта до найближчого кластера на основі функції відстані. Потім центроїди перераховуються як середнє значення об'єктів у кожному кластері. Цей процес продовжується доти, доки центроїди та належність об'єктів до кластерів не стабілізуються.

Після завершення кластеризації можна виявити аномалії, аналізуючи розмір і форму кластерів. Аномальні події здатні виявлятися як одиночні об'єкти, що не належать жодному з головних кластерів, або як малі, відмінні кластери.

Нехай маємо набір метрик про використання ресурсів мікросервісної системи. Можемо впроваджувати кластерний аналіз для групування днів або годин, коли спостерігалися незвичайні патерни використання ресурсів у відповідних кластерах. Це дає змогу ідентифікувати часові аномалії у застосуванні ресурсів та реагувати на них для підтримки нормальної роботи системи.

Отже, кластерний аналіз є потужним інструментом для виявлення аномалій у мікросервісних системах, даючи змогу ідентифікувати відхилення в поведінці системи та реагувати на них для забезпечення надійності та ефективності.

Метод головних компонент

Метод головних компонент (PCA) є потужним інструментом у сфері аналізу інформації, що впроваджується для зменшення розмірності даних та виявлення в них складних структур.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

PCA розглядається як техніка без завантаження, оскільки не вимагає маркування даних або заздалегідь визначених класів. Замість цього, PCA шукає лінійні комбінації ознак, які максимізують дисперсію даних, а також зменшують кореляцію між ознаками [19]. Розглянемо кроки PCA та його математичні аспекти.

Почнемо з обчислення середнього значення для кожної ознаки в наборі даних. Нехай маємо n ознак та m спостережень. Тоді середнє значення x_j для кожної ознаки j обчислюється як

$$x_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij} \ , \qquad (7)$$

де x_{ij} – значення ознаки j для спостереження i.

Потім проводиться центрування даних, тобто віднімається середнє значення кожної ознаки від відповідного значення даних: $\mu_{ij} = x_{ij} - x_j$.

Після центрування обчислюється коваріаційна матриця, що містить коваріації між усіма парами ознак. Коваріація між ознаками j та k обчислюється як

$$\mathrm{cov}\left(x_j, x_k\right) = \frac{1}{m}\sum_{i=1}^{m}\left(\mu_{ij}\cdot\mu_{ik}\right) \ . \qquad (8)$$

Головні компоненти обчислюються як власні вектори коваріаційної матриці. Вони вказують напрямки максимальної дисперсії даних. Нехай $v_1, v_2, \ldots, v_n$ – це власні вектори коваріаційної матриці, відсортовані за принципом спадання власних значень.

Головні компоненти обираються залежно від власних значень. Типово обираються перші k головних компонент, які пояснюють значну частину дисперсії даних, де k – кількість ознак, що має бути збережена.

Коли головні компоненти обрані, дані проєктуються на новий простір ознак, утворений цими компонентами. Позначимо матрицю головних компонент:

$$v_k = \left[v_1, v_2, \ldots, v_k\right] \ . \qquad (9)$$

Тоді проєкція даних xi на простір головних компонент обчислюється як

$$y_i = V_k^t \cdot \mu_i \ , \qquad (10)$$

де y_i – вектор проєкції для спостереження i.

PCA є важливим інструментом для аналізу даних та виявлення складних залежностей. Використання цього методу дає змогу зменшити розмірність даних і водночас зберегти важливу інформацію, а також полегшує виявлення аномалій або розбіжностей у наборі даних [20].

Новий запропонований метод

Запропонований метод полягатиме у використанні трьох основних статистичних методів: регресійного аналізу, кластерного аналізу та методу головних компонент (PCA). Кожен із згаданих методів застосовується для виявлення аномалій у мікросервісних системах, після чого їх результати об’єднуються за допомогою інтеграційного показника аномалій (IPA).

Інтеграційний показник аномалій (IPA)

Метою буде об’єднання результатів усіх трьох методів для комплексного оцінювання аномалій. Інтеграційний показник аномалій обчислюється за формулою

$$IPA_i = a_1 d_i + a_2 d_{c_i} + a_3 d_{PCA_i} \ , \qquad (11)$$

де a_1, a_2, a_3 – вагові коефіцієнти, що визначають внесок кожного методу.

У процесі встановлення порогу аномалії з’ясовується порогове значення IPA, перевищення якого свідчить про аномалію.

Для нового методу властиві кілька особливостей методології.

- Комбінований підхід: використання трьох різних методів дає змогу брати до уваги різні аспекти аномалій у мікросервісних системах.
- Адаптивність: вагові коефіцієнти a_1, a_2, a_3 можуть бути налаштовані для різних систем, що забезпечує гнучкість методу.
- Ефективність: інтеграційний показник аномалій (IPA) забезпечує більш точне виявлення аномалій, порівняно з використанням окремих методів.

Порівняння алгоритмів

У цій роботі порівнюються алгоритми за допомогою методу лінійної адитивної згортки. Цей підхід дає змогу інтегрувати результати кількох методів в один загальний показник з огляду на вагові коефіцієнти, призначені для кожного з методів. Такий підхід сприяє отриманню більш точного та комплексного оцінювання ефективності кожного алгоритму.

Адитивна згортка обчислюється за формулою

$$C(x) = \sum_{j=1}^{n} a_j C_j(x) \ , \qquad (12)$$

де C(x) – загальний критерій для альтернативи $x \in X$;

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

$\left(C_1(x),\ldots,C_j(x),\ldots,C_n(x)\right)$ – набір вихідних критеріїв;

n – кількість вихідних критеріїв;

a_j – нормуючий множник, що визначає вагу кожного критерію.

Найкраща альтернатива серед усіх можливих варіантів задачі визначається за допомогою такої формули:

$$x^* = \arg\max_{x \in x} C(x) \ . \qquad (13)$$

Іншими словами, результатом є альтернатива, що має найвищий показник, добутий за допомогою адитивної згортки.

Мета завдання полягає в тому, щоб вирішити багатокритеріальну проблему: визначити, який з методів статичного аналізу найкраще прогнозує можливість виявлення аномалій у мікросервісній архітектурі для знаходження способів їх попередження.

Спочатку визначимо набір альтернатив – моделей, що найчастіше використовуються в мікросервісній архітектурі, серед яких обиратимемо найбільш відповідну для цього дослідження.

Наші моделі:

– кластерний аналіз;
– регресійний аналіз;
– метод головних компонент;
– інтеграційний показник аномалій.

Дані, що використовуються для виявлення аномалій у мікросервісній архітектурі, можуть бути нестабільними. Вони часто містять аномальні, неповні, пропущені або нелінійні значення у великій кількості, адже інформація про мікросервіси отримується з різноманітних логів, метрик і трасувань. Тому оптимальний підхід має брати до уваги особливість цих даних, тобто модель має бути здатною обробляти значний обсяг вхідної інформації, яка може бути нелінійною, зважати на пропущені дані та бути стійкою до шуму. Тому критеріями вибору стали такі:

– вимоги до даних: модель має обробляти пропущені та нелінійні дані, бути стійкою до шуму;

– складність: висока, оскільки обробляються значні обсяги інформації різної структури;

– адаптивність: модель має адаптуватися до змін даних і системи;

– стійкість до шуму: висока стійкість до шумових даних;

– інтерпретація результатів: вони мають бути зрозумілими для аналізу та прийняття рішень;

– обчислювальна складність: модель має бути ефективною щодо обчислювальних ресурсів, щоб обробляти значні обсяги інформації у реальному часі;

– надійність: стійкість алгоритму до різних типів аномалій і здатність правильно їх ідентифікувати;

– масштабованість: здатність алгоритму ефективно працювати в умовах збільшення обсягів інформації та кількості мікросервісів;

– часова ефективність: швидкість виконання алгоритму в реальному часі;

– легкість у впровадженні: складність інтеграції алгоритму в уже наявну систему;

– витрати на обслуговування: ресурси, необхідні для підтримки та оновлення алгоритму.

У табл. 1 подані моделі, а також наші критерії.

Таблиця 1. *Показники для порівняння*

Критерій	**Регресійний аналіз**	**Кластерний аналіз**	**Метод головних компонент (PCA)**	**Інтеграційний показник аномалій (IPA)**
Вимоги до даних	Оброблення пропущених даних	Оброблення пропущених даних	Оброблення пропущених даних	Оброблення пропущених даних
Складність	висока	висока	висока	висока
Адаптивність	низька	середня	низька	висока
Стійкість до шуму	середня	висока	середня	висока
Інтерпретація результатів	середня	висока	середня	висока
Обчислювальна складність	середня	висока	висока	висока
Надійність	середня	висока	середня	висока
Масштабованість	середня	висока	висока	висока
Часова ефективність	висока	середня	середня	середня
Легкість у впровадженні	висока	середня	середня	низька
Витрати на обслуговування	низькі	середні	середні	високі

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Для подальшого аналізу необхідно перевести значення критеріїв у числові показники. Розглянемо кожен з них детальніше.

1. Складність моделі оцінюється за складністю алгоритмів, кількістю методів і структурою моделі. Відповідно, значення "низька", "середня" та "висока" отримують 1, 2 та 3 бали відповідно.

2. Адаптивність моделі визначається її здатністю адаптуватися до змін даних та системи. Моделі з високою адаптивністю отримують 2 бали, середньою – 1 бал, низькою – 0 балів.

3. Стійкість до шуму відтворює здатність моделі працювати із шумовими даними. Якщо модель стійка до шуму – 2 бали, середня стійкість – 1 бал, низька – 0 балів.

4. Інтерпретація результатів важлива для розуміння та аналізу. Моделі з легкою інтерпретацією отримують 2 бали, середньою – 1 бал, важкою – 0 балів.

5. Надійність визначається стійкістю до різних типів аномалій. Висока надійність отримує 2 бали, середня – 1 бал, низька – 0 балів.

6. Масштабованість оцінюється здатністю моделі працювати з великими обсягами інформації. Висока масштабованість – 2 бали, середня – 1 бал, низька – 0 балів.

7. Часова ефективність відтворює швидкість виконання алгоритму. Висока ефективність – 2 бали, середня – 1 бал, низька – 0 балів.

8. Легкість у впровадженні показує, наскільки просто інтегрувати алгоритм у систему. Висока легкість – 2 бали, середня – 1 бал, низька – 0 балів.

9. Витрати на обслуговування визначаються ресурсами, необхідними для підтримки та оновлення алгоритму. Низькі витрати – 2 бали, середні – 1 бал, високі – 0 балів.

10. Вимоги до даних передбачають оброблення пропущених і нелінійних даних. Моделі, здатні обробляти такі дані, отримують 2 бали, ті, що обробляють з обмеженнями, – 1 бал, а ті, що не обробляють, – 0 балів.

11. Обчислювальна складність визначається ефективністю використання обчислювальних ресурсів. Низька складність – 2 бали, середня – 1 бал, висока – 0 балів.

Після переведення значень критеріїв у кількісні показники можна вилучити деякі альтернативи за принципом Парето, якщо вони поступаються іншим варіантам за всіма критеріями (див. табл. 2).

Таблиця 2. *Оновлена таблиця за принципом Парето*

Критерій	Кластерний аналіз	Інтеграційний показник аномалій (IPA)
Вимоги до даних	1	1
Складність	3	3
Адаптивність	1	2
Стійкість до шуму	2	2
Інтерпретація результатів	2	2
Обчислювальна складність	0	0
Надійність	2	2
Масштабованість	2	2
Часова ефективність	1	1
Легкість у впровадженні	1	0
Витрати на обслуговування	1	1

Фінальним етапом є обчислення значень лінійної адитивної згортки для кожного варіанта із попередньо визначеним значенням нормуючого множника для всіх критеріїв (див. табл. 3).

Таблиця 3. *Результати згортки*

Модель	Показник
Кластерний аналіз	4.5
Інтеграційний показник аномалій (IPA)	7.5

Відповідно до результатів, наведених вище, бачимо, що кращою моделлю є інтеграційний показник аномалій.

Результати досліджень та їх обговорення

Розглянемо результати досліджень, що порівнюють ефективність різних методів виявлення аномалій у мікросервісах із використанням статистичного аналізу. Основну увагу було приділено трьом методам: регресійному аналізу, кластерному аналізу та методу головних компонент (PCA).

Регресійний аналіз показав свою ефективність у прогнозуванні значень метрик мікросервісів на основі історичних даних. Цей метод дає змогу виявляти аномалії з допомогою аналізу відхилень фактичних значень від прогнозованих. Регресійні моделі, особливо лінійні, відрізняються відносно низькою обчислювальною складністю та простою інтерпретацією результатів. Однак регресійний аналіз

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

потребує наявності явних трендів або залежностей у даних, що може бути обмеженням у складних і динамічних системах мікросервісів. Залежно від складності моделі адаптивність зазначеного методу може бути забезпечена завдяки оновленню моделі з новими даними, що дає змогу ефективно реагувати на зміни в системі.

Кластерний аналіз виявився дуже корисним для ідентифікації груп подій або об'єктів, що демонструють схожі патерни поведінки. Цей метод особливо ефективний для виявлення пікових навантажень і кластеризації подій за схожими характеристиками. Алгоритми кластерного аналізу, зокрема *k*-середніх та *DBSCAN*, показали високу стійкість до шуму та допомогли у виявленні як окремих аномальних об'єктів, так і відхилених груп. Водночас обчислювальна складність цього методу може бути високою, особливо для великих наборів даних. Крім того, для забезпечення адаптивності потрібне перенавчання кластерів за умови зміни даних, що може вимагати значних ресурсів.

Метод головних компонент (PCA) продемонстрував свою ефективність у зменшенні розмірності даних та виявленні основних структур і відхилень. PCA дає змогу виокремити основні напрямки варіації в даних, що допомагає ідентифікувати аномалії, які можуть бути неочевидними в оригінальному просторі ознак. Цей метод є особливо корисним для аналізу даних з високою розмірністю, де можлива кореляція між ознаками. Незважаючи на високу обчислювальну складність для великих наборів даних, PCA забезпечує інтуїтивно зрозумілі результати у вигляді головних компонент. Адаптивність методу може бути досягнута способом перевизначення головних компонент із новими даними, що дає змогу ефективно реагувати на зміни в системі.

Запропонований новий комбінований метод об'єднує переваги трьох розглянутих методів: регресійного аналізу, кластерного аналізу та PCA. Такий підхід покращує виявлення аномалій завдяки інтеграції різних аспектів аналізу даних. Комбінований метод забезпечує високу адаптивність і стійкість до шуму, об'єднуючи результати всіх трьох методів, що дає змогу ефективно реагувати на зміни в системі та забезпечувати надійність.

Порівняння кластерного аналізу та інтеграційного показника аномалій (IPA) показує, що обидва методи мають однакові вимоги до даних, високу складність, стійкість до шуму, інтерпретацію результатів, обчислювальну складність, надійність, масштабованість, часову ефективність та витрати на обслуговування. Кластерний аналіз є більш легким у впровадженні порівняно з IPA та потребує інтеграції результатів різних методів, тоді як IPA має вищу адаптивність, що є ключовою перевагою у динамічних середовищах. Обидва методи мають високу стійкість до шуму, що дає змогу ефективно працювати з даними, що містять шум. Інтерпретація результатів обох методів є зрозумілою, що допомагає легко аналізувати отримані дані. Обчислювальна складність розглянутих методів є високою, вимагає значних ресурсів для виконання аналізу. Обидва методи є надійними, забезпечують високу стійкість до різних типів аномалій і добре масштабуються, здатні працювати з великими обсягами інформації. Часова ефективність обох методів є середньою та потребує значного часу для аналізу. Витрати на обслуговування порівняних методів є середніми й вимагають певних ресурсів для підтримки та оновлення. Отже, основні розбіжності між кластерним аналізом та IPA полягають у їх адаптивності та легкості у впровадженні, що робить кожен метод привабливим для різних сценаріїв використання.

Висновки й перспективи подальшого розвитку

Зважаючи на проведене дослідження, можна зробити кілька важливих висновків. Насамперед виявлення аномалій у мікросервісних системах є критично важливим для забезпечення їх стабільної роботи та надійності. Дослідження підтвердило, що існують різні методи виявлення аномалій, зокрема регресійний аналіз, кластерний аналіз та метод головних компонент, кожен з яких має свої переваги та обмеження.

Регресійний аналіз, наприклад, ефективний у виявленні аномалій у системах з явними трендами, але може бути менш ефективним у складних і динамічних системах. Кластерний аналіз довів свою стійкість до шуму та здатність виявляти як окремі аномалії, так і групи аномальних подій, але потребує значних обчислювальних ресурсів. Метод головних компонент є потужним інструментом для аналізу високорозмірних даних, але може бути обмеженим щодо обчислень високої складності та інтерпретації результатів.

Як доповнення до вже наявних методів було запропоновано новий комбінований підхід, що передбачає впровадження регресійного аналізу, кластерного аналізу та методу головних компонент

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

(PCA). Цей підхід дає змогу об'єднувати результати трьох методів за допомогою інтеграційного показника відхилень (IPA), що забезпечує більш точне виявлення аномалій. Використання вагових коефіцієнтів a_1, a_2, a_3 дає змогу адаптувати метод під різні системи, забезпечуючи гнучкість і ефективність.

Отже, комбінований підхід має кілька ключових переваг:

– комбінований підхід: застосування трьох різних методів дає змогу взяти до уваги різні аспекти аномалій у мікросервісних системах;

– адаптивність: вагові коефіцієнти a_1, a_2, a_3 можуть бути налаштовані для різних систем, що забезпечує гнучкість методу;

– ефективність: інтеграційний показник аномалій (IPA) забезпечує більш точне виявлення аномалій, порівняно з використанням окремих методів.

З огляду на перелічені фактори рекомендується обирати метод виявлення аномалій залежно від конкретних характеристик системи та потреб користувача. Також важливо зважати на можливість комбінування різних методів для досягнення більш точних результатів.

Нарешті, розвиток методів виявлення аномалій у мікросервісних системах є актуальною та перспективною сферою досліджень, що може сприяти подальшому вдосконаленню та розширенню можливостей у цьому напрямі.

Список літератури

1. Ghani A., Ahmad S., Khan M. A., Khalid S., Zohaib S. Огляд технік виявлення відхилень для мікросервісів. *IEEE Access.* №9. 2021. P. 122766–122805. DOI: https://ieeexplore.ieee.org/document/10479425
2. Andrzejak M., Moniruzzaman M., Schumann D., Winkler S., Biffl C. Виявлення аномалій у даних веб-трафіку за допомогою статистичного навчання. *Міжнародна конференція IEEE з видобутку даних. IEEE*, 2009. P. 423–430. DOI: https://ieeexplore.ieee.org/document/6211951
3. Singh R., Paul A. Виявлення аномалій у мікросервісах за допомогою статистичного навчання. *Журнал хмарних обчислень,* 2021. 9(1), 21 p.
4. Carvalho F. D., Gama J., Rocha R. Моніторинг потоків мережевого трафіку для виявлення аномалій за допомогою однокласових опорних векторних машин. *Міжнародна конференція IEEE з видобутку даних. IEEE,* 2010. P. 77–84. DOI: https://ieeexplore.ieee.org/document/9751518
5. Alarcar J., Nguyen H. Q., Nguyen S. V., Gaber M. Виявлення аномалій в архітектурах хмарних мікросервісів. *ACM Computing Surveys.* №52. 2019. P. 1–38. DOI: https://www.sciencedirect.com/science/article/pii/S0164121223003126
6. Jiang X., Chen C., Zhou Y., Zhang J., Liu X. Виявлення аномалій для мікросервісів на основі глибокого навчання. *Міжнародна конференція IEEE з веб-сервісів (ICWS). IEEE,* 2021. P. 574–581. DOI: https://ieeexplore.ieee.org/document/9461271
7. Li Z., Zhang W., Li Z., Zhao M. Виявлення аномалій для мікросервісів. *ACM Computing Surveys.* №53. 2020. P. 1–42. DOI: https://dl.acm.org/doi/10.1145/3639478.3643535
8. He S., Jin H., Dai W., Hu X. Виявлення аномалій в системах на основі мікросервісів. *IEEE Access.* №6. 2018. P. 8459–8469. DOI: https://doi.org/10.1109/ACCESS.2018.2805848
9. Liu F. T., Ting K. M., Zhou Z. H. Outlier Detection and Classification: A Review. In: Aggarwal, C. C., Zhai, Y. (Eds.), *Outlier Analysis. Springer,* Berlin, Heidelberg, 2012. P. 1–21. DOI: https://www.sciencedirect.com/science/article/pii/S2772662223000048
10. Luo J., Wu J., Zhang Y., Sun L., Liu Y. Виявлення аномалій для мікросервісів на основі федеративного навчання. *Міжнародна конференція IEEE з веб-сервісів (ICWS). IEEE,* 2022. P. 582–589. DOI: https://ieeexplore.ieee.org/document/9461272
11. He P., Ranjan R., Nogueira J., Veiga L. M., Zhao W. Огляд виявлення аномалій для мікросервісів в хмарному обчисленні. *Журнал системного та програмного забезпечення.* № 182. 2021. 111318 p. DOI: https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-022-00296-4
12. Pimentel A., Clifton L., Clifton L., Lee Y., Kang A. (2014), Огляд технік виявлення аномалій у даних часових рядів. *Сигнальна обробка.* №92. 2014. P. 67–81. DOI: https://link.springer.com/chapter/10.1007/978-3-030-73100-7_60
13. Jiang Y., Tan K. M. C., Lam S. W., Chen P. Виявлення аномалій за допомогою однокласових опорних векторних машин у високорозмірних просторах. *Міжнародна конференція IEEE з видобутку даних.* 2017. P. 622–631. DOI: https://doi.org/10.110
14. Xu W., Huang C., Li W., Dong Z., Xu D. Виявлення аномалій в системах хмарного обчислення: огляд. *Міжнародна конференція IEEE з хмарного обчислення. IEEE,* 2014. P. 986–993. DOI: https://ieeexplore.ieee.org/document/5331755

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

15. Hawkins D. M. Викиди: виявлення аномалій у даних. Чам: *Springer International Publishing.* 2016. DOI: https://link.springer.com/10.1007/978-1-4899-7502-7_912-1
16. Patcha A., Park J.-M. Огляд технік виявлення аномалій: існуючі рішення та останні технологічні тенденції. *Комп'ютерні мережі.* №51. 2007. P. 3448–3470. DOI: https://doi.org/10.1016/j.comnet.2007.02.001
17. Thompson R., Williams L. Прикладні моделі лінійної регресії. *Sage Publications.* 2019. DOI: https://doi.org/10.4135/9781412993882
18. Kim Y., Lee H. Аналіз кластерів: концепції та практика. *Springer.* 2017.
19. Stewart D., DeCoster J. Аналіз головних компонент і пов'язані техніки. *Routledge.* 2020.
20. Yu S., Zhao L., Zhang Y. Anomaly Detection in Microservices Using Principal Component Analysis (PCA). *Visual Studio Magazine.* 2019. DOI: http://dx.doi.org/10.1109/ACCESS.2020.3044610

References

1. Ghani, A., Ahmad, S., Khan, M. A., Khalid, S., Zohaib, S. (2021), "A Survey on Anomaly Detection Techniques for Microservices". *IEEE Access.* №9. P. 122766–122805. DOI: https://ieeexplore.ieee.org/document/10479425
2. Andrzejak, M., Moniruzzaman, M., Schumann, D., Winkler, S., Biffl, C. (2009), "Detecting Anomalies in Web Traffic Data Using Statistical Learning". *International Conference on Data Mining. IEEE,* P. 423–430. DOI: https://ieeexplore.ieee.org/document/6211951
3. Singh, R., Paul, A. (2020), "Anomaly Detection in Microservices Using Statistical Learning". *Journal of Cloud Computing,* 9(1), 21.
4. Carvalho, F. D., Gama, J., Rocha, R. (2010), "Monitoring Streams of Network Traffic for Anomaly Detection Using One-Class Support Vector Machines". *In: 2010 IEEE International Conference on Data Mining. IEEE,* P. 77–84. DOI: https://ieeexplore.ieee.org/document/9751518
5. Alarcar, J., Nguyen, H. Q., Nguyen, S. V., Gaber, M. (2019), "Anomaly Detection in Cloud-Based Microservice Architectures: A Survey". *ACM Computing Surveys.* №52. P. 1–38. DOI: https://www.sciencedirect.com/science/article/pii/S0164121223003126
6. Jiang, X., Chen, C., Zhou, Y., Zhang, J., Liu, X. (2021), "Anomaly Detection for Microservices Based on Deep Learning". *In: 2021 IEEE International Conference on Web Services (ICWS). IEEE*, P. 574–581. DOI: https://ieeexplore.ieee.org/document/9461271
7. Li, Z., Zhang, W., Li, Z., Zhao, M. (2020), "Anomaly Detection for Microservices: A Survey". *ACM Computing Surveys.* №53. P. 1–42. DOI: https://dl.acm.org/doi/10.1145/3639478.3643535
8. He, S., Jin, H., Dai, W., Hu, X. (2018), "Anomaly Detection in Microservices-Based Systems: A Survey". *IEEE Access.* №6. P. 8459–8469. DOI: https://doi.org/10.1109/ACCESS.2018.2805848
9. Liu, F. T., Ting, K. M., Zhou, Z. H. (2012), "Outlier Detection and Classification: A Review". In: Aggarwal, C. C., Zhai, Y. (Eds.), *Outlier Analysis. Springer,* Berlin, Heidelberg, P. 1–21. DOI: https://www.sciencedirect.com/science/article/pii/S2772662223000048
10. Luo, J., Wu, J., Zhang, Y., Sun, L., Liu, Y. (2022), "Anomaly Detection for Microservices Based on Federated Learning". *In: 2022 IEEE International Conference on Web Services (ICWS). IEEE*, P. 582–589. DOI: https://ieeexplore.ieee.org/document/9461272
11. He, P., Ranjan, R., Nogueira, J., Veiga, L. M., Zhao, W. (2021), "A Survey on Anomaly Detection for Microservices in Cloud Computing". *Journal of Systems and Software.* №182. 111318 p. DOI: https://journalofcloudcomputing.springeropen.com/articles/10.1186/s13677-022-00296-4
12. Pimentel, A., Clifton, L., Clifton, L., Lee, Y., Kang, A. (2014), "A Review of Techniques for Detecting Anomalies in Time Series Data". *Signal Processing.* №92. P. 67–81. DOI: https://link.springer.com/chapter/10.1007/978-3-030-73100-7_60
13. Jiang, Y., Tan, K. M. C., Lam, S. W., Chen, P. (2017), "Anomaly Detection Using One-Class SVM in High-Dimensional Spaces". *IEEE International Conference on Data Mining.* P. 622–631. DOI: https://doi.org/10.1109/ICDM.2017.103
14. Xu, W., Huang, C., Li, W., Dong, Z., Xu, D. (2014), "Anomaly Detection in Cloud Computing Systems: A Survey". *In: 2014 IEEE International Conference on Cloud Computing. IEEE,* P. 986–993. DOI: https://ieeexplore.ieee.org/document/5331755
15. Hawkins, D. M. (2016), "The Outliers: Detecting Anomalies in Data". *Cham: Springer International Publishing.* DOI: https://link.springer.com/10.1007/978-1-4899-7502-7_912-1
16. Patcha, A., Park, J.-M. (2007), "An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends". *Computer Networks.* № 51. P. 3448–3470. DOI: https://doi.org/10.1016/j.comnet.2007.02.001
17. Thompson, R., Williams, L. (2019), "Applied Linear Regression Models". *Sage Publications.* DOI: https://doi.org/10.4135/9781412993882

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

18. Kim, Y., Lee, H. (2017), "Cluster Analysis: Concepts and Practice". *Springer.*
19. Stewart, D., DeCoster, J. (2020), "Principal Components Analysis and Related Techniques". *Routledge.*
20. Yu, S., Zhao, L., Zhang, Y. (2019), "Anomaly Detection in Microservices Using Principal Component Analysis (PCA)". *Visual Studio Magazine.* DOI: http://dx.doi.org/10.1109/ACCESS.2020.3044610

Надійшла (Received) 29.05.2024

Відомості про авторів / About the Authors

Перетяга Максим Юрійович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії, Харків, Україна; e-mail: maksym.peretiaha@nure.ua; ORCID ID: https://orcid.org/0000-0002-9675-1305

Peretiaha Maksym – Kharkiv National University of Radio Electronics, Postgraduate Student at the Department of Software Engineering, Kharkiv, Ukraine.

METHODS FOR DETECTING ANOMALIES IN MICROSERVICES USING STATISTICAL ANALYSIS

The **subject** of the study is methods of detecting anomalies in microservices using statistical analysis. Microservices is a popular software development architecture that allows for flexible and scalable systems. However, due to their complexity, such systems can be vulnerable to various types of anomalies that can affect their performance and reliability. The **goal** of the work is an analytical review of existing methods of detecting anomalies in microservice systems using statistical analysis methods. Detection of anomalies is critical to ensure stable system operation and quick response to possible problems. To achieve the purpose, the following **tasks** are defined: review of methods for detecting anomalies in microservices; description of the principles of regression analysis, cluster analysis and the method of principal components; comparison of methods according to the criteria of efficiency, computational complexity, resistance to noise and adaptability; recommendations for choosing a method and the possibility of combining them; summary of results and identification of directions for future research. A method for detecting anomalies in microservices is considered, which includes regression analysis, cluster analysis, and the method of principal components (PCA). The **results** of the study confirmed that each method has its advantages and limitations. Regression analysis is effective in systems with clear trends, but less effective in complex and dynamic systems. Cluster analysis has proven to be robust to noise and capable of detecting both individual anomalies and groups of anomalous events but requires significant computational resources. The method of principal components (PCA) is a powerful tool for the analysis of high-dimensional data, but it has limitations in the high complexity of calculations and interpretation of results. Each of the considered methods has its pros and cons, so the study proposed a new method that would consist in combining them. The **conclusions** emphasize the importance of statistical analysis for monitoring microservice systems. Well-chosen data analysis techniques facilitate the detection of anomalies in complex environments such as microservices. The use of regression analysis, cluster analysis and the method of principal components allows you to get a deep insight into the operation of the system. However, for best results, it is recommended to combine different methods and analyze their results in the context of a specific system. This approach provides greater resistance to anomalies and faster response to them in microservice architectures.

Keywords: anomaly detection; microservices; statistical analysis; regression analysis; clustering.

Бібліографічні описи / Bibliographic descriptions

Перетяга М. Ю. Методи виявлення аномалій у мікросервісах із використанням статистичного аналізу. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 121–132. DOI: https://doi.org/10.30837/2522-9818.2024.2.121

Peretiaha, M. (2024), "Methods for detecting anomalies in microservices using statistical analysis", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 121–132. DOI: https://doi.org/10.30837/2522-9818.2024.2.121

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 005.93:005.71-021.131
DOI: https://doi.org/10.30837/2522-9818.2024.2.133

Ю. Полупан, О. Малєєва

СИСТЕМНА МОДЕЛЬ РИЗИКІВ ТА ДЕРЕВА АЛЬТЕРНАТИВНИХ РІШЕНЬ З УДОСКОНАЛЕННЯ ЛОГІСТИЧНОГО ЛАНЦЮГА ВИРОБНИЧОГО ПІДПРИЄМСТВА

Предметом дослідження статті є процеси прийняття рішень з удосконалення елементів логістичного ланцюга (процесів постачання та збуту) виробничого підприємства в умовах невизначеності та ризиків. **Мета роботи** – зменшення часу й вартості постачання та збуту продукції виробничого підприємства з огляду на можливі ризики за допомогою прийняття раціональних управлінських рішень. У статті розв'язуються такі **завдання:** розгляд особливостей елементів логістичного ланцюга; дослідження основних проблем постачання та збуту й визначення способів їх вирішення; розроблення системної моделі логістичних ризиків виробничого підприємства; формування дерев альтернативних рішень у стратегічному управлінні ланцюгом постачання на виробничих підприємствах. Упроваджуються такі **методи:** системний підхід, методи структурної декомпозиції, ризик-орієнтований підхід. Досягнуто таких **результатів:** розглянуто особливості та виокремлено проблеми елементів логістичного ланцюга (постачання, виробництво, складування та збут); досліджено основні завдання управління ланцюгом постачання та визначено способи їх виконання; окреслено проблеми складського управління; ідентифіковано внутрішні та зовнішні ризики постачання; сформовано системну модель логістичних ризиків виробничого підприємства, основними складниками якої є логістичні проблеми, часткові ризики, їх наслідки та можливі управлінські рішення з парирування ризиків; побудовано дерева рішень для визначених проблем нестабільності постачання сировини й транспортних заторів і затримань; побудовано діаграму альтернативних рішень для ілюстративного прикладу. **Висновки.** Для подолання проблеми нестабільності постачання сировини рекомендується розвивати диверсифікацію джерел постачання, резервування запасів та використання альтернативних транспортних маршрутів. Побудова альтернативних дерев рішень у стратегічному управлінні є ефективним інструментом прийняття раціональних рішень виробничим підприємством у складних умовах невизначеності та ризиків. Вони допомагають аналізувати альтернативи та їх наслідки, щоб обрати спосіб оптимізації логістичного ланцюга.

Ключові слова: логістичний ланцюг; постачання; збут; виробниче підприємство; ризики; альтернативні рішення.

Вступ

В умовах сучасного світу, де глобалізація та швидкість постачання важливі для багатьох галузей, розвиток ефективних логістичних процесів є критично важливим для забезпечення виробництва й постачання продукції. Зростання складності високотехнологічних виробів і їх виробництва вимагає сучасних методів управління логістикою та інформаційних технологій для оптимізації процесів.

Логістика виробничого підприємства є важливим аспектом управління ланцюгом постачань і передбачає такі етапи: постачання, виробництво, складування та збут [1]. Усі перелічені етапи необхідні для нормального функціонування логістичної системи підприємства та сприяють її конкурентоспроможності на ринку. Прийняття оптимальних рішень з управління процесами на кожному із зазначених етапів підвищить ефективність управління ланцюгом постачань і забезпечить успіх діяльності підприємства.

Важливо наголосити на невизначеності та ризиках у логістичних процесах, що зумовлено нестабільністю як економічних, так і політичних факторів сьогодення. Неспроможність виявлення ризиків та ефективного управління ними є серйозним обмеженням для успішного функціонування ланцюга постачання підприємства. Недостатня увага до ідентифікації потенційних ризиків може призвести до непередбачених проблем, які надалі суттєво вплинуть на виробничий процес і конкурентоспроможність. Тому актуальним є завдання формування та аналізу альтернативних рішень з управління логістичним ланцюгом сучасного виробничого підприємства. Завдяки прийняттю раціональних рішень буде досягнуто мету дослідження – зниження часу та вартості постачання та збуту продукції виробничого підприємства, зважаючи на можливі ризики.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Аналіз останніх досліджень і публікацій

Аналіз публікацій з теми статті вказує на значний інтерес до цієї проблематики в академічному та професійному середовищах. Дослідники активно вивчають різні аспекти прийняття рішень у логістиці та їх вплив на виробничі підприємства, особливо в умовах невизначеності та ризиків.

У роботі [2] запропоновано класифікації, що дають змогу визначити основні елементи логістичних витрат. Крім того, описано інструменти, які можуть полегшити управління витратами. Автори статті [3] розглядають процеси логістичного управління виробничо-господарською організацією. Запропоновано підхід до проєктування логістичної системи, що передбачає формування організаційної структури підприємства за процесно-матричним принципом та створення ефективної логістичної служби підприємства, яка виконує роль координатора та інтегратора його бізнес-систем.

Праця [4] присвячена розробленню науково-методичного підходу до управління стійкістю каналів зв'язку в умовах розвитку промислового підприємства. Тут визначено особливості різних типів внутрішніх і зовнішніх каналів зв'язку та розроблено методики оцінювання їх стійкості.

У статті [5] емпірично досліджується вплив практик *SSCM* на динамічні можливості ланцюга постачань і продуктивність підприємства. Автори дослідження [6] проаналізували важливість протиепідемічного ланцюга постачання під час пандемії та вплив технологічних інновацій на цей процес.

У публікації [7] запропоновано систему управління організаційними ризиками. Наведено основи управління ризиками ланцюга постачань і стратегії пом'якшення. Стаття [8] розглядає вплив війни на логістичні ланцюги постачання та пропонує стратегії вдосконалення управління в умовах невизначеності та ризиків.

Методи оцінювання та управління ризиками комунікацій в транспортних проєктах запропоновано в роботі [9]. Систематизовано ризики у вигляді відносин між учасниками проєкту, формалізовано подано комунікації зацікавлених сторін з огляду на причини та можливе парирування ризиків, розроблено модель кількісного оцінювання вартості ризиків проєкту.

З проведеного огляду можна зробити висновок про актуальність вивчення логістичних ризиків виробничих підприємств. Недоліками проаналізованих методів та підходів є обмежена сфера дослідження, зосередження уваги на економічних показниках діяльності підприємства, відсутність деталізації моделей логістичних процесів і брак конкретних рекомендацій для практичної реалізації результатів. Однак вони важливі для розуміння впливу факторів ризику (зокрема воєнних конфліктів, епідемій) та стратегічних рішень на логістичні та економічні процеси, які і є причинами нестабільності та ризиків у логістичному ланцюгу.

Мета й завдання роботи

Отже, метою дослідження є підвищення якості логістичних процесів з огляду на можливі ризики в постачанні та збуті продукції на виробничих підприємствах завдяки прийняттю раціональних управлінських рішень.

У статті передбачено виконання таких завдань:

1) розгляд особливостей елементів логістичного ланцюга;

2) дослідження основних проблем постачання та збуту й визначення способів їх розв'язання;

3) розроблення системної моделі логістичних ризиків виробничого підприємства;

4) формування дерев альтернативних рішень у стратегічному управлінні ланцюгом постачання на виробничих підприємствах.

Матеріали та методи

Розглянемо особливості елементів логістичного ланцюга для виявлення можливих проблем, ризиків та їх наслідків.

Важливим моментом етапу постачання є вибір постачальника, укладення контракту, контроль якості матеріалів і вчасне доправлення. Організація та контроль розподілу забезпечує безперебійне постачання матеріалів на підприємство. Успішне завершення цього етапу забезпечує стабільність виробництва, знижує витрати та підвищує задоволеність споживачів [4].

На етапі зберігання матеріали або готова продукція утримуються відповідно до встановлених вимог і стандартів для подальшого використання або реалізації. З метою оптимізації складських процесів і мінімізації втрат важливо ефективно організувати системи управління складом і запасами [10].

На етапі збуту продукти або послуги фізично передаються або продаються споживачам за

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

допомогою різних каналів розподілу. Вирішуються завдання планування та реалізації маркетингових стратегій, встановлення політики ціноутворення, організації роздрібних операцій та управління ними, керування запасами та логістикою для забезпечення відповідних запасів і доправлення продукції, а також підтримка та обслуговування споживачів. Основна мета цього етапу – забезпечити успішне впровадження продукту чи послуги та задовольнити потреби споживачів [11].

Усі перелічені етапи важливі для нормального функціонування логістичної системи підприємства та забезпечення його конкурентоспроможності на ринку. Оптимізація кожного з них підвищить ефективність управління ланцюгом постачань і забезпечить успіх діяльності підприємства.

Розглянемо проблеми, що виникають в управлінні логістичними процесами підприємства, викликають певні ризики та вимагають відповідних стратегічних рішень.

1. Нестабільність постачання сировини є значущою проблемою для виробничих підприємств і може бути викликана різними факторами, наприклад геополітичними або економічними змінами. Ці аспекти передбачають політичні конфлікти, торговельні обмеження, зміни в законодавстві та енергетичні конфлікти. Окреслені фактори можуть призводити до обмежень у вивезенні сировини та впливати на умови постачання [12].

Погодні умови є також важливим чинником, що суттєво впливає на стабільність постачання сировини для виробничих підприємств. Це може викликати недостатність сировини та спричинити зростання цін або навіть відсутність необхідної сировини для виробництва.

Зі свого боку транспортні питання є важливим аспектом стабільності постачання сировини для підприємств. Проблеми з транспортом можуть виникати на різних етапах ланцюга постачання, зокрема: перевезення сировини від постачальників, вивезення сировини із зони добування, а також внутрішньодержавні та міжнародні перевезення до підприємства.

Низька стабільність у постачальницькому ланцюгу є серйозною проблемою для виробничих підприємств. Це може передбачати низьку надійність постачальників, зміни у власності чи управлінні, фінансові труднощі та інші чинники, що впливають на надійність та ефективність постачання сировини.

Пандемії та глобальні кризи є серйозними викликами для виробничих підприємств, що позначаються на процесах постачання та збуту продукції. Пандемії, як, наприклад, COVID-19, спричиняють обмеження виробництва, перерви в ланцюгу постачання, а також проблеми з діяльністю персоналу через введення карантинних заходів та обмежень.

Політичні чинники, такі як блокування на кордоні та оголошення воєнного стану, можуть суттєво ускладнити стабільність постачання сировини та мати серйозний вплив на виробничий процес. Блокування на кордоні, зокрема, виникає внаслідок різних політичних конфліктів або рішень уряду. Це може призвести до перерв у постачанні, затримань у доправленні сировини та обмежень у вивезенні готової продукції на зовнішні ринки.

Дія воєнного стану також може серйозно вплинути на постачання сировини. Війна, зокрема, призводить до знищення інфраструктури, перешкоджає нормальному функціонуванню транспортних маршрутів і порушує зв'язки з постачальниками. Крім того, воєнний стан може спричинити евакуацію працівників підприємства та перерви у виробництві.

Для вирішення окреслених проблем підприємства можуть упроваджувати різні методи:

– диверсифікація джерел постачання;
– резервування запасів;
– використання альтернативних транспортних маршрутів;
– розроблення бізнес-планів з огляду на ризики;
– установлення партнерських відносин із постачальниками;
– упровадження передових технологій у логістиці.

Диверсифікація джерел постачання є ключовою стратегією, що дає змогу розподіляти ризики між різними постачальниками та регіонами.

Технології прогнозування попиту дозволяють аналізувати та передбачати коливання в потребі на ринку. Розвиток альтернативних джерел сировини й використання фінансових інструментів, зокрема фіксованих контрактів чи страхування цін, також спроможні зменшити ризики від нестабільності постачання.

Загальне управління ризиками, спільно з постачальниками, може стати ефективним засобом мінімізації негативних впливів нестабільності постачання. Оцінювання ризиків і вибір відповідних стратегій вирішення є важливим складником успішного управління окресленою проблемою для забезпечення неперервного виробництва та збуту продукції в умовах мінливого середовища бізнесу [13].

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Також необхідно розглядати внутрішні ризики, що можуть серйозно позначитися на стабільності постачання сировини або її зберігання чи оброблення. Зокрема йдеться про такі види ризиків:

– технічні (збій обладнання);
– зниження якості продукції;
– нестача або низька кваліфікація персоналу;
– неефективне управління запасами;
– неякісне зберігання.

Технічні проблеми або збої обладнання можуть серйозно впливати на виробничий процес і логістичні операції підприємства. Важливо мати запасні частини та матеріали для швидкого відновлення роботи обладнання.

Ризики, пов'язані з якістю продукції, також можуть негативно позначитися на діяльності виробничого підприємства. Низька якість матеріалів та сировини спричиняють дефекти продукції та, як наслідок, утрату довіри клієнтів і ринкової позиції.

Недостатня кваліфікація персоналу може призвести до помилок в обробленні замовлень, що спричинить затримання товару та незадоволення клієнтів.

Усі ці ризики є серйозними перешкодами для діяльності підприємства. Проте за допомогою ретельного планування, аналізу та прийняття рішень з ефективного управління можна зменшити їх вплив і забезпечити безперебійну роботу логістичного ланцюга.

2. Неефективне керівництво ланцюгом постачання є серйозною проблемою, що може вплинути на ефективність і конкурентоспроможність підприємства. Зазначена проблема може бути наслідком недоліків в організації та управлінні ланцюгом постачання.

Недооцінювання або переоцінювання попиту, неправильне передбачення тенденцій ринку й недостатня увага до ризиків можуть спричинити серйозні наслідки. Неналежне планування здатне призвести до надмірних запасів або, навпаки, дефіциту товарів; обидві ситуації негативно впливають на фінансові результати підприємства. Крім того, недооцінювання ризиків, пов'язаних із ланцюгом постачання, може спричинити непередбачені перешкоди, зокрема припинення постачання сировини, катастрофічні природні події або геополітичні труднощі.

Недостатня комунікація та взаємодія між виробниками, постачальниками та дистриб'юторами, імовірно, викличе затримання в постачанні, непередбачені перерви в постачанні сировини або деталей для комплектування, а також неефективне використання ресурсів. Відсутність чітких ланок співпраці може призвести до простою обладнання та втрати прибутковості.

Низька гнучкість і неналежна реактивність управління ланцюгом постачання є важливим та потенційно шкідливим аспектом для підприємства. Гнучкість є ключовою для успішного пристосування середовища, що швидко змінюється, а реактивність визначає здатність оперативно реагувати на виклики та зміни в ланцюгу постачання [14].

Відсутність систематичного аналізу є причиною ізольованого управління окремими аспектами ланцюга постачання та недостатності загального стратегічного огляду ризиків. Брак системності у виявленні ризиків може ускладнити їх взаємозалежність та потенційний каскадний ефект на різні етапи постачального ланцюга.

3. Проблеми зі складським управлінням. Брак чіткого уявлення про рівні та місця розташування запасів може призвести до низки негативних наслідків:

1) недооцінювання чи переоцінювання обсягів потрібних запасів;

2) ускладнення виявлення проблем і ризиків у ланцюгу постачання;

3) збільшення ризику помилок і неправильних рішень.

Недооцінювання або переоцінювання запасів, імовірно, спричинить невідповідність між попитом і пропозицією, що призводить до втрат через недостатність товарів або переповнення складів. Для боротьби з цим необхідно впроваджувати ефективні системи управління запасами, використовувати передові технології прогнозування попиту й забезпечення ресурсами, а також підтримувати постійний моніторинг та аналіз запасів.

Неспроможність ефективного управління запасами може призвести до нестабільності в ланцюгу постачання, що зі свого боку викличе проблеми з виконанням замовлень, затримання в постачанні та втрату репутації підприємства [15].

Недоліки в організації простору на складі є значущим викликом для ефективного функціонування ланцюга постачання та оптимізації виробничих процесів. Відсутність оптимальної організації простору може спричинити затримання та неефективність у виборі та комплектації товарів. Незручне розташування продукції призводить до труднощів у виявленні та відстеженні запасів, ускладнює ефективність виробничих і постачальних процесів [16].

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Проблеми з безпекою та втратами на складі є важливим аспектом управління ланцюгом постачання. Недостатній контроль над безпекою може стати причиною різноманітних проблем, що впливають на якість обслуговування та фінансовий стан підприємства [17]. Крім того, порушення безпеки позначається на здоров'ї працівників, може призвести до травм і втрат робочого часу.

Недостатній контроль і неефективні процеси управління поверненнями здатні спричинити такі негативні наслідки [18]:

1) знижувати фінансові показники підприємства;

2) впливати на задоволення клієнтів та їхню лояльність;

3) породжувати проблеми в ланцюгу постачання, зокрема збільшення кількості повернених товарів.

Результати дослідження

На основні вивчення основних проблем логістичного ланцюга, з огляду на можливі наслідки та управлінські рішення, сформовано системну модель логістичних ризиків виробничого підприємства (рис. 1).

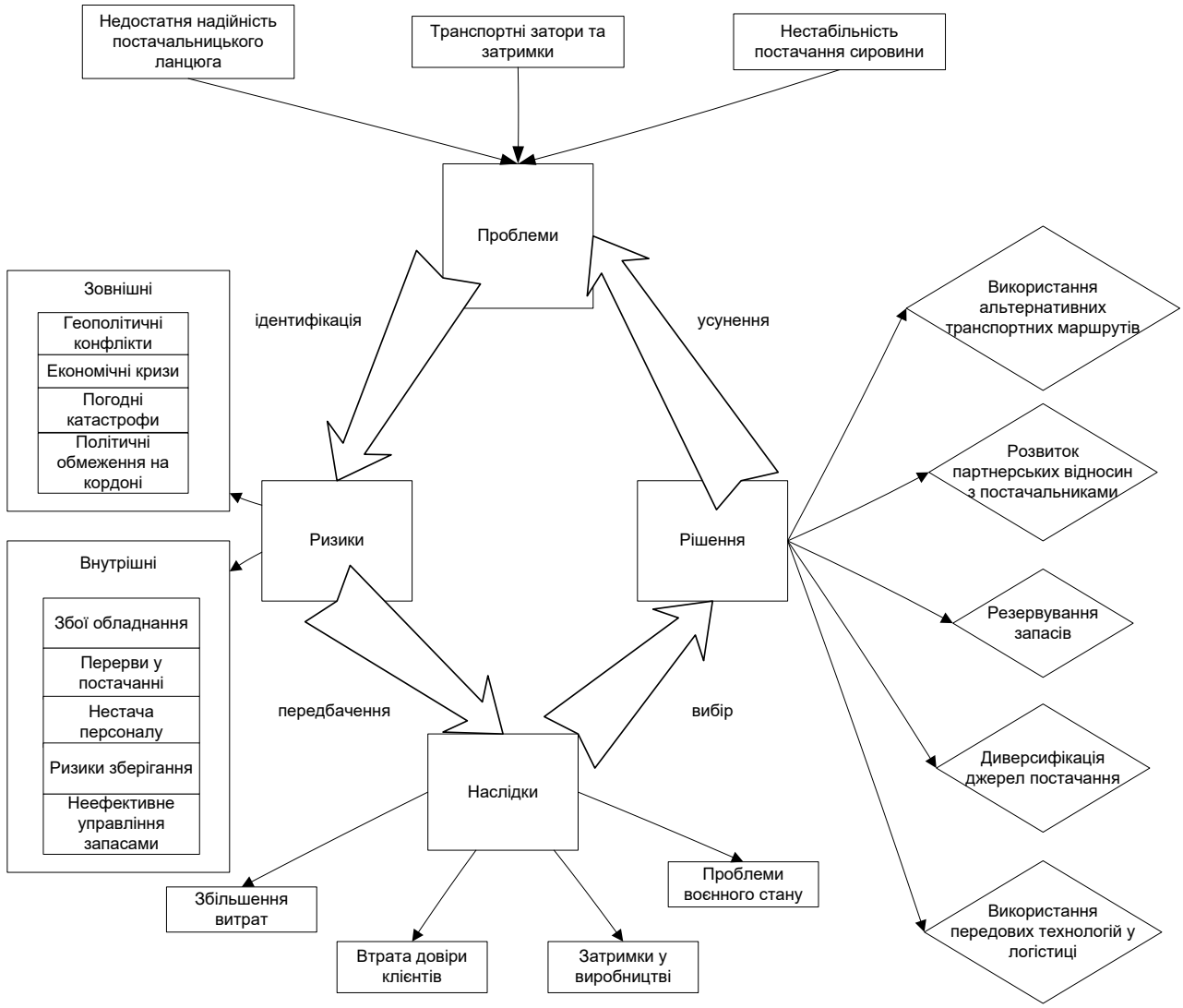


Рис. 1. Системна модель управління логістичними ризиками виробничого підприємства

Способи вирішення проблем, наведених вище, можна подати за допомогою дерева альтернативних рішень. Дерево рішень розглядає альтернативні способи, які виробниче підприємство може обрати для вирішення питань нестабільності постачання сировини (рис. 2). Початковий вузол позначає саму проблему – нестабільність постачання. Потім з'являються різні альтернативи рішень: диверсифікація джерел постачання, резервування запасів, використання альтернативних транспортних маршрутів, розроблення

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

бізнес-планів з огляду на ризики, встановлення партнерських відносин із постачальниками та впровадження передових технологій у логістиці.

Кожне альтернативне рішення має свої наслідки, які подані на гілках дерева. Для кожної альтернативи рішення розглядаються два параметри: вартість і надійність постачання. Наприклад, за умови диверсифікації джерел постачання може збільшитися вартість закупівлі, але покращиться надійність постачання завдяки розподілу ризиків між різними постачальниками. Отже, після оцінювання всіх можливих альтернатив та їх наслідків підприємство може обрати найбільш ефективне рішення для розв'язання проблеми нестабільності постачання.

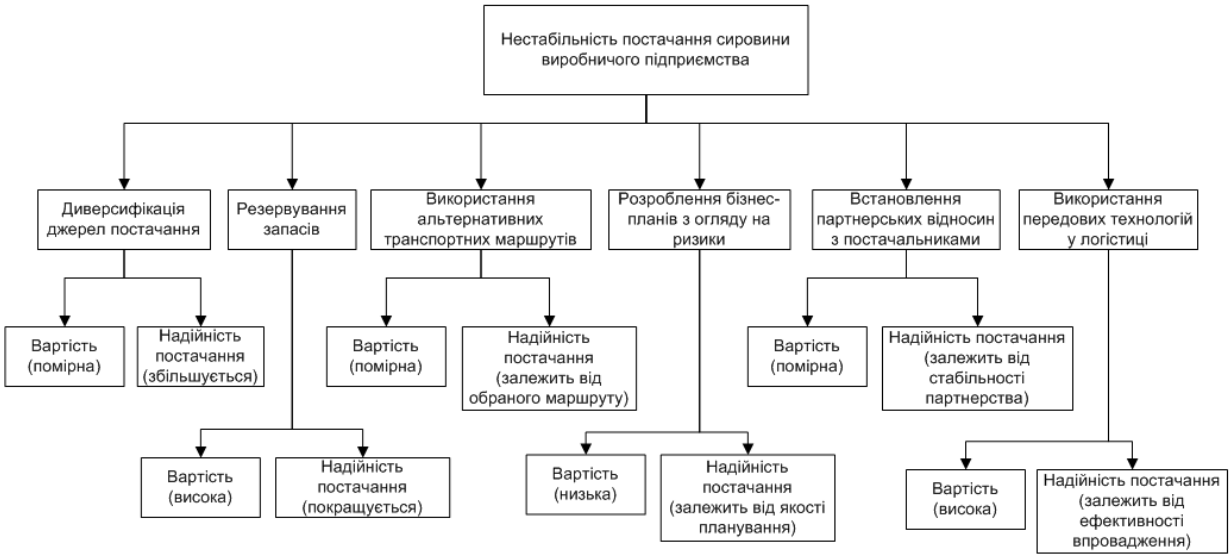


Рис. 2. Дерево альтернативних рішень для проблеми нестабільності постачання сировини

Крім першої гілки, що описує диверсифікацію джерел постачання, існують ще кілька інших альтернативних рішень. Резервування запасів може зменшити ризик нестабільності постачання, але водночас збільшить вартість утримання запасів і може викликати проблеми з обіговими коштами. Використання альтернативних транспортних маршрутів скоротить час постачання, але може бути дорожчим або менш надійним порівняно зі звичайним маршрутом. Розроблення бізнес-планів з огляду на ризики може допомогти ідентифікувати потенційні проблеми й забезпечити стратегічне планування для їх управління. Установлення партнерських відносин із постачальниками підвищує надійність постачання завдяки покращеній комунікації та спільному вирішенню проблем. Упровадження передових технологій у логістиці зменшує витрати та покращує ефективність управління постачанням, але може вимагати значних інвестицій.

Кожна з цих альтернатив має свої переваги й недоліки, що необхідно брати до уваги під час прийняття рішення щодо розв'язання проблеми нестабільності постачання.

Також можна побудувати дерево рішень для проблеми транспортних заторів і затримань, що містить певні альтернативи (рис. 3). Одна з можливих стратегій – це оптимізація маршрутів, що передбачає використання обхідних способів для уникнення транспортних заторів, але здатна збільшити вартість через додаткові витрати на дорожні послуги. Друга альтернатива – це застосування альтернативних видів транспорту, наприклад залізничного або водного, що зменшує час доправлення, але може бути дорожчим через особливості цього транспорту. Третя альтернатива – це встановлення партнерських відносин із логістичними компаніями для покращення доправлення, що, імовірно, підвищить вартість послуг, але забезпечить пріоритетні умови доправлення завдяки співпраці з партнерами. Кожна альтернатива потребує уважного аналізу вартості та часу доправлення для вибору оптимального рішення.

Розглянемо приклад з узагальненими показниками про вартість і час для кожної альтернативи, що дасть змогу візуально порівняти переваги рішень.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Оптимізація маршрутів

– вартість: +10 % до загальних витрат;

– час доправлення: збереження стандартного часу або незначне зменшення.

Використання альтернативних видів транспорту

– вартість: +15 % до загальних витрат;

– час доправлення: зменшення на 20 % стандартного часу.

Установлення партнерських відносин з логістичними компаніями

– вартість: +5 % до загальних витрат;

– час доправлення: забезпечення пріоритетних умов, що зменшує час на 10 %.

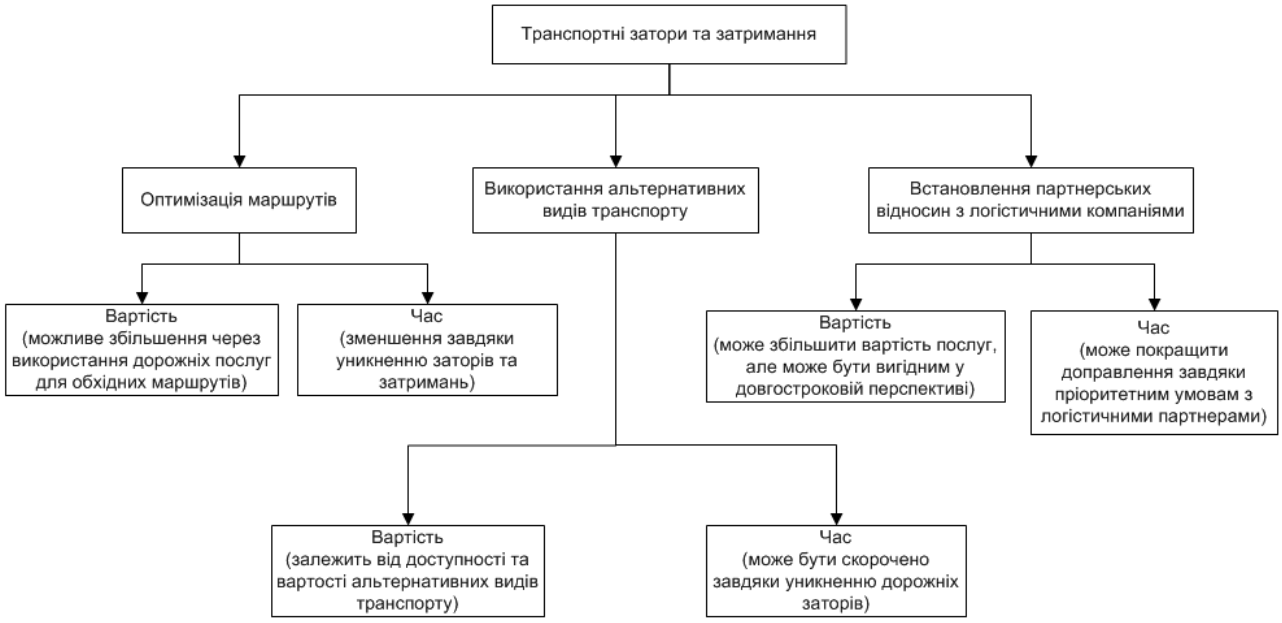


Рис. 3. Дерево альтернативних рішень проблеми транспортних заторів та затримань

На діаграмі (рис. 4) кожен стовпець подає одну з альтернатив:

– стовпець A відповідає "Оптимізації маршрутів" зі збереженням стандартного часу та додатковими витратами на 10%;

– стовпець B відповідає "Використанню альтернативних видів транспорту" зі зменшенням часу доправлення на 20% та додатковими витратами на 15%;

– стовпець C відповідає "Установленню партнерських відносин з логістичними компаніями" зі зменшенням часу доправлення на 10% та додатковими витратами на 5%.

З огляду на наведені показники підприємство має обрати оптимальну стратегію, зважаючи на баланс між вартістю та часом доправлення, що найбільше відповідає його потребам.

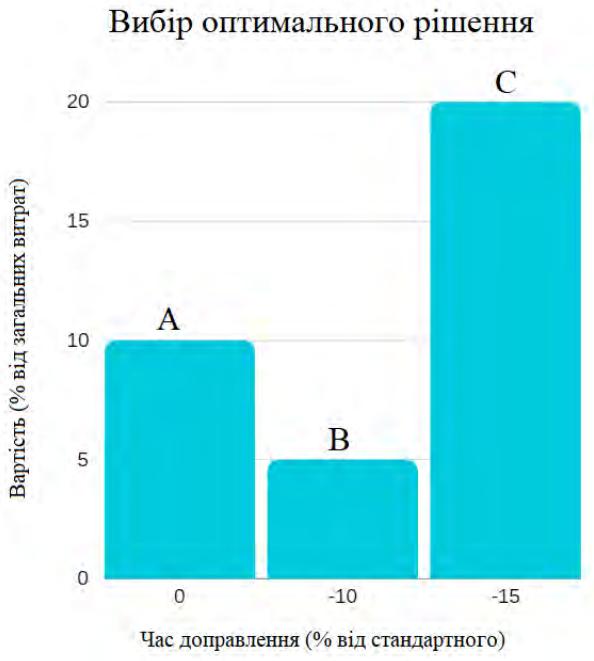


Рис. 4. Діаграма альтернатив дерева рішень

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Висновки

У статті розглянуто ключові аспекти логістичного ланцюга на виробничих підприємствах в умовах постійних викликів, таких як нестабільність постачання сировини, конкурентоспроможність, недооцінювання ризиків та інші проблеми, пов'язані з управлінням ланцюгом постачання.

Для подолання проблеми нестабільності постачання сировини рекомендується розвивати диверсифікацію джерел постачання, резервування запасів і використання альтернативних транспортних маршрутів. З метою збереження конкурентних переваг необхідно впроваджувати інновації в логістичних процесах та вдосконалювати системи управління якістю. Вирішення проблеми неефективного управління ланцюгом постачання передбачає встановлення партнерських відносин із постачальниками і застосування передових технологій у логістиці. Проблеми зі складським управлінням потребують уваги до організації простору, ефективного впровадження технологій та вдосконалення систем управління запасами.

Науковою новизною є розроблення системної моделі логістичних ризиків виробничого підприємства, у якій, на відміну від наявних моделей, сформовано дерева альтернативних рішень для проблем логістичного ланцюга. Вони допомагають аналізувати альтернативи та їх наслідки, щоб обрати найбільш ефективний спосіб управління процесами логістичного ланцюга та зменшення ризиків. Застосування запропонованої моделі в стратегічному управлінні дає змогу виробничим підприємствам підвищити ефективність логістичного ланцюга за допомогою вибору управлінських рішень (спрямованих на поліпшення параметрів часу, вартості та надійності доправлення матеріальних ресурсів і товарів) у складних умовах невизначеності та ризиків.

Напрямом подальших досліджень є розроблення моделей комунікації учасників логістичного ланцюга в процесі реалізації зазначених альтернативних рішень.

Список літератури

1. Федорович О. Є., Сломчинський О. В., Пуйденко В. О. Дослідження логістики управління виробництвом високотехнологічної продукції віртуального підприємства. *Авіаційно-космічна техніка та технологія*. 2018. № 4. С. 107 – 115. DOI: 10.32620/aktt.2018.4.13
2. Santos T. F., Gonçalves A. T. P., Leite M. S. A. Logistics cost management: insights on tools and operations. *International Journal of Logistics Systems and Management*. 2016. Vol. 23. No. 2. P. 171–188. DOI: 10.1504/IJLSM.2016.073967
3. Cherchata A., Popovychenko I., Andrusiv U., Gryn V., Shevchenko N., Shkuropatskyi O. Innovations in Logistics Management as a Direction for Improving the Logistics Activities of Enterprises. *Management Systems in Production Engineering.* 2022. Vol. 30. No.1. P. 9–17. DOI: 10.2478/mspe-2022-0002
4. Bezchasnyi O., Khobta V., Pushak Ya., Kotkalova-Litvin I., Dorovska I. Modeling of control stability of communication channels in development management conditions. *Фінансово-кредитна діяльність: проблеми теорії і практики.* 2018. № 27. С. 282–295. DOI: 10.18371/fcaptp.v4i27.154116.
5. Hong J., Zhang Y., Ding M. Sustainable supply chain management practices, supply chain dynamic capabilities, and enterprise performance. *Journal of cleaner production.* 2018. Vol. 172. P. 3508–3519. DOI: 10.1016/j.jclepro.2017.06.093
6. Malin Song, Sai Yuan, Hongguang Bo. Robust optimization model of anti-epidemic supply chain under technological innovation: learning from COVID-19. *Annals of Operations Research.* 2022. Vol. 335. P. 1332–1360. DOI: 10.1007/s10479-022-04855-5
7. Olson D. L., Wu D. Enterprise Risk Management in Supply Chains. In: *Enterprise Risk Management Models*. Springer Texts in Business and Economics. Springer, Berlin, Heidelberg. 2023. P. 1–14. DOI: 10.1007/978-3-662-68038-4_1
8. Мартинець В. Б., Кабан О. В., Полянська А. С. Оптимізація ланцюга постачання на підприємстві в умовах кризових явищ. Актуальні проблеми розвитку економіки регіону. 2022. № 18. С. 112–125. DOI: 10.15330/apred.2.18.112-127
9. Lytvynenko D., Malyeyeva O. Risk management in projects of restoration the regional transport structure on the basis of participants' commubication. Сучасний стан наукових досліджень та технологій в промисловості. 2022. № 2 (20), С. 44-51. DOI: https://doi.org/10.30837/ITSSI.2022.20.044
10. Бондарчук О. М., Темченко Г. В. Астаф'єва К. О. Використання принципів бенчмаркінгу для забезпечення підвищення ефективності діяльності. Ефективна економіка. 2021. № 3. С. 8–20. DOI: 10.32702/2307-2105-2021.3.69
11. Багорка М. О., Квасова Л. С., Кравець О. В. Формування маркетингової системи збуту продукції аграрного підприємства. *Економіка та управління підприємствами*. 2023. № 1(33). С. 14–21. DOI: 10.32782/2522-4263/2023-1-3
12. Костецька Н. І. Актуальні проблеми планування діяльності підприємств. *Сталий розвиток економіки*. 2018. № 1(38). С. 74–80.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

13. Лизунова О. М., Іщенко Я. Г., Кондрашова Г. В. Використання інноваційних методів управління персоналом підприємства. *Економіка та суспільство*. 2018. № 14. С. 448–456.
14. Григорак М. Ю. *Інтелектуалізація ринку логістичних послуг: концепції, методологія, компетентність*. Київ: Сік Груп Україна. 2017. 513с.
15. Іващук О. В. Управління запасами як складова методології керування підприємством. *Глобальні та національні проблеми економіки.* 2015. № 4. С. 404–407.
16. Костецька М. *Ефективність управління запасами на підприємстві*. Тернопіль: ТНЕУ. 2019. 358 с.
17. Sanjoy K., Ruhul S., Daryl E. Managing risk and disruption in production-inventory and supply chain systems. A review. *Journal of Industrial and Management Optimization,* 2016, No. 12(3). P. 1009–1029. DOI:10.3934/jimo.2016.12.1009
18. Шишкін В. О., Решетньова А. В. Особливості оптимізації системи управління логістичними бізнес-процесами на промислових підприємствах. *Ефективна економіка.* 2016. № 7. С. 536–540.

References

1. Fedorovych, O. E., Slomchynskyi, O. V., Puydenko, V. O. (2018), "Investigation of logistics to manage high-tech technology production of virtual enterprise" ["Doslidzhennya lohistyky upravlinnya vyrobnytstvom vysokotekhnolohichnoyi produktsiyi virtual′noho pidpryyemstva"], *Aerospace technic and technology,* No. 4, P. 107–115. DOI: 10.32620/aktt.2018.4.13
2. Santos, T. F., Gonçalves, A. T. P., Leite, M. S. A. (2016), Logistics cost management: insights on tools and operations. *International Journal of Logistics Systems and Management,* Vol. 23, No. 2, P. 171–188. DOI: 10.1504/IJLSM.2016.073967
3. Cherchata, A., Popovychenko, I., Andrusiv, U., Gryn, V., Shevchenko, N., Shkuropatskyi, O. (2022), Innovations in Logistics Management as a Direction for Improving the Logistics Activities of Enterprises, *Management Systems in Production Engineering,* Vol. 30, No.1, P. 9–17. DOI: 10.2478/mspe-2022-0002
4. Bezchasnyi, O., Khobta, V., Pushak, Ya., Kotkalova-Litvin, I., Dorovska, I. (2018), Modeling of control stability of communication channels in development management conditions, *Financial and credit activity: problems of theory and practice,* No. 27, P. 282–295. DOI: 10.18371/fcaptp.v4i27.154116
5. Hong, J., Zhang, Y., Ding, M. (2018), Sustainable supply chain management practices, supply chain dynamic capabilities, and enterprise performance, *Journal of cleaner production,* Vol. 172, P. 3508–3519. DOI: 10.1016/j.jclepro.2017.06.093
6. Malin Song, Sai Yuan, Hongguang Bo (2022), Robust optimization model of anti-epidemic supply chain under technological innovation: learning from COVID-19, *Annals of Operations Research*, Vol. 335, P. 1332–1360. DOI: 10.1007/s10479-022-04855-5
7. Olson, D. L., Wu, D. (2023), Enterprise Risk Management in Supply Chains, In: *Enterprise Risk Management Models*, Springer Texts in Business and Economics. Springer, Berlin, Heidelberg. DOI:10.1007/978-3-662-68038-4_1
8. Martynets, V. B., Kaban, O. V., Polyanska, A. S. (2022), "Optimization of the supply chain at the enterprise in the conditions of crisis phenomena" ["Optymizatsiya lantsyuha postachannya na pidpryyemstvi v umovakh kryzovykh yavyshch"]. *Actual problems of the development of the economy of the region*, 2022, No. 18, P. 112–125. DOI: 10.15330/apred.2.18.112-127
9. Lytvynenko, D., Malyeyeva, O. (2022), Risk management in projects of restoration the regional transport structure on the basis of participants' commubication, *Innovative technologies and scientific solutions for industries*, No. 2 (20), P. 44–51. DOI: https://doi.org/10.30837/ITSSI.2022.20.044
10. Bondarchuk, O. M., Temchenko, G. V. Astaf'eva, K. O. (2021), "Using the principles of benchmarking to ensure the improvement of activity efficiency" ["Vykorystannya pryntsypiv benchmarkinhu dlya zabezpechennya pidvyshchennya efektyvnosti diyal′nosti"], *Efficient economy*, No. 3, P. 8–20. DOI: 10.32702/2307-2105-2021.3.69
11. Bagorka, M.O., Kvasova, L.S., Kravets, O.V. (2023), "Formation of a marketing system for the sale of products of an agricultural enterprise" ["Formuvannya marketynhovoyi systemy zbutu produktsiyi ahrarnoho pidpryyemstva"], *Economics and enterprise management,* No. 1(33), P. 14–21. DOI: 10.32782/2522-4263/2023-1-3
12. Kostecka, N. I. (2018), "Actual problems of enterprise activity planning" ["Aktual′ni problemy planuvannya diyal′nosti pidpryyemstv"], *Sustainable economic development*, No. 1(38), P. 74–80.
13. Lyzunova, O. M., Ishchenko, Y. G., Kondrashova, G. V. (2018), "Use of innovative methods of enterprise personnel management" [Vykorystannya innovatsiynykh metodiv upravlinnya personalom pidpryyemstva], *Economy and society*, No. 14, P. 448–456.
14. Hryhorak, M. Yu. (2017), "*Intellectualization of the logistics services market: concepts, methodology, competence*" ["Intelektualizatsiya rynku lohistychnykh posluh: kontseptsiyi, metodolohiya, kompetentnist′"], Kyiv: Sik Group Ukraine.
15 vashchuk, O. V. (2015), "Inventory management as a component of enterprise management methodology" ["Upravlinnya zapasamy yak skladova metodolohiyi keruvannya pidpryyemstvom"], *Global and national economic problems*, No. 4, P. 404–407.
16. Kostecka, M. (2019), "*Effectiveness of inventory management at the enterprise*" ["Efektyvnist′ upravlinnya zapasamy na pidpryyemstvi"], Ternopil: TNEU.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

17. Sanjoy, K., Ruhul, S., Daryl, E. (2016), Managing risk and disruption in production-inventory and supply chain systems. A review, *Journal of Industrial and Management Optimization*, No. 12(3), P. 1009–1029. DOI:10.3934/jimo.2016.12.1009
18. Shishkin, V.O., Reshetnyova, A.V. (2016), "Peculiarities of the optimization of the logistics business process management system at industrial enterprises" ["Osoblyvosti optymizatsiyi systemy upravlinnya lohistychnymy biznes-protsesamy na promyslovykh pidpryyemstvakh"], *Efficient economy,* No. 7, P. 536–540.

Надійшла (Received) 25.05.2024

Відомості про авторів / About the Authors

Полупан Юрій Володимирович – Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", аспірант кафедри комп'ютерних наук та інформаційних технологій, Харків, Україна; e-mail: yuriypolupan6@gmail.com; ORCID ID: 0009-0009-3030-8448

Малєєва Ольга Володимирівна – доктор технічних наук, професор, Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", професор кафедри комп'ютерних наук та інформаційних технологій, Харків, Україна; e-mail: o.maleyeva@khai.edu; ORCID ID: 0000-0002-9336-4182

Polupan Yuriy – National Aerospace University "Kharkiv Aviation Institute", PhD Student at the Department of Computer Science and Information Technologies, Kharkiv, Ukraine.

Malyeyeva Olga – Doctor of Sciences (Engineering), Professor, National Aerospace University "Kharkiv Aviation Institute", Professor at the Department of Computer Science and Information Technologies, Kharkiv, Ukraine.

SYSTEM MODEL OF RISKS AND TREES OF ALTERNATIVE SOLUTIONS FOR IMPROVING THE LOGISTICS CHAIN AT A MANUFACTURING ENTERPRISE

The subject of this article is the decision-making processes for improving the elements of the supply chain of a manufacturing enterprise under conditions of uncertainty and risks. The purpose of the research is to reduce the time and cost of in supply and distribution of products of manufacturing enterprises by considering possible risks through making rational management decisions. The article addresses the following tasks: examination of the characteristics of supply chain elements; investigation of the main problems of supply and distribution management and identification of ways to solve them; development of a systemic model of logistic risks for a manufacturing enterprise; and formation of decision trees for alternative decisions in strategic supply chain management at manufacturing enterprises. **The methods** applied include a systems approach, structural decomposition methods, a risk-oriented approach. The following **results** were obtained: features and problems of supply chain elements such as supply, production, storage, and distribution were examined and identified; main supply chain management tasks were investigated and solutions identified; problems in warehouse management were pinpointed; internal and external supply risks were identified; a systemic model of logistic risks for a manufacturing enterprise was formed, which main components are logistic problems, partial risks, their consequences, and possible management decisions to mitigate risks; decision trees were built for problems of raw material supply instability and transportation bottlenecks and delays; an alternative decision diagram was constructed for an illustrative example. **Conclusions**: To overcome the problem of raw material supply instability, it is recommended to develop diversification of supply sources, stock reservation, and the use of alternative transport routes. Building alternative decision trees in strategic management is an effective tool for making rational decisions by manufacturing enterprises in complex conditions of uncertainty and risks. They help to analyze alternatives and their consequences to choose a path to optimize the supply chain.

Keywords: supply chain; supply; distribution; manufacturing enterprise; risks; alternative decisions.

Бібліографічні описи / Bibliographic descriptions

Полупан Ю. В., Малєєва О. В. Системна модель ризиків та дерева альтернативних рішень з удосконалення логістичного ланцюга виробничого підприємства. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 133–142. DOI: https://doi.org/10.30837/2522-9818.2024.2.133

Polupan, Y., Malyeyeva, O. (2024), "System model of risks and trees of alternative solutions for improving the logistics chain at a manufacturing enterprise", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 133–142. DOI: https://doi.org/10.30837/2522-9818.2024.2.133

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

UDC 005.93:005.71-021.131 DOI: https://doi.org/10.30837/2522-9818.2024.28.143

S. SEMENOV, S. YENHALYCHEV, M. POCHEBUT, O. SITNIKOVA

MODELS OF DATA PROCESSING AND LOGICAL ACCESS SEGREGATION CONSIDERING THE HETEROGENEITY OF ENTITIES IN INFORMATION SYSTEMS

The subject of the research is the process of logical access segregation to data in information systems. The aim of the article is to improve the accuracy and reliability of modeling processes for data processing and logical access segregation considering the heterogeneity of entities in information systems. The tasks to be solved include: conducting a comparative analysis of modern data access distribution models, integrating simpler role-based models, synthesizing hierarchical role-based models, developing enforced typing models based on trust relationships, and presenting the main provisions of the security policy integration process. The **methods** used are: systems analysis, component design, logical and simulation modeling in the form of role-based access segregation models. The **results** obtained include: development of data processing models and logical access segregation in information systems that take into account the heterogeneity of entities and the multi-level structure of information systems. The models differ from known ones by considering the heterogeneity of entities and the multi-level structure of information systems. This has increased scalability by up to 35% due to a modular approach to defining security policies. Additionally, the developed model demonstrates 25% higher implementation practicality as it easily integrates with existing access control systems and adapts to various platforms and environments. The proposed models are effective for large information systems and distributed environments due to their modularity and ability to adapt to different operational conditions. This ensures reliable access control in systems with numerous subjects and objects. The implementation of multi-level RBAC models has improved the accuracy and reliability of **results**.

Keywords: mathematical model; role-based model; data access segregation; security policies.

Introduction

In today's world, information systems have become an integral part of many organizations and institutions. Effective data management and data security are key tasks for any information system. Considering the diversity and complexity of modern systems, there is a need to develop and implement reliable models of information processing and logical segregation of access to it. Particular attention should be paid to the multi-level LASDE (logical access segregation and distribution of entities) models, which allow avoiding incorrect information flows even when attackers control privileged accounts.

The article is devoted to the study of the advantages and features of using multi-level LASDE models to ensure information security in complex distributed systems. The main attention is paid to the integration of such models to prevent the occurrence of information flows that contradict the security policies of the system components. The conditions for the correct integration of various information systems based on LASDE models and the corresponding trust relations between the subjects of these systems are also considered.

Literature analysis

Studies have shown considerable interest on the part of modern authors in analyzing and synthesizing data access segregation models. For example, article [1] offers a thorough review of multi-level security models and the specifics of their application in distributed systems. The authors describe in detail various approaches to data access and processing segregation, including the use of lattice structures. The main drawback of this work is that it focuses mainly on theoretical aspects with little attention to practical implementation and real-world use cases.

Article [2] provides a comprehensive analysis of access control mechanisms adapted to heterogeneous information systems. Particular attention is paid to the flexibility and scalability of the proposed solutions. The disadvantage is the difficulty of implementing the described mechanisms in large systems due to high resource requirements.

Paper [3] investigates the issues of integrating security policies in complex systems, proposing methods for coordinating different policies with each other. The main drawback is the lack of attention to dynamic

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

changes in systems, which can lead to problems with maintaining the relevance of security policies.

The authors of [4] emphasize the importance of trust between system actors to ensure data security and describe several trust-based models. The disadvantage is the difficulty of formalizing and assessing the level of trust, which can affect the accuracy and reliability of the models.

Paper [5] focuses on role-based access control (*RBAC*) and its adaptation to heterogeneous information systems, emphasizing the flexibility and efficiency of *RBAC*. Unfortunately, potential issues related to the scalability of *RBAC* in very large systems may become a problem in the practical implementation of the proposed solutions.

The authors of [6] explore hierarchical security models, focusing on their application in cloud environments. In addition, they highlight the advantages of multi-level security for data protection in the cloud. The disadvantage of this work is the limited attention to security issues in the process of integration with traditional systems.

Paper [7] proposes dynamic access control models for Internet of Things (IoT) systems and analyzes their ability to adapt to changes in real time. Unfortunately, the high complexity and low accuracy of the results in the context of limited resources of IoT devices hinder the practical implementation of this work.

Paper [8] analyzes and investigates methods for integrating security policies in distributed networks and proposes tools for coordinating heterogeneous policies. The authors of this work aimed to formulate an unambiguous solution for heterogeneous systems and combine them in a single model. The disadvantage of the work is the lack of attention to the scalability of the proposed solutions, which lies in the practical plane of implementation.

Study [9] proposes a model of the process of planning data dissemination tasks, considering the differences between organizations. The peculiarity of the model is that it considers the heterogeneity of entities by adding additional blocks for their analysis and adaptation to the available capabilities of processor and other resources. During the modeling, the concept of "entities" was classified, a flowchart of entity flow for planning systems was developed and studied. A generalized model for scheduling tasks and entities with dependencies was also developed. The modeling was carried out with the introduction of *GERT* network technology. As a result, we obtained *GERT* networks of the distribution task planning process for a separate *n*-th set of data types. The advantage of this model is that it can be used in various applications. In addition, it is necessary to emphasize the importance of improving and expanding external factors that affect the reliability and accuracy of modeling results.

Article [10] illustrates the results of a study of policy-oriented access control models, drawing attention to their effectiveness in distributed environments. The disadvantage is the high complexity of setting up and maintaining security policies in changing environments.

Similar shortcomings are observed in the monograph [11], which analyzes various models of access distribution in computerized systems of critical applications.

Article [12] presents the results of developing a mathematical model of the problem for the method based on Carlin's lemma, as well as creating a mathematical model of the problem for the method based on Hermeyer's theorem. Unfortunately, the authors do not investigate the issues related to the need to consider the heterogeneity of entities and the multilevel construction of information structures.

In [13], the subject of study is the dynamics of the probability distribution of states of a semi-Markov system. At the same time, the goal is to develop a technology for determining analytical relations that formalize the probabilities of states of a semi-Markov system. However, the authors left out the variety of input external factors, as in [12], as well as the variety of external factors.

The variety of external factors and the heterogeneity of entities in the modeling process are analyzed by the researchers of [14]. However, the complexity was not taken into account.

Another interesting example of mathematical modeling is [15]. Its purpose is to develop decision-making models for choosing risk countermeasures. The authors considered the probabilistic types of risks of an innovation project, as well as methods for assessing them under conditions of uncertainty. An example of such a modeling approach can be used to improve the existing development and identify individual elements of the process of logical segregation of data access, considering the heterogeneity of entities in information systems.

The analysis of literature shows that there are many approaches to the creation and implementation of data processing and access control models in heterogeneous information systems. Each of these approaches has its advantages and disadvantages, which requires careful

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

selection of the model depending on the specific requirements and operating conditions of the system. Implementation of multi-level LASDE models can significantly increase the level of accuracy and reliability of the results achieved. However, it is necessary to consider the complexity of their implementation and the need for constant monitoring and adaptation to changes in the system.

Thus, the synthesis and integration of role-based models for processing and logical segregation of data access, considering various factors, including the heterogeneity of entities in information systems, to improve the accuracy of modeling results is an important scientific task.

Main part

1. Combining the simplest role models

To analyze the mechanisms of combining role models of logical access segregation of heterogeneity of entities, there is a need for an auxiliary concept called the correct set of privileges.

Let the following sets be given in the information system A:

- P, called the set of privileges;
- R, called the set of roles;
- U, called the set of users;
- S, called the set of subjects;

and the following relationships:

- $RP \subseteq R \times P$;
- $RU \subseteq R \times U$;
- $RS \subseteq R \times S$;

and reproduction $u : S \to U$.

For any $s \in S$ and $r \in R$ the following condition is met: $(r, s) \in RS$ means that $(r, u(s)) \in RU$.

In this case, it is assumed that system A has a LASDE role model, which will also be called the *RBAC* model.

In the proposed model, for any user u, subject s, and role r, we denote:

$$R(u) = \{r \in R : (r, u) \in RU\}; \quad (1)$$

$$R(s) = \{r \in R : (r, s) \in RS\}; \quad (2)$$

$$P(r) = \{p \in P : (r, p) \in RP\}. \quad (3)$$

It is obvious that $R(s) \subseteq R(u(s))$. If the privilege $p \subseteq P(r)$, we assume that the role r has the privilege p.

Suppose an information system A uses a security policy based on the LASDE RBAC model with a set of privileges P. The set of privileges $P' \subseteq P$ is called correct if there exist roles $r_1, ..., r_n$ such that $P' = P(r_1) \cup \ldots \cup P(r_n)$.

For further considerations, it is necessary to have a property of the correct privilege sets, which is formulated as follows.

Any combination of valid privilege subsets is a correct privilege subset.

Let us prove this statement. Let $P_1, \ldots, P_n$ be correct sets of privileges, and let $P' = P_1 \cup \ldots \cup P_n$ be their combination. Let $P_1, \ldots, P_n$ be the sets of roles such that for any j $P_j = \cup r \in R_j P(r)$.

Let $R' = R_1 \cup \ldots \cup R_n$, then $P' = \cup r \in R' P(r)$. This means that P' is a correct set of privileges.

Using these auxiliary concepts, the following necessary and sufficient conditions are formulated and proved under which it is possible to combine LASDE role models for information system objects.

Let information subsystems A and B have security policies based on the LASDE RBAC model. Let system C contain objects of subsystems A and B and have a security policy based on the LASDE RBAC model. Suppose that the set of privileges of system C is $P(C) = P(A) \cup P(B)$, and the restriction of the LASDE model of system C on each of the subsystems coincides with the local LAS model of this subsystem. In this case, the combination of the LASDE models of subsystems A and B in the LASDE model of system C can be expressed by means of trust relations if and only if when for any role r_c of system C, the set of its privileges $P(r_c)$ has the form $P(r) = P_A(r) \cup P_B(r)$, where the sets of privileges $P_A(r) = P(r) \cap P(A)$ and $P_B(r) = P(r) \cap P(B)$ are correct from the standpoint of local LASDE models of systems A and B.

The proof of this statement, if "necessary," is as follows. Let $T_{A,B}$ and $T_{B,A}$ be the trust relation between systems A and B. Let SA be an arbitrary subject of system A with a single role $r_c(S_A)$. Let $P_C(S_A)$ be the set of privileges of this subject (and hence the specified role) in system C. Let $P_A(S_A) = P_C(S_A) \cap P(A)$ and $P_B(S_A) = PC(S_A) \cap P(B)$ be the restrictions of the set

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

of privileges of the subject S_A to each of the subsystems. According to the condition of the theorem, the set of privileges $P_A(S_A)$ is correct, since it is the set of privileges of the subject S_A in system A. Let $S_{B,1},...,S_{B,n}$ be the subjects of system B that trust S_A. Let $r_{B,1},...,r_{B,N}$ be all the roles of all subjects $S_{B,j}$, numbered in any order. Then the set $P_B(S_A)$ has the form $P_B(S_A)=\cup_j P(r_{B,j})$ and is the correct set of privileges of system B.

The proof of sufficiency can be formulated as follows. Given that the sets of privileges of each role of system C in each of the subsystems of this system are correct, it follows that the sets of privileges of each subject of system C in each of the subsystems are also correct. Take an arbitrary subject S_A in system A. Let $r_1,\ldots,r_n$ be the roles in system B such that $P_B(S_A)=\cup_j P(r_j)$. The set $PB(SA)=PC(SA)\cap P(B)$ is the set of privileges of system B possessed by the subject S_A. According to the condition of the theorem, such roles exist. Let's add to system B a subject S_B that has the roles $r_1,\ldots,r_n$, and no others. Suppose that in this case S_B trusts S_A. In this way, the trust relation $T_{A,B}$ is constructed. Similarly, the trust relation $T_{B,A}$ is constructed.

Thus, the necessary and sufficient conditions have been achieved that guarantee the possibility of expressing the LASDE role model of a distributed information system through the LASDE models of its components using trust relations.

2. Synthesis of hierarchical role models

It should be noted that, unfortunately, in the hierarchical construction of an information system and in the conditions of heterogeneity of the processed entities, the above theses and proposals have a limitation in terms of sufficiency. Let us prove this limitation. Let us assume that system B has three privileges P_1, P_2 and P_3 and three roles r_1, r_2 and r_3. Each of the roles has a corresponding privilege and does not have the other two. Suppose in this case $r_3 < r_2$, and the role r_1 cannot be compared with the other two. Suppose that, according to the combined LASDE model of systems A and B, subject A of system A has privileges (P_1,P_2), where P_1 is some privilege of system A. Under this condition, the set $P_C(S_A)\cap P(B)$ containing one privilege P_2 is correct. However, such a LASDE model cannot be derived from the LASDE models of systems A and B using trust relations. In fact, for entity a to have privilege P_2, there must be an entity b in system B that trusts him and has privilege P_2, and hence role r_2. However, in this case, entity b also has privilege P_3, which entity a does not have. Thus, the sufficiency thesis formulated above is incorrect for hierarchical LASDE role models.

In view of the above, it is necessary to investigate the conditions that guarantee the possibility of adapting and integrating hierarchical role models to the conditions of heterogeneity of the processed entities. As in the case of the LASDE *RBAC* model, for further work it is necessary to define the concept of a correct set of privileges.

A set of privileges $P' \subseteq P$ is called correct in the hierarchical sense if there are such roles $r_1,\ldots,r_n$, that $P=\cup_r \in R_j \cup_r < r_j' P(r)$.

Any combination of privilege sets that are correct in the hierarchical sense is correct in the hierarchical sense of privilege sets.

We can prove this statement. Let $P_1,\ldots,P_n$ be sets of privileges correct in the hierarchical sense, $P'=P_1\cup\ldots\cup P_n$ their combinations. Let $R_1,\ldots,R_n$ – such sets of roles that for any j $P_j=\cup r\in Rj\cup r'<rP(r')$. That means that P' is a hierarchically correct set of privileges.

Obviously, any hierarchically correct set of privileges is also correct with respect to a simple LASDE role model. However, the opposite statement is incorrect.

Taking into account the proposed intermediate concepts, we formulate and prove a necessary and sufficient condition under which hierarchical LASDE role models for information system objects can be combined.

Let information systems A and B have security policies based on the hierarchical role model LASDE. Let system C contain the objects of systems A and B, and also have a security policy based on the hierarchical role model LASDE. Suppose that the set of privileges of system C is $P(C)=P(A)\cup P(B)$, and the

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

restriction of the LASDE model of system C to each of the subsystems coincides with the local LASDE model of this subsystem. In this case, the combination of LASDE models of systems A and B in the LASDE model of system C can be expressed by means of trust relations if and only if when for any role r_C of system C, the set of its privileges $P(r_C)$ has the form $P(r_C) = P_A(r_C) \cup P_B(r_C)$, where the sets of privileges $P_A(r_C) = P(r_C) \cap P(A)$ and $P_B(r_C) = P(r_C) \cup P(B)$ are correct in the hierarchical sense with respect to the local LASDE models of systems A and B.

Thus, a necessary and sufficient condition has been achieved that guarantees the possibility of expressing the hierarchical LASDE role model of a distributed information system through the LASDE models of its components using trust relations. This condition is similar to the corresponding condition for LASDE *RBAC* models and is also valid in most practically used information systems.

3. Mandatory typing models based on trust relationships

Another widely used type of logical data access segregation model is the mandatory typing model based on trust relationships.

Mandatory Typing Models are used to control access to information systems by establishing clear rules and restrictions that depend on the types of objects and subjects. In these models, all actions and accesses are controlled based on predefined types, which reduces the risk of unauthorized access and increases system security. Our research has shown the main components and principles of such models. Let's list them.

1. Types of objects and subjects. All objects and subjects of the system are classified by type. Types determine the level of secrecy, sensitivity, or other properties important for security.

2. Mandatory access control. Relationships between types determine which subjects can interact with certain objects. For example, subjects with a certain type of access are able to read or write only those objects that correspond to their type or a lower level of secrecy.

3. Security policies. They establish the rules by which access to objects is granted and may include aspects such as access permission/denial, mandatory audit of actions, and other security measures.

4. Integration of policies. Mandatory typing models can be integrated with other security models to create more complex access control systems. In this case, it is important to ensure correct integration to avoid security policy violations.

5. Determinism. Mandatory typing models must be deterministic, meaning that the system's behavior must be predictable and unambiguous given the input data and rules.

6. Protection against information leaks. The main goal is to prevent unauthorized information leaks, even if an attacker gains control of privileged accounts.

Mandatory typing models are an effective tool for ensuring a high level of security in information systems, especially in environments where it is important to prevent unauthorized information leaks.

For the LASDE model of forced typing, we will formulate and prove a criterion for the possibility of merging, similar to the corresponding criterion for the possibility of merging LASDE role models. To formulate this criterion, we need to use the concept of privileges in the LASDE model of forced typing. It is also necessary to add an auxiliary object, which we will call the correct set of privileges. In this case, a set of privileges P is called correct if there exist types $t_1,\ldots,t_n$ such that all privileges of each type t_j are contained in the set P, and each privilege in P belongs to at least one of the types t_j.

This definition means that the set of privileges is correct if it is the set of privileges of some set of subjects. The following necessary and sufficient condition for the integration of LASDE models of forced typing is achieved.

Let information subsystems A and B have security policies based on the LASDE model of forced typing. Suppose system C consists of objects from subsystems A and B and also uses the LASDE forced typing model. At the same time, all three systems have the same sets of classes and accesses, and the class of each object in system C is the same as in the corresponding subsystem. Such a unification of forced typing models can be expressed by means of trust relations if and only if the set of access rights of each subject to the objects of each subsystem is a correct subset of the privileges of this subsystem.

The proof of necessity within the framework of this thesis can be as follows. Let $T(A,B)$ and $T(B,A)$ be the trust relation between systems A and B. Let S_A be any subject of system A. The set of access rights of the subject S_A to the objects of system A corresponds to

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

the set of privileges of the type of this subject and, therefore, is correct. It remains to prove the statement for the access of the subject S_A to the objects of system B. Let $S_{B1},\ldots,S_{Bu}$ be the subjects of system B that trust S_A, and let $t_1,\ldots,t_n$ be their types. Then the set of accesses of the subject S_A to the objects of system B corresponds to the combined set of privileges of these types.

The proof of sufficiency is as follows. Let us take an arbitrary subject S_A in system A. Let $t_1,\ldots,t_n$ be such types in system B that the set of access of the subject S_A to the objects of system B is the combination of the sets of privileges of these types. By the terms of the theorem, such types exist. In this case, the subject S_A must be trusted by the subjects of system B that have types $t_1,\ldots,t_n$ and no other types. Note that all types that have any privileges are domains, so for each type t_j in system B there is at least one entity that has this type. Thus, the trust relation $T(A,B)$ is built. Similarly, the trust relation $T(B,A)$ is constructed.

Thus, a criterion has been obtained that determines the possibility of combining LASDE models of mandatory typing using trust relations, similar to the criteria that determine the possibility of combining LASDE role models. This condition makes it possible to substantiate the correctness of the functioning of the mechanisms for controlling the access of subjects to remote objects of information systems, the components of which use software tools for access control, implementing the LASDE model of forced typing, considered in the future.

4. The main provisions of the process of security policies integration

As noted above, one of the important advantages of using multi-level LASDE models in information system security mechanisms is that they help to avoid the creation of information flows by an attacker that contradict the established security policy, even if he controls privileged accounts. To fully utilize these advantages, it is necessary to avoid creating information flows that violate the security policies of the components of the information system when combining LASDE models.

As a result of combining multi-level LASDE models for information system objects distributed in a network environment, special attention should be paid to measures to prevent the emergence of top-down information flows that use objects of other components. Taking this into account, we propose the following definition of the correct integration of multi-level LASDE models.

Let information systems A and B have security policies based on a multi-level LASDE model. Let system C, which contains objects of systems A and B, also have a security policy based on the LASDE multilevel model. In systems A and B, there should be no bottom-up information flows that contradict the LASDE model of the security policy of system C. In this case, system C can be considered a correct combination of systems A and B.

This statement allows us to formulate the following assumption: let systems A and B have security policies based on the LASDE multi-level model, and both systems have at least one subject at each level of secrecy. Suppose that the correct association of the multi-level LASDE models of systems A and B can be expressed through the trust relation between the subjects of systems A and B. Then the value lattices of systems A and B are isomorphic to each other, and the value lattice of system C, which is the combination of systems A and B, is also isomorphic to them.

Let us prove this statement. Let S_1 and S_2 be any subjects of system A. If in system C subjects S_1 and S_2 are at the same level of secrecy, then in system A they are also at the same level of secrecy.

Suppose that information systems A and B have security policies based on the LASDE multi-level model. Suppose that system C, which contains objects of systems A and B, also has a security policy based on the LASDE multi-level model. In systems A and B, there should be no top-down information flows that contradict the LASDE model of the security policy of system C. In this case, system C can be considered a correct combination of systems A and B.

This statement allows us to make the following assumption: let systems A and B have security policies based on the LASDE multi-level model, and both systems have at least one subject at each level of secrecy. Suppose that the correct association of the multi-level LASDE models of systems A and B can be expressed through the trust relation between the subjects of systems A and B. Then the value lattices of systems A and B are isomorphic to each other, and the value lattice of

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

system C, which is the combination of systems A and B, is also isomorphic to them.

Let us prove the statement. Let S_1 and S_2 be any subjects of system A. If in system C subjects S_1 and S_2 are at the same level of secrecy, then in system A they are also at the same level of secrecy.

Proof from the opposite. Suppose that subjects S_1 and S_2 are at different levels of secrecy in system A. Without limiting the generality, we assume that if the levels of secrecy of subjects S_1 and S_2 are comparable, then S_1 is higher than S_2. Then in system A there is an object O, read access to which is allowed to subject S_1, but denied to subject S_2. However, object O is also an object of system C, in which subjects S_1 and S_2 have the same access rights to it as in system A. Therefore, in system C, subjects S_1 and S_2 are at different levels of secrecy. This contradiction proves the statement.

Let S_1 and S_2 be subjects of system A that are at the same level of secrecy in system A. Then in the LASDE model of system C, the access rights to all objects of system B are the same for subjects S_1 and S_2.

Proof of the opposite. Let O_B be an object of system B. Let the subject S_1, in accordance with the security policy of system C, has read access to the object O_B, and the subject S_2 does not have such a right. Let O_A be an object of system A that is located at the same level of secrecy as S_1 and S_2. Subject S_1 has read access to object O_B, so in the value lattice of system C, subject S_1 is either above or at the same level as object O_B. Entity S_1 has write access to object O_A, and therefore object O_A is either above or at the same level as entity S_1. Finally, entity S_2 has read access to the O_A object. This means that S_2 is located in the value lattice of the system C either above or at the same level as the object O_A. Let the function $L(O)$ define the level of secrecy of object O in system C, then $L(O_B) < L(S_2)$. This means that S_2 must have read access to the object O_B. Similarly, we consider the situation when the subject S_1, in accordance with the security policy of system C, has write access to the object O_B, and the subject S_2 does not have such access.

Let S_1 and S_2 be subjects of system A. Then subjects S_1 and S_2 have the same level of secrecy in system A if and only if they have the same level of secrecy in system C.

Let's prove this statement. It is known that if S_1 and S_2 are at the same level of secrecy in system C, then they are at the same level of secrecy in system A. It is necessary to prove the opposite statement. Suppose that subjects S_1 and S_2 are at the same level of secrecy in system A. Then all access rights to the objects of system A are the same for subjects S_1 and S_2. However, by the previous lemma, all access rights to all objects in system B are also the same for subjects S_1 and S_2. For this reason, subjects S_1 and S_2 have the same access rights in system C, i.e., they are at the same level of secrecy in this system.

Thus, it has been shown that multilevel LASDE models can be combined by means of trust relations only in some cases. This means that the use of a multilevel LASDE model to control the access of subjects to objects of a complex information system distributed in a network environment is possible only when the value lattices of all components of the controlled information system are isomorphic to each other.

5. Comparative studies

The proposed multilevel LASDE model has a number of advantages. Fig. 1 shows diagrams comparing the main characteristics: accuracy, reliability, flexibility, scalability, and practicality of implementation.

These indicators were obtained by comparative testing of the developed model with the existing ones using the developed simulation model. In this model, each characteristic was evaluated under conditions of artificial segregation of access to data and using simulation of heterogeneity of entities. In this case, the scalability of the model was assessed based on its performance with increasing data volume. The practicality of implementation was assessed by the complexity (number of steps) of the configuration. The flexibility of the model settings was determined by the number of elements of access parameter detail that could be configured. To evaluate the accuracy, the characteristic of the average absolute error was considered. The coefficient of variation was used to assess the reliability.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

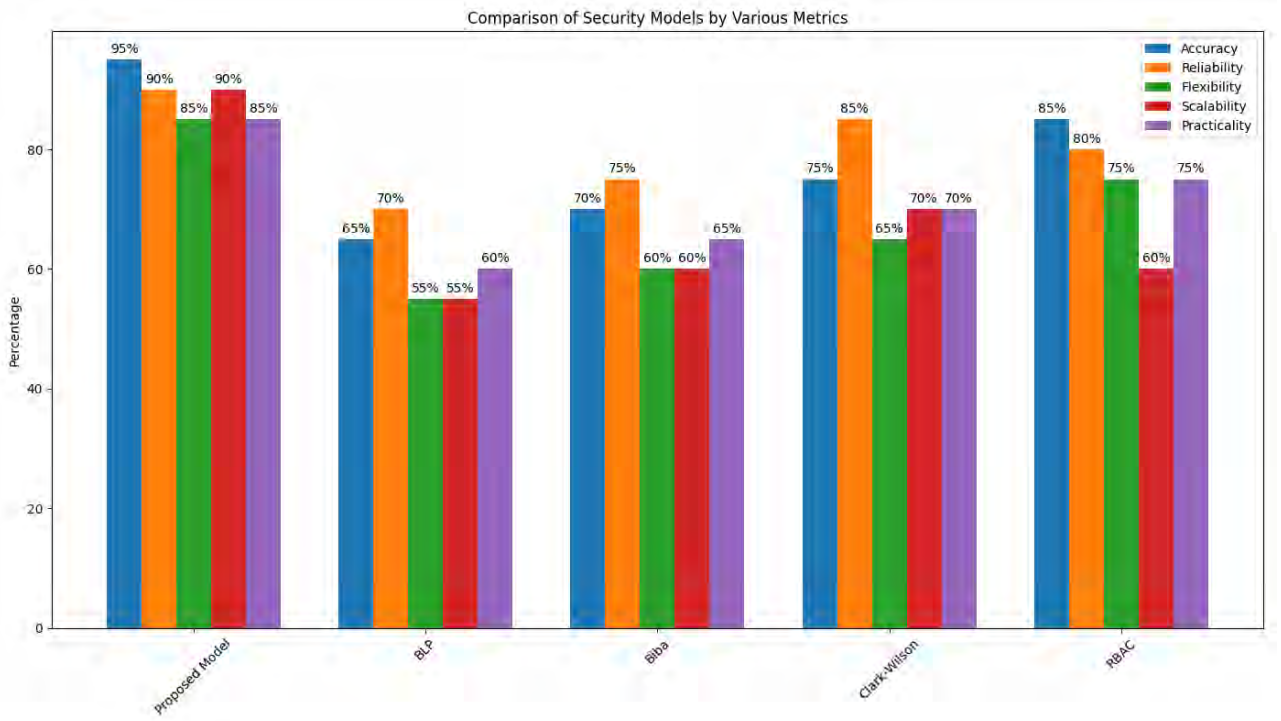


Fig. 1. Diagrams comparing the main characteristics of the developed model

As can be seen from Fig. 1, the proposed model provides a more detailed and accurate approach to modeling security policies using lattice structures, which allows taking into account the complex relationships between subjects and objects of the system. In contrast to *BLP* and *Biba*, which focus on only one aspect (confidentiality or integrity), the model proposed in this paper provides a comprehensive approach to both aspects. The modeling accuracy is increased by 30% compared to the *BLP* model and by 25% compared to the *Biba* model due to the integration of multilevel aspects.

Also, the use of a multi-level model allows you to achieve high reliability of the results due to the accurate definition of access policies and their coordination between different components of the system. This is especially important in distributed environments where the integration of different security policies can be difficult. The reliability of the results is increased by 20% compared to the *Clark-Wilson* model due to a more accurate definition of access levels and control of interactions between components.

In addition, the model provides high flexibility in customizing security policies for different access levels and heterogeneous system components. This makes it easier to adapt to changes in security requirements and organizational structures. The flexibility of the model is increased by 40% compared to *RBAC* due to the ability to customize access levels for each entity in detail.

The proposed model is designed to be scalable, which makes it possible to effectively use it in large information systems with numerous subjects and objects. Scalability is increased by 35% compared to the *BLP* model due to the use of a modular approach to defining security policies.

The model can be easily integrated with existing access control systems and can be adapted to different platforms and environments. This ensures its versatility and ease of use. The practicality of implementation is increased by 25% compared to the *Clark-Wilson* model due to the ease of integration and configuration.

Conclusion

Thus, a model for processing and logical segregation of access to data in information systems has been developed. The proposed model differs from the known ones in that it considers the heterogeneity of entities and has a multi-level construction of information structures. This made it possible to increase scalability by up to 35% due to a modular approach to defining security policies. The developed model also demonstrates a 25% higher practicality of implementation, as it can be easily integrated with

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

existing access control systems and adapted to different platforms and environments.

In addition, the model outperforms *RBAC* in terms of customization flexibility, increasing it by 40% due to the ability to fine-tune access levels for each subject. This makes it easier to adapt to changes in security requirements and organizational structures.

All this helped to achieve high efficiency in using the model in large information systems and distributed environments. The proposed model is effective for distributed systems due to its modularity and ability to adapt to different operating conditions. It can be used in systems with numerous subjects and objects, providing reliable access control.

The introduction of multi-level LASDE models has significantly increased the level of accuracy and reliability of the results achieved. However, it is necessary to consider the complexity of their implementation and the need for constant monitoring and adaptation to changes in the system, as well as to improve and expand external factors that affect the reliability and accuracy of modeling results.

References

1. "Ming-xin Ma,guo-zhen Shi,ya-qiong Wang,hao-jie Wang,wen-wen Cheng. Multilevel secure access control policy for distributed systems. Chinese Journal of Network and Information Security", 2017, 3(8). P. 28-3-4. available at: https://www.infocomm-journal.com/cjnis/EN/10.11959/j.issn.2096-109x.2017.00184
2. Poniszewska-Maranda, A. (2010), "Conception approach of access control in heterogeneous information systems using UML". *Telecommun Syst* 45, P. 177–190. DOI: https://doi.org/10.1007/s11235-009-9243-0
3. Buccafurri, F., Angelis, V., Lazzaro, S, Pugliese, A, (2024), "Enforcing security policies on interacting authentication systems", *Computers & Security*, Vol. 140, 103771 p. DOI: https://doi.org/10.1016/j.cose.2024.103771
4. Mythili, K., Haldorai, A. (2013), "Trust management approach for secure and privacy data access in cloud computing". *International Conference on Green Computing, Communication and Conservation of Energy (ICGCE),* P. 923–927. 10.1109/ICGCE.2013.6823567
5. Singh, M., Sural, S., Vaidya, J. et al. (2021), "A Role-Based Administrative Model for Administration of Heterogeneous Access Control Policies and its Security Analysis". *Information Systems Frontiers*, DOI: https://doi.org/10.1007/s10796-021-10167-z
6. Manavi, S., Mohammadalian, S., Udzir, N., Abdullah, A. (2012), "Hierarchical Secure Virtualization Model for Cloud". *International Conference on Cyber Security, Cyber Warfare and Digital Forensic*. DOI: 10.1109/CyberSec.2012.6246117
7. Aftab, Muhammad Umar, Oluwasanmi, Ariyo, Alharbi, Abdullah, Sohaib, Osama, Nie, Xuyun, Qin, Zhiguang, Son, Ngo (2021). "Secure and dynamic access control for the internet of things (IoT) based traffic system". *PeerJ Computer Science*. DOI: 7.e471.10.7717/peerj-cs.471
8. Lewis, G., Paolo, M., Rémi, G., Victor, C., (2023), "Securing distributed systems: A survey on access control techniques for cloud, blockchain, IoT and SDN", *Cyber Security and Applications,* Volume 1, 100015 p., DOI: https://doi.org/10.1016/j.csa.2023.100015
9. Semenov, S., Lymarenko, V., Yenhalychev, S., Gavrilenko, S. (2022), "The data dissemination planning tasks process model into account the entities differentity," *12th International Conference on Dependable Systems, Services and Technologies (DESSERT),* Athens, Greece, 2022, P. 1–6, DOI: 10.1109/DESSERT58054.2022.10018695
10. Ayedh, M, Wahab, A., Idris, M. (2023), "Enhanced adaptable and distributed access control decision making model based on machine learning for policy conflict resolution in BYOD environment". *MDPI Journal*, 13, 7102 p. DOI: https://doi.org/10.3390/app13127102
11. Semenov, S., Davydov, V., Gavrilenko, S. "Data protection in computerised control systems. LAP Lambert academic publishing GmbH & Co. KG. Germany", 2014 available at: https://scholar.google.com.ua/citations?view_op=view_citation&hl=ru&user=4Vn1dBkAAAAJ&citation_for_view=4Vn1dBkAAAAJ:0izLItjtcgwC
12. Beskorovainyi, V., Kolesnyk, L., Dr. Chinwi Mgbere. (2023), "Mathematical models for determining the Pareto front for building technological processes options under the conditions of interval presentation of local criteria", I*nnovative Technologies and Scientific Solutions for Industries*, No. 2 (24), P. 16–26. DOI: https://doi.org/10.30837/ITSSI.2023.24.016
13. Raskin, L., Sira, O., Sukhomlyn, L., Korsun, R. (2021), "Development of a model for the dynamics of probabilities of states of Semi-Markov systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (17), P. 62–68. DOI: https://doi.org/10.30837/ITSSI.2021.17.062
14. Fedorovich, O., Kosenko, V., Lutai, L., Zamirets, I. (2022), "Methods and models of research of investment attractiveness and competitiveness of project-oriented enterprise in the process of creating innovative high-tech", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (21), P. 51–59. DOI: https://doi.org/10.30837/ITSSI.2022.21.051

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

15. Kosenko, V. (2019), "Models of making decisions to select the techniques for countering innovative project risks". *Advanced Information Systems*, 3(1), P. 13–18. DOI: https://doi.org/10.20998/2522-9052.2019.1.03

Надійшла (Received) 25.06.2024

Відомості про авторів / About the Authors

Семенов Сергій Геннадійович – доктор технічних наук, професор, Університет Комісії національної освіти, Краків, Польща; приватна установа "Університет науки, підприємництва та технологій", Київ, Україна; e-mail: s_semenov@ukr.net; ORCID ID: http://orcid.org/0000-0003-4472-9234

Енгаличев Сергій Олександрович – Харківський національний економічний університет ім. С. Кузнеця, аспірант, Харків, Україна; e-mail: Ser.engalichev@gmail.com; ORCID ID: https://orcid.org/0000-0001-5298-2251

Почебут Максим Валентинович – кандидат технічних наук, приватна установа "Університет науки, підприємництва та технологій", Київ, Україна; e-mail: pochebutmaxim@gmail.com; ORCID ID: http://orcid.org/0000-0002-4412-2478

Сітнікова Оксана Олександрівна – кандидат технічних наук, приватна установа "Університет науки, підприємництва та технологій", Київ, Україна; e-mail: oasitnikova11@gmail.com; ORCID ID: https://orcid.org/0000-0002-2417-8220

Semenov Serhii – Doctor of Sciences (Engineering), Professor, University of the National Education Commission, Krakow, Poland; Private Institution "University of Science, Entrepreneurship and Technology", Kyiv, Ukraine.

Yenhalychev Serhii – Simon Kuznets Kharkiv National University of Economics, PhD Student, Kharkiv, Ukraine.

Pochebut Maxim – PhD, Private Institution "University of Science, Entrepreneurship and Technology", Kyiv, Ukraine; e-mail: pochebutmaxim@gmail.com

Sitnikova Oksana – PhD, Private Institution "University of Science, Entrepreneurship and Technology", Kyiv, Ukraine.

МОДЕЛІ ОПРАЦЮВАННЯ ТА ЛОГІЧНОГО РОЗМЕЖУВАННЯ ДОСТУПУ ДО ДАНИХ З ОГЛЯДУ НА РІЗНОРІДНОСТІ СУТНОСТЕЙ В ІНФОРМАЦІЙНИХ СИСТЕМАХ

Предметом дослідження є процес логічного розмежування доступу до даних в інформаційних системах. **Мета статті** – підвищення точності та достовірності результатів моделювання процесів опрацювання та логічного розмежування доступу до даних, зважаючи на різнорідність сутностей в інформаційних системах. **Завдання**, що необхідно виконати: порівняти сучасні моделі розподілу доступу до даних; об'єднати простіші рольові моделі; синтезувати ієрархічні рольові моделі; розробити моделі примусової типізації на основі відношень довіри; запропонувати основні положення процесу об'єднання політик безпеки. **Застосовані методи**: системний аналіз, компонентне проєктування, логічне та імітаційне моделювання у вигляді рольових моделей розмежування доступу. **Досягнуті результати**: розроблено моделі опрацювання даних та логічного розмежування доступу в інформаційних системах, що беруть до уваги різнорідність сутностей та багаторівневу побудову інформаційних структур. Моделі відрізняються від відомих тим, що зважають на різнорідності сутностей та багаторівневість побудови інформаційних структур. Це дало змогу підвищити масштабованість до 35% завдяки модульному підходу до визначення політик безпеки. Також розроблена модель демонструє вищу практичність реалізації на 25%, оскільки легко інтегрується з наявними системами контролю доступу та адаптується для різних платформ і середовищ. **Висновки**. Запропоновані моделі ефективні для великих інформаційних систем і розподілених середовищ завдяки своїй модульності та здатності адаптуватися до різних умов експлуатації. Це забезпечує надійний контроль доступу в системах з численними суб'єктами та об'єктами. Упровадження багаторівневих моделей ЛРДРС підвищило рівень точності та достовірності результатів.

Ключові слова: математична модель; рольова модель; розмежування доступу до даних; політики безпеки.

Бібліографічні описи / Bibliographic descriptions

Семенов С. Г., Енгаличев С. О., Почебут М. В., Сітнікова О. О. Моделі опрацювання та логічного розмежування доступу до даних з огляду на різнорідності сутностей в інформаційних системах. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 143–152. DOI: https://doi.org/10.30837/2522-9818.2024.28.143

Semenov, S., Yenhalychev, S., Pochebut, M., Sitnikova, O. (2024), "Models of data processing and logical access segregation considering the heterogeneity of entities in information systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 143–152. DOI: https://doi.org/10.30837/2522-9818.2024.28.143

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

УДК 004.41
DOI: https://doi.org/10.30837/2522-9818.2024.2.153

І. Соловей, О. Ворочек

УПРОВАДЖЕННЯ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ В ПРОЦЕСИ АВТОМАТИЗОВАНОГО ПРОГНОЗУВАННЯ ПОКАЗНИКІВ ПРОЄКТІВ ІЗ РОЗРОБЛЕННЯ ПРОГРАМНИХ СИСТЕМ

Предметом дослідження є процеси автоматизованого прогнозування показників проєктів із розроблення програмних систем, що зазвичай підлягають оцінюванню, а також методи й моделі штучного інтелекту, які можуть бути застосовані для генерації базових шаблонів дорожніх карт і післяопераційних переліків робіт та альтернативних оцінок залежно від контексту. **Мета роботи** – дослідження можливості впровадження та ефективності методів штучного інтелекту у створенні системи для автоматизованого прогнозування альтернативних оцінок програмного продукту. У статті розв'язуються такі **завдання**: визначення етапів, пов'язаних з оцінюванням альтернатив у життєвому циклі проєкту з розроблення програмного продукту; вивчення проблем прогнозування та основних факторів, що впливають на кінцеві показники; дослідження методів прогнозування, що можуть бути впроваджені для реалізації багатоваріантного оцінювання проєкту з розроблення програмного продукту. Стаття присвячена визначенню концептуальних засад створення систем автоматизованого оцінювання та прогнозування на підставі аналізу ефективності обраних моделей машинного навчання. Застосовуються такі **методи**: оцінювання та прогнозування трудовитрат у проєктах із розроблення програмного забезпечення, машинного та глибокого навчання й оцінювання їх ефективності для вирішення проблеми прогнозування. **Досягнуті результати**: визначено концептуальні засади створення систем автоматизованого оцінювання та прогнозування на підставі аналізу ефективності обраних моделей машинного навчання, сфери застосування методів штучного інтелекту в процесі оцінювання показників проєктів із розроблення програмного забезпечення; оцінено показники продуктивності різних моделей машинного навчання за певними параметрами оцінки моделі, які характеризують точність прогнозів; запропоновано концептуальну архітектуру програмного засобу генерації дорожніх карт проєкту з використанням мовної моделі *GPT*. **Висновки**: використання методів машинного та глибокого навчання може підвищити точність прогнозів основних показників проєкту, забезпечити можливість гнучкої генерації різних альтернативних варіантів шаблонів дорожніх карт і післяопераційних переліків робіт, що зробить процес планування та управління більш ефективним і прозорим за умови високого рівня невизначеності вимог до проєкту.

Ключові слова: оцінювання проєкту; програмне забезпечення; машинне навчання; генеративні моделі.

Вступ

У наш час у сфері IT-індустрії розробляється все більше й більше проєктів, стартапів і різноманітних продуктів, але далеко не всі з них досягають успіху, тобто завершуються виведенням на ринок продукту або послуги вчасно й у межах бюджету, надають очікувану цінність і відповідають вимогам якості. Основною причиною невдачі проєктів є невідповідна оцінка основних показників розроблення і, як наслідок, вартості. Оскільки оцінка вартості програмного забезпечення є природно складною, люди часто помиляються в прогнозуванні абсолютних результатів. Жодні два проєкти не є однаковими: кожен унікальний за своєю метою, за контекстом виконання та ключовими вимогами. Розв'язання проблем під час розроблення завжди призводить до змін обсягу робіт, часу реалізації, вартості тощо. Унаслідок різноманіття ризиків у проєкті завжди будуть "невідомі", які можна виявити лише тоді, коли вони виникають.

Крім того, учасники проєктів (зацікавлені особи та сторони, причетні сторони, стейкголдери) не є однаковими. Кожний з них має власний набір знань, досвіду, цінностей, очікувань, ставлення до ризику та здатності адаптуватися. І цей контекст також змінюється як під час виконання одного проєкту, так і послідовної реалізації різних завдань.

Оцінювання – це ітеративний процес; точність оцінки коливається від дуже приблизної на етапі ідеї та до максимально точної після завершення, і одним із найважливіших завдань менеджера та інших зацікавлених осіб є отримання найбільш вірогідних оцінок на будь-якому етапі життєвого циклу. Занижений бюджет може переконати керівництво розробити нові системи, що згодом перевищать їх бюджет і не досягнуть очікуваної вигоди. Багато вартих уваги проєктів скасовуються через

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

перевитрати внаслідок неякісних і нереальних кошторисів. У такий спосіб компанія втрачає прибуток і зазнає репутаційних ризиків; неефективною стає майбутня співпраця з клієнтом, а продукт так і не виходить у використання.

З іншого боку, завищені витрати можуть переконати керівництво не розробляти потенційно корисні системи. Коли оцінювачі прогнозують нереально високі витрати, що перевищують максимально допустимі для виправдання нової системи із значними перевагами, керівництво, як правило, відмовляється це схвалювати й втрачає вигоди.

Ще одним питанням, яке має вирішуватись під час оцінювання, – це передбачення впливу різних підходів до розроблення програмного продукту на отримані оцінки основних показників проєкту.

В оцінюванні часу, вартості та обсягів робіт необхідно брати до уваги також різні можливі способи реалізації тих чи інших складників програмних систем. Так, наприклад, повторне використання компонентів може як збільшити, так і зменшити трудомісткість і вартість розроблення продукту. Отже, прогнозування застосовується для оцінювання того, якими будуть ці витрати. Прогнози ніколи не можуть бути абсолютно точними, але дають гарну уяву про те, на що чекати, якщо застосовуються відповідні техніки. Деякі техніки прогнозування менш точні, але використовуються для початку планування. Коли стає більш доступною інформація, можуть впроваджуватися більш точні техніки прогнозування. Мета цієї роботи полягає в дослідженні методів прогнозування, що можуть застосовуватися у створенні програмного забезпечення для автоматизованого прогнозування оцінок імовірних варіацій реалізації та модифікування програмного продукту.

Аналіз останніх досліджень і публікацій

Загальновідомо, що кількість ІТ-проєктів завершується, так і не досягаючи заявлених цілей, і тому вважаються неуспішними. Це пояснюється неефективним управлінням проєктом, поганим оцінюванням витрат, низькими вимогами тощо [1].

Нещодавня стаття *SaasList* показала, що один із шести ІТ-проєктів перевищує витрати на 200%, а ІТ-проєкти з бюджетом щонайменше $ 1 млн мають на 50% більшу ймовірність невдачі в досягненні бізнес-цілей [2].

Оцінки процесу інженерії програмного забезпечення відіграють важливу роль у розрахунку вартості проєкту, водночас оцінка зусиль на розроблення є найважливішим складником для визначення вартісних показників. Протягом останніх десятиліть було проведено доволі значну кількість різноманітних досліджень, що відтворюють складну природу моделі оцінювання та прогнозування ймовірних змін у досягнутих показниках завдяки варіаціям підходів до розроблення та модифікування елементів програмного продукту.

Оцінювання зусиль щодо створення програмного забезпечення (ОЗРПЗ, або *SDEE – Software Development Efforts Estimation*) – це процес, що використовують керівники проєктів або розробники програмного забезпечення для прогнозування трудовитрат, необхідних для створення системи програмного забезпечення.

Найбільш відомі методи оцінювання різних показників у проєктах, пов’язаних із розробленням програмного забезпечення, можна класифікувати таким чином:

а) *алгоритмічні методи*

– вартість конструктивної моделі (COCOMO) [3]. Це підхід до визначення вартості програмного забезпечення, що використовує математичні формули та розрахунки для оцінювання вартості проєкту. Він дає приблизну кількість необхідних зусиль, а також розклад проєкту програмного забезпечення;

– функціональний точковий аналіз [4]. Оцінює показники системи з функціонального погляду, таким чином розв’язуючи проблеми, пов’язані з технологічною залежністю в життєвому циклі розроблення. Ефективність у програмній інженерії досягається завдяки комплексному аналізу застосунків у три етапи. Перший етап аналізу функціональних точок стосується визначення форм транзакцій, які мають бути здійснені в програмних застосунках. По-друге, інженери оцінюють компоненти програмної системи. Нарешті, процес передбачає оцінювання загальних характеристик системи;

– модель Путнама [5]. Забезпечує простий і надійний спосіб прогнозування витрат на програмне забезпечення. Він упроваджується з метою розрахунку зусиль і часу, необхідних для завершення роботи над програмним забезпеченням, на основі заданого розміру проєкту;

б) *неалгоритмічні методи*

– експертне оцінювання. Під час оцінювання витрат здебільшого покладаються на експертизу та

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

досвід експерта. Це залежить від його предметних знань, а не від історичних показників. Досвідчений фахівець несе відповідальність за оцінку вартості програмного забезпечення на основі того факту, що він має достатні знання, які гарантують максимально точну оцінку вартості;

– оцінювання "згори вниз" [6]. Зазначений підхід до оцінювання витрат зосереджується на визначенні вартості проєкту на основі глобальних його властивостей загалом і використання або алгоритмічних (наприклад, модель Патнема), або неалгоритмічних методів. Потім оцінка пропорційно розбивається на різні компоненти;

– оцінювання "знизу вгору" [7]. Повна протилежність попереднього методу. У ньому визначається вартість кожного компонента програмного забезпечення, а потім кінцевий результат досягається за допомогою поєднання цих елементів для отримання загальної оцінної вартості проєкту;

– оцінювання ціни до виграшу [8]. У зазначеному підході оцінка програмного проєкту прямо пропорційна бюджету замовника. Більше уваги приділяється фінансовим можливостям клієнта, ніж функціональності програмного забезпечення, і проєкт коштує стільки, скільки замовник має на нього витратити;

в) *моделі, орієнтовані на навчання*

– штучні нейронні мережі. Це один з основних підходів, що використовуються в секторі моделей машинного навчання. Як випливає з назви, це зазвичай надихається нейронною частиною системи мозку з наміром імітувати розумний живий організм. Він складається з двох шарів – вхідного та вихідного; усередині шарів є прихований шар, що містить блоки, основною метою яких є призначення вагових коефіцієнтів для даних, які надходять із вхідних показників. Ці ваги призначаються даним випадково;

– генетичні алгоритми [9]. Визначаються як адаптивні та евристичні алгоритми пошуку, що є предметом теорії природного відбору Дарвіна. Це чи не найактивніша царина досліджень, розроблених за допомогою метаевристики, натхненної природою. Генетичний алгоритм є одним із методів програмного обчислення в процесі оцінювання вартості програмного забезпечення, до того ж його основна роль полягає в зміні певних параметрів класичних методів, таких як підхід COCOMO, для більш точного прогнозування вартості проєкту;

– нечітка логіка. Розгортається для прийняття рішень, завдяки чому її можна реалізувати з різними розмірами та можливостями, починаючи від невеликих мікроконтролерів і завершуючи розробленням програмного забезпечення на основі значних робочих станцій;

– баєсівські мережі [10]. Застосовують графічні моделі для подання наборів змінних у поєднанні з їх умовними залежностями через спрямований ациклічний графік. Іншими словами, задіяний баєсівський висновок для виконання ймовірнісних обчислень; вони спрямовані на моделювання умовної залежності під час розроблення оцінки вартості програмного забезпечення, яка зазвичай позначається ребрами в орієнтованому графі;

– регресія опорних векторів [11]. Це концепція набору пов’язаних методів навчання під наглядом, які зазвичай упроваджуються для аналізу даних у поєднанні з розпізнаними шаблонами. Зазначена модель, застосована в оцінюванні вартості програмного забезпечення, бере набір вхідних показників, а потім дає прогноз кожного вхідного;

– дерево регресії [12]. Поширений підхід, що використовується в оцінюванні вартості програмного забезпечення на основі низки факторів та їх наслідків. Модель, як правило, має форму дерева з різними внутрішніми вузлами та призначена для перевірки атрибута;

– за аналогами. Один із найефективніших підходів в оцінюванні вартості програмного забезпечення завдяки його видатній продуктивності та здатності обробляти складні набори даних. Модель використовує порівняння як основну форму предмета для зіставлення проєкту програмного забезпечення, що розглядається, з минулими проєктами, які мають попередні відомі характеристики, графік і зусилля.

Цікаво, що, незважаючи на широкий спектр методів і моделей оцінювання, нині найбільш уживаними залишаються методи оцінювання "згори вниз", "знизу вгору", за аналогами та їх різновиди.

Оцінювання за аналогами – це техніка для визначення різноманітних параметрів проєкту та показників масштабу. Цей тип оцінювання може бути корисним у разі наявності досвіду проєктів в одній сфері, який можна застосувати до іншої галузі. Параметри проєкту, які можна виміряти, передбачають його вартість, бюджет, обсяг та очікувану тривалість. Показники проєкту, які можна оцінити за допомогою цієї техніки, коливаються залежно від розміру, вагомості та складності завдання. Оцінки отримуються способом зіставлення поточної діяльності з діяльністю, яка мала місце раніше, і проведення порівнянь пропорційно цьому.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Техніка часто впроваджується для оцінювання розміру певного параметра, коли інформація щодо цього параметра в межах поточного проєкту обмежена або не доступна до пізнішої дати. Оцінювання за аналогами здебільшого є формою експертного судження, яке є найбільш надійним не лише тоді, коли попередні дії схожі на поточну діяльність фактично, але також традиційно є найбільш ефективним, коли члени команди, що готує оцінювання, мають високий рівень технологічної експертизи та здатні проводити часткові аналогії на підставі досвіду та/або загальнодоступної інформації.

Оцінювання "знизу вгору" є надзвичайно корисною технікою в управлінні проєктами, оскільки дає змогу отримати більш точний показник конкретного компонента роботи. У процесі такого оцінювання кожне завдання розбивається на більш дрібні складники. Потім розробляються індивідуальні кошториси, щоб визначити, що конкретно необхідно для задоволення вимог кожного з цих менших компонентів роботи. Оцінки для менших окремих компонентів потім агрегуються з метою розроблення більшого оцінювання для всього завдання загалом. У цьому разі оцінка для всього завдання найчастіше є набагато точнішою, оскільки дає змогу ретельно розглянути кожну меншу частину, а потім об’єднати ці ретельно продумані оцінки, а не лише отримати одну велику оцінку, яка зазвичай не настільки детально розглядатиме всі окремі компоненти завдання.

Основною проблемою окреслених підходів є те, що вони не передбачають гнучкого механізму багатоваріантного оцінювання. Якщо необхідно розглянути різні варіанти етапів виконання проєкту з розроблення програмного продукту, це призводить до необхідності проведення повного циклу оцінювання для кожного варіанта.

Щодо передбачення наслідків повторного використання коду або внесення змін (модифікацій) у процес розроблення, то найбільш уживаним залишається баєсівський підхід.

Варто зазначити, що наявні на ринку комерційні засоби оцінювання, які застосовуються в реальних проєктах зі створення програмного забезпечення, наразі майже не використовують широкі можливості відповідних моделей і методів, фактично підміняючи процес прогнозування процесом оцінювання, у найкращому випадку визначаючи певний ступінь точності показників.

Найбільш поширеним підходом в абсолютній більшості систем є впровадження методу *PERT* [13], що може бути в таких програмних продуктах, як, наприклад, *Microsoft Project* (www.microsoft.com/project). У більшості продуктів основна увага зосереджена на автоматизації побудови календарних графіків проєкту на підставі оцінок, розрахованих у ручному режимі, та на аналізі прогресу певних показників проєкту щодо запланованих значень. Прикладами таких систем є *Wrike* (www.wrike.com), *ClickUp* (www.clickup.com), *Zoho Projects* (www.zoho.com/projects/), *Jira* (www.atlassian.com/software/jira), *Trello* (www.trello.com), *Smartsheet* (www.smartsheet.com) тощо.

Проведений аналіз продемонстрував, що проблеми ефективного засобу визначення та прогнозування альтернативних оцінок основних показників проєктів, пов’язаних із розробленням програмного забезпечення, не існує.

Мета й завдання роботи

Мета цієї статті полягає в дослідженні засад створення системи для автоматизованого прогнозування альтернативних оцінок програмного продукту.

Основними завданнями є такі:

– визначення етапів, пов’язаних з оцінюванням альтернатив у життєвому циклі проєкту з розроблення програмного продукту;

– висвітлення проблем прогнозування та основних факторів, що впливають на кінцеві показники;

– дослідження методів прогнозування, що можуть бути використані для реалізації багатоваріантного оцінювання проєкту з розроблення програмного продукту.

Матеріали й методи

У цьому дослідженні припускається, що проєкт із розроблення програмного забезпечення, для якого аналізується процедура оцінювання та прогнозування показників трудовитрат, тривалості або вартості, має типовий життєвий цикл відповідно до моделі, яку запропонував Інститут управління проєктами (PMI). Ця модель містить п’ять основних фаз: ініціація, планування, виконання, продуктивність та контроль і завершення проєкту.

Аналіз основних фаз дає змогу виявити місця в життєвому циклі, що потенційно передбачають прийняття рішень за наявності кількох альтернатив і завдань, які необхідно виконати. Визначаються

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

підходи до оцінювання, що типово впроваджуються на кожному етапі, та наявні обмеження.

Оскільки кожен з етапів має свої особливості, то й алгоритмічний і математичний апарати, які можуть бути використані на кожному етапі, імовірно, відрізняються. У цьому разі аналізується множина факторів, що впливають на оцінки.

Якщо поглянути на життєвий цикл (рис. 1), то можна побачити, що принаймні три його етапи здатні передбачати альтернативне оцінювання залежно від застосованого контексту та, як наслідок, викликають потребу спрогнозувати основні показники й на їх підставі прийняти рішення щодо подальших етапів проєкту.

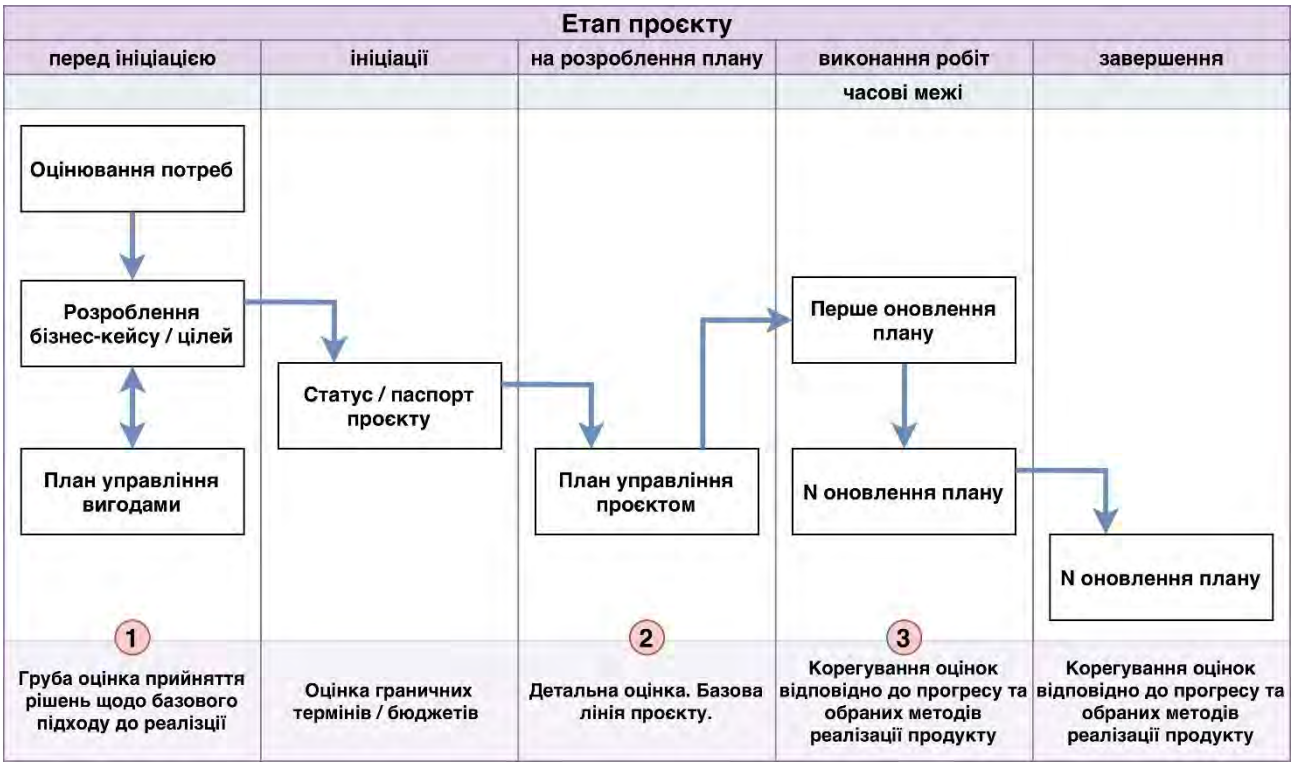


Рис. 1. Оцінювання проєкту залежно від етапу

Так, на етапі концепції (1), особливо якщо проєкт пов'язаний з новою розробкою, основною проблемою є відсутність чіткої інформації щодо структури продукту. Як наслідок, оцінки, якими оперують особи, які приймають рішення, мають низький ступінь вірогідності та точності. Фактично на цьому етапі відсутня можливість побудувати післяопераційний перелік робіт (*WBS*, *Work breakdown structure*), що призводить до неможливості застосування наявних методів оцінювання, таких як "знизу вгору". З іншого боку, для формування концепції, за умови наявності історичного досвіду виконання подібних проєктів, а також детального визначення контексту системи (як бізнесового, так і технічного), отримання прогнозів (якісних – таких що відтворюють потенційні варіанти реалізації проєкту; кількісних – прогнозних оцінок трудовитрат, вартості, часу залежно від контексту) можливе. Крім того, зважаючи на сучасний стан розвитку технологій, на цьому етапі доцільно використовувати засоби генеративного штучного інтелекту для формування альтернативних маршрутних карт проєктів з огляду на контекстні характеристики. Необхідно зауважити, що всі прогнозовані оцінки, отримані на цьому етапі, мають низьку точність, але саме вони визначають граничні значення вартості й термінів, які на етапі ініціації фіксуються в статуті проєкту.

На етапі планування (2) проводиться детальна декомпозиція робочого процесу. Природно, що у формуванні оцінок цього рівня менеджери проєктів зосереджуються на діяльності, безпосередньо пов'язаній з основними процесами програмної інженерії. Деталізація післяопераційного переліку робіт може бути різного рівня, але здебільшого в базовому плані проєкту вже береться до уваги

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

архітектура продукту та дії, які потрібно виконати для забезпечення якісного результату, – такі як робота з вимогами, проєктування, розроблення, тестування. Типово виникають питання щодо доцільності повторного використання компонентів, розроблених у минулому, або сторонніх компонентів; можливості розроблення власними силами або передавання робіт на субпідряд; призначення ресурсів на роботи з огляду на баланс продуктивності / якості / вартості тощо. Фактично особа, яка укладає детальний план, має справу з екстремально багатокритеріальним завданням, розв'язання якого на практиці зводиться до формування єдиної детальної маршрутної карти проєкту. Її якість залежить тільки від кваліфікації менеджера, який оцінює, і скоріш за все, вона не є оптимальною.

Упровадження методів машинного навчання (*ML*) на цьому етапі не тільки забезпечить можливість генерації більш точних післяопераційних переліків робіт, але й різних альтернативних декомпозицій, зважаючи на глибокий контекст і наявні припущення та обмеження.

На етапі виконання робіт, окрім проблем, властивих попередньому етапу, також додається необхідність змінити попередні оцінки та передбачити показники варіантів реалізації окремих компонентів для забезпечення вимог щодо термінів і бюджетів основних етапів проєкту. У цьому разі може бути потрібне повернення до оцінок будь-якої попередньої ітерації, уточнення прогнозів різного рівня ієрархії післяопераційного переліку робіт.

З огляду на вищезазначене у створенні програмної системи автоматизованого оцінювання та прогнозування основних показників проєкту з розроблення програмного забезпечення особлива увага має приділятися визначенню методів глибокого (*DL*) та машинного навчання (*ML*), ефективних для прогнозування обсягу зусиль, термінів виконання та вартості різних варіантів реалізації визначеного функціоналу.

Аналіз можливості застосування *ML*-методів оцінювання та прогнозування передбачав розроблення концептуального прототипу програмного продукту оцінювання показників проєктів, пов'язаних з інженерією програмного забезпечення, що надає інструменти прогнозування альтернативних оцінок багатоваріантного процесу розроблення.

Передбачалося, що базовими методами оцінювання в продукті впроваджуватимуться за аналогіями та "знизу вгору", й адаптованість до альтернатив забезпечуватиметься за допомогою використання *DL* та *ML* [14].

Основними функціональними можливостями такого продукту, пов'язаними безпосередньо з оцінюванням, мають бути:

– формування контексту проєкту (новий проєкт чи модифікація наявного, маршрутна карта розроблення продукту, елементи, реалізація яких передбачає альтернативні рішення);

– вибір технології проєкту та визначення залежностей оцінок проєкту від властивостей застосування відповідних технологій;

– вибір ролей і визначення альтернативних оцінок компетентності / продуктивності / вартості доступних людських ресурсів на кожну роль;

– визначення додаткових необхідних типів ресурсів, що відрізняються від людських;

– генерація альтернативних маршрутів проєкту та відповідних переліків робіт;

– розрахунок прогнозних оцінок для кожного з варіантів.

Існує три основних сценарії використання прототипу, пов'язаних із досліджуваними процесами: попередня генерація узагальненої структури проєкту відповідності контексту з прогнозуванням передпроєктних оцінок трудовитрат (1), вартості та термінів реалізації (2); оцінювання під час побудови базового плану проєкту (*project baseline*); оцінювання та прогнозування змін у процесі реалізації (3). *Service Blueprint* для перших двох сценаріїв наведено на рис. 2 і 3.

Контекст у межах цієї роботи – це певний набір параметрів, за допомогою яких може здійснюватися відбір відповідних наявних дорожніх карт або післяопераційних переліків робіт, а для моделювання контексту можуть використовуватися різні підходи, наприклад метаконтекстний обмін інформацією [15].

У межах цього підходу система оперує двома видами знань: контекстними та онтологічними. Семантика того чи іншого об'єкта визначається як $S(Sk, So) = So^{Sk}$, де Sk – контекстні знання про об'єкти; So – онтологічні описи. Контекстні знання такої системи описують властивості об'єктів предметної галузі, тоді як онтологічні знання описують взаємозв'язки між об'єктами та вплив цих зав'язків на формування семантики об'єктів. Фактично онтологічний складник системи формально описує знання про можливі схеми концептуального опису дорожніх карт / структур *WBS* / опису робіт.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

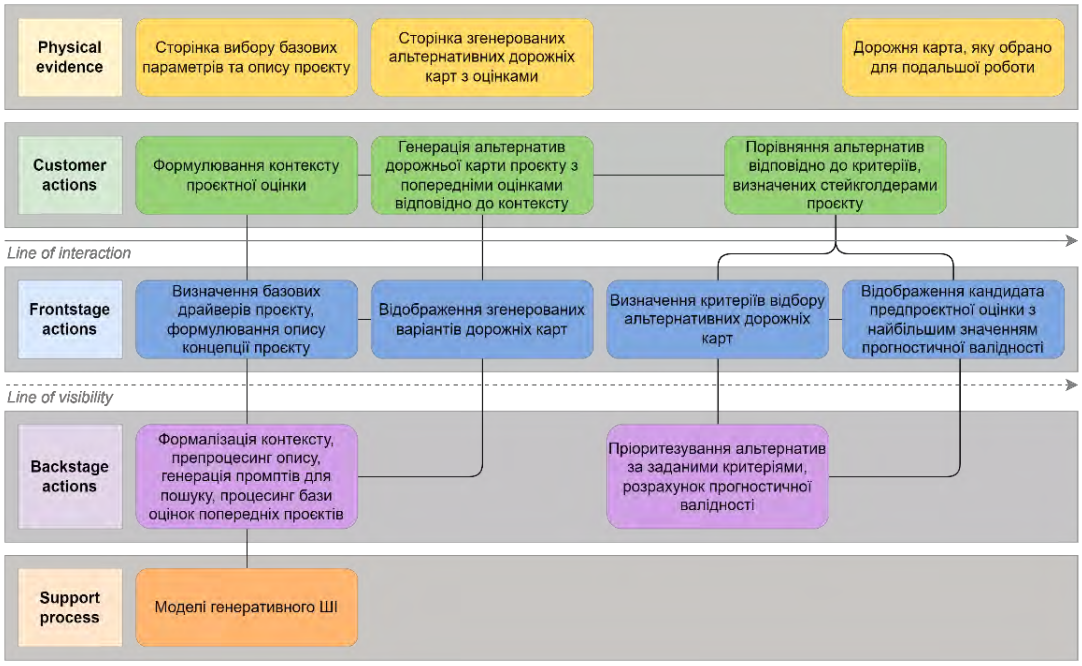


Рис. 2. Сценарій попередньої генерації узагальненої структури проєкту та прогнозування передпроєктних оцінок

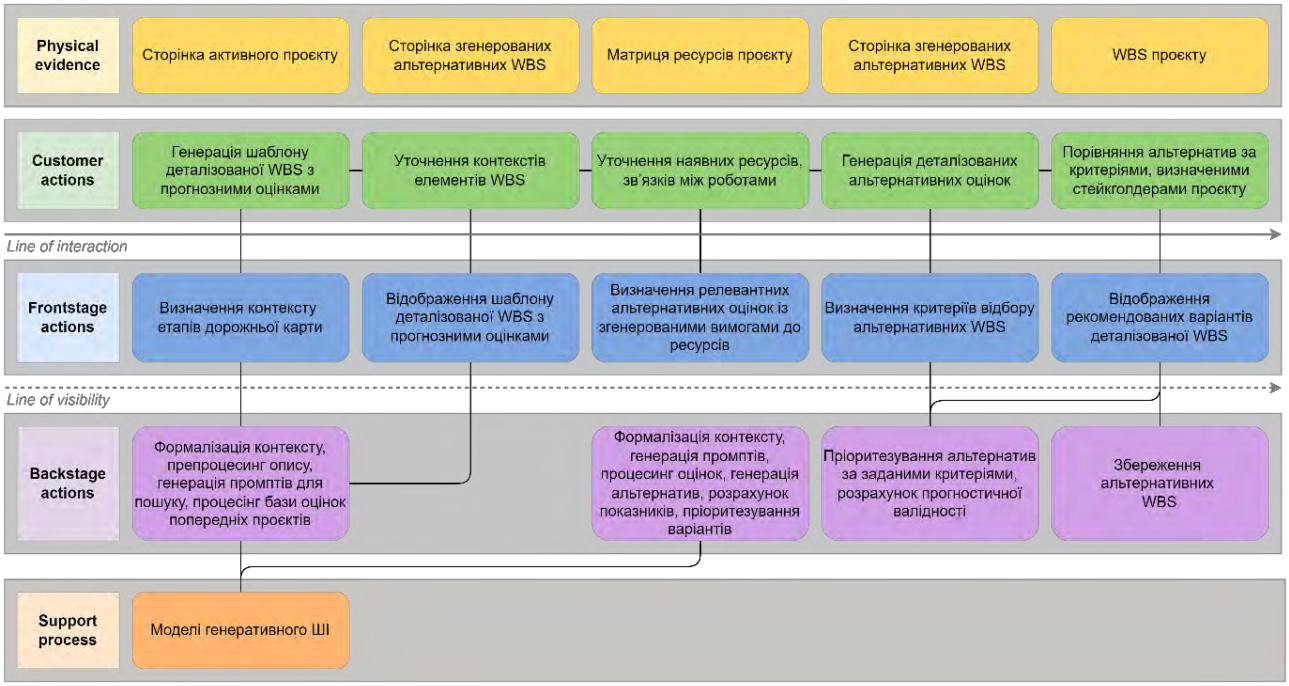


Рис. 3. Сценарій оцінювання показників у процесі побудови базового плану проєкту

Препроцесинг описів полягає в обробленні текстових визначень проєкту / продукту та у формуванні на їх основі структурованих формальних виразів, що уточнюють параметри контексту.

Щодо процесингу бази попередніх проєктів у дослідженні передбачається, що система зберігає попередні оцінки в модульному вигляді, супроводжуючи їх формальними контекстами, для забезпечення можливості повторного використання як у повному вигляді, так і окремих елементів.

Можливості моделей генеративного ШІ дають змогу отримувати прогнози навіть за умов відсутності історичної бази оцінок. Для ефективного застосування модель має бути донавчена. Сучасний стан розвитку

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

сфери управління проєктами пропонує значну кількість наявних датасетів з оцінками.

Обчислення показника прогностичної валідності є також одним із важливих аспектів створення цієї системи, оскільки саме цей параметр має суттєвий вплив на прийняття рішень щодо використання того чи іншого варіанта отриманих оцінок.

Прогнозовані оцінки в цьому сценарії ґрунтуються на наявних значеннях попередніх проєктів та загальному показнику продуктивності компанії-розробника, якщо він вказаний як контекстний параметр і використовується як фактор оцінювання.

Для другого сценарію властиві дві схеми оцінювання: побудова детального плану на базі *WBS* з повністю новим кодом та WBS, частина елементів якої пов'язана з повторним використанням. У першому випадку методи й алгоритми фактично поглиблюють згенеровану карту, альтернативні оцінки можливі завдяки знанням про наявні ресурси. У другому випадку маємо справу з розгалуженням оцінок структурних елементів різного рівня та прогнозуванням кінцевих показників для кожного з варіантів.

Прогнозовані оцінки базуються на історичній інформації та обчислюються, зважаючи на коефіцієнти продуктивності, доступності та вартості потенційних виконавців для ролей проєкту.

Особливістю третього сценарію є те, що зміни в розробленні можуть потребувати перегляду оцінок (не тільки обраного елемента *WBS*, а й інших, пов'язаних із ним). Має бути спрогнозовано вплив різних варіантів доданих змін на показники проєкту для поточної структури, а також наново обчислено майбутні етапи з огляду на поточну продуктивність проєкту та згенеровані оцінки. На цьому етапі мають обчислюватися коефіцієнти часової та фінансової ефективності поточного процесу, а також продуктивність виконання робіт кожного з елементів післяопераційного переліку робіт як драйверів показників.

Як уже згадувалося вище, під час дослідження особливостей створення системи оцінювання та прогнозування показників проєктів із розроблення програмного забезпечення передбачалося використання методів, моделей та алгоритмів глибокого й машинного навчання, що порівнювалися за різними доступними наборами даних за допомогою восьми параметрів.

У процесі перевірки гіпотези можливості використання методів машинного навчання для автоматизованого оцінювання робіт було вирішено дослідити три типові моделі: дерева рішень, наївний баєсівський класифікатор та багатошаровий персептрон.

Дерева рішень (ДР) – це непараметричні контрольовані алгоритми навчання, що застосовуються для розв'язання проблем класифікації та регресії. Метою використання цього алгоритму може бути підготовка моделі прогнозування цільових змінних способом вивчення правил прийняття рішень, отриманих на підставі характеристик даних. Наївний баєсівський класифікатор (НБК) – це класифікатор, що на підставі теореми Баєса визначає ймовірність належності елемента вибірки до певного класу за умови припущення незалежності змінних. Багатошаровий персептрон (БШП) – це штучна нейронна мережа, контрольований алгоритм навчання. Мережа містить нейрони, розподілені по трьох шарах (вхідні, приховані та вихідні). Дані подаються з нейронів вхідного рівня, прогнозуються нейронами вихідного рівня, рівень абстракції забезпечується прихованими шарами.

У дослідженні розглядалися ці моделі як представники різних типів моделей навчання для перевірки загальної гіпотези щодо можливості їх застосування.

У статті порівнюються результати різних моделей машинного навчання на трьох загальнодоступних наборах даних за допомогою восьми параметрів. Параметри оцінювання моделі містять такі показники: середня абсолютна помилка (САП), середньоквадратична помилка (СКвП), коренева середньоквадратична помилка (КСКвП) та коефіцієнт детермінації (*R*-квадрат). Середня точність оцінки у визначенні зусиль розроблення програмного забезпечення вимірюється за допомогою середньої величини відносної похибки (СВВП), відносної середньої величини помилки (ВВСП), медіанної величини відносної помилки (МВВП) та точності прогнозування.

1. Середня абсолютна похибка (САП, англ. *MAE*) – це середня сума абсолютних похибок.

$$САП = \frac{1}{n}\sum_{i=1}^{n}\left|y - \overrightarrow{y_i}\right|, \qquad (1)$$

де $y - \overrightarrow{y_i}$ – помилка прогнозування, абсолютна помилка – модуль помилки прогнозування, а САП – середнє суми помилок. У зазначеному рівнянні n відповідає за загальну кількість даних, y_i є фактичним значенням, а $\overrightarrow{y_i}$ – передбачуваним значенням.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

2. Середньоквадратична помилка (СКвП, англ. *MSE*) – це середнє значення квадратів помилок у наборі даних.

$$СКвП = \frac{1}{n}\sum_{i=1}^{n}\left(y - \overrightarrow{y_i}\right)^2. \quad (2)$$

3. Коренева середньоквадратична помилка (КСКвП, англ. *RMSE*) – це міра стандартного відхилення передбачених помилок.

$$КСКвП = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y - \overrightarrow{y_i}\right)^2}. \quad (3)$$

4. R -квадрат – це метрика відповідності моделі актуальним даним. Це статистичний показник того, наскільки добре апроксимуються фактичні дані в контексті регресії. Формула для обчислення R -квадрата має такий вигляд:

$$R^2 = 1 - \frac{ЗалСК}{ЗагСК}, \quad (4)$$

де сума квадратів помилок між початковими значеннями y та $\overrightarrow{y_i}$ відома як залишкова сума квадратів $\left(ЗалСК\right)$. Загальна сума квадратів $\left(ЗагСК\right)$ вимірює суму квадратів помилок між оригінальним значенням y і сумою всіх y. Значення R -квадрата можуть коливатися від 0 до 1. Якщо значення R -квадрата близьке до 1 або дорівнює 1, моделі надається перевага. Якщо R -квадрат негативний, це свідчить про відсутність асоціації між даними та моделлю. Цей показник також відомий як коефіцієнт детермінації.

5. Середня величина відносної похибки (СВВП, англ. *MMRE*). Використовує значення відносної помилки (ЗВП – англ. *MRE*), щоб визначити середнє значення відносної помилки. Відносна помилка (ЗВП – англ. *MRE*) обчислюється за такою формулою:

$$ЗВП = \frac{\left|ДійсніТрудовитрати - ПрогнозованіТрудовитрати\right|}{ДійсніТрудовитрати}, \quad (5)$$

$$СВВП = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|ДійсніТрудовитрати - ПрогнозованіТрудовитрати\right|}{ДійсніТрудовитрати}\right)_i \quad (6)$$

або

$$СВПП = \frac{1}{n}\sum_{i=1}^{n}\left(ЗВП\right)_i. \quad (7)$$

6. Медіанна величина відносної помилки (МВВП – англ. *MdMRE*).

$$МВВП = медіана\left(ЗВП\right)_i. \quad (8)$$

7. Відносна середня величина помилки (ВСВП – англ. *MMER*). Відносно використовує значення помилки (ЗПВ – англ. *MER*), що обчислюється за такою формулою:

$$ЗПВ = \frac{\left|ДійсніТрудовитрати - ПрогнозованіТрудовитрати\right|}{ПрогнозованіТрудовитрати}, \quad (9)$$

$$ВСВП = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|ДійсніТрудовитрати - ПрогнозованіТрудовитрати\right|}{ПрогнозованіТрудовитрати}\right)_i \quad (10)$$

або

$$ВСВП = \frac{1}{n}\sum_{i=1}^{n}\left(ЗПВ\right)_i. \quad (11)$$

8. Точність прогнозування. Середній відсоток оцінок, що перебували в межах N % від фактичних значень.

$$Прогн\left(N\right) = \frac{КількістьВідповідних}{ЗагальнаКількість} \times 100\%. \quad (12)$$

Аналіз проводився з використанням таких датасетів, як
China (https://zenodo.org/records/268446),
Finnish (https://figshare.com/articles/dataset/Finnish_Effort_Estimation_Dataset/1334271) та
Kitchenham (https://zenodo.org/records/268457).

Результати досліджень та їх обговорення

Дослідження виявило низку проблем, пов’язаних із створенням системи автоматизованого оцінювання основних показників проєкту та прогнозування впливу на ці показники різних модифікацій як самого продукту, для якого будуються оцінки, так і альтернативних процесів у межах життєвого циклу.

Як уже згадувалося вище, важливим питанням була перевірка продуктивності різних моделей машинного навчання за певними параметрами оцінки моделі. Ці показники характеризують точність прогнозів, що можна отримати. Порівняльні результати для дерев рішень, наївного баєсівського класифікатора та багатошарового персептрона

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

наведено в табл. 1. Як бачимо, найвища точність передбачення для датасету *China* властива БШП, для *Finnish* – НБК, для *Kemerer* – ДР. Детальний аналіз досягнутих результатів продемонстрував, що показники загалом залежать від типу даних, але найбільш придатними для використання в системах прогнозування зусиль на розроблення є ймовірнісні моделі та моделі, основані на нейронних мережах.

Таблиця 1. *Показники продуктивності моделей для обраних датасетів*

Модель	Прогн (25)	Прогн (50)	САП	СВВП	ВСВП	МВВП	R -квадрат	СКвП	КСКвП
Датасет *China*									
ДР	22.66%	46.66%	0.0366	1.0713	0.5456	0.5011	0.6409	0.0065	0.0807
НБК	21.33%	47.33%	0.0415	0.9399	0.3359	0.4545	0.6239	0.0068	0.0826
БШП	27.33%	49.33%	0.0357	0.9481	0.3734	0.5025	0.7015	0.0054	0.0735
Датасет *Finnish*									
ДР	19.67%	45.90%	0.0511	1.3536	0.5668	0.5188	-0.1736	0.0123	0.1109
НБК	23.77%	45.90%	0.0362	1.2171	–2.4677	0.5560	0.6766	0.0034	0.0582
БШП	17.21%	44.26%	0.0403	1.3961	0.0086	0.5225	0.6444	0.0037	0.0611
Датасет *Kemerer*									
ДР	0%	60%	0.0801	0.2071	0.3295	0.1047	-0.1736	0.0083	0.0911
НБК	20%	40%	0.1046	0.8173	0.8792	0.4589	0.6766	0.0156	0.1247
БШП	20%	40%	0.0703	0.3848	0.3641	0.2510	0.6444	0.0057	0.0757

Дослідження також показало, що в розробленні подібних систем необхідно брати до уваги вплив сетів даних, які використовуються для навчання моделі, а саме специфіку проєктів датасету, їх тип, складність, кількість проєктів у сеті, атрибути тощо.

Поза межами дослідження також залишилися методології оцінювання, основані не на традиційних показниках зусиль (наприклад, у людино-годинах), а тих, що часто застосовуються в *Agile*-проєктах (наприклад, сторі поінт). Як наслідок, питання ефективності моделей штучного інтелекту для роботи із зазначеними типами проєктів залишається дискусійним.

Другим результатом дослідження став спроєктований для демонстрації можливості використання генеративного штучного інтелекту *GPT*-засіб, архітектура якого зображена на рис. 5.

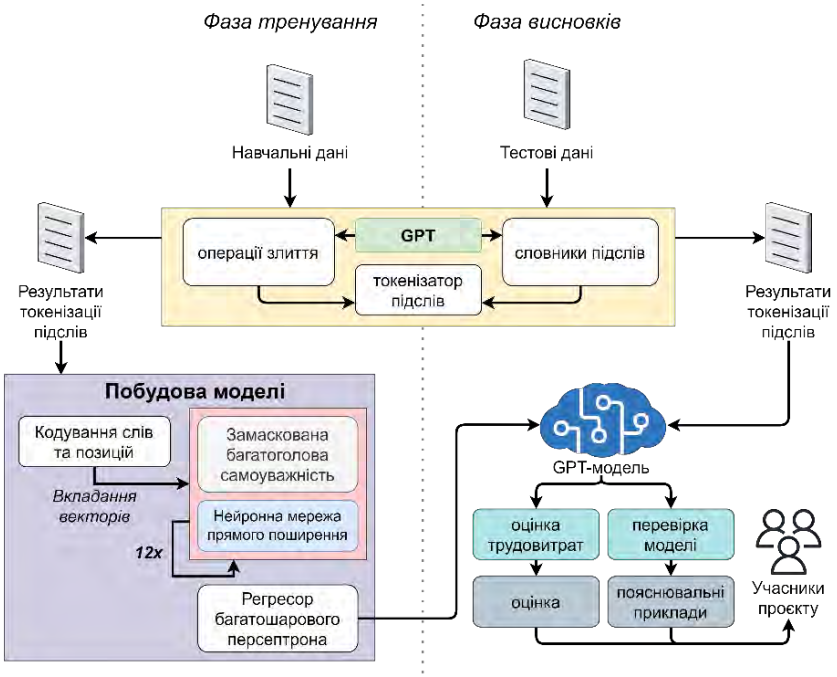


Рис. 5. Огляд архітектури програмного засобу генерації дорожніх карт і відповідних оцінок

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Концептуально програмний засіб працює таким чином. Спочатку проводиться токенізація підслів на основі попередньо навченої моделі *GPT*, а далі створюється модель генерації на основі архітектури *GPT*. Для кожного результату токенізації підслова ця модель кодує слова й позиції, створюючи цим вектор вбудовування кожного слова й позиції в результат. Далі вектор подається в архітектуру *GPT*, що містить блоки декодера. Після оброблення декодером вихідний вектор подається в багатошаровий персептрон для оцінювання елемента дорожньої карти цього результату.

Використання зазначеного засобу дозволило протестувати можливості генеративного ШІ для формування передпроєктних оцінок. Важливо зауважити, що відкритим питанням для обговорення залишається вибір мовної моделі (*BERT, Gemini, LLaMA*) та їх ефективність у розв'язанні поставленого завдання.

Висновки й перспективи подальшого розвитку

Запропоноване в статті дослідження присвячено концептуальним засадам створення програмних систем автоматизованого обчислення показників проєктів із розроблення програмного забезпечення та прогнозування впливу на показники різних підходів до проєкту. Особливу увагу приділено аналізу можливості використання моделей штучного інтелекту як найбільш трендового підходу в сучасній інженерії систем оброблення інформації та прийняття рішень.

Досягнуті результати підтвердили гіпотезу про те, що методи машинного та глибокого навчання доцільно впроваджувати для підвищення об'єктивності оцінок проєктів. Саме вони забезпечують можливість отримувати значущі та правильні характеристики заданих у процесі генерації результатів. Проблемним питанням водночас залишається побудова ефективних інструментів для оброблення складних даних та їх високорозмірних варіацій.

Ще одним напрямом подальших досліджень є створення універсального датасету, оскільки через відносно незначні розміри та специфічність наявних датасетів може значно збільшитися час на навчання моделі та виникнути суттєві похибки в прогнозах. Це особливо важливо для сфер діяльності, що мають надзвичайно мінливі контекстні умови, зокрема програмна інженерія, і датасети не є репрезентативними для всього діапазону продуктових та проєктних характеристик. Це призводить до необхідності постійної підтримки тренувального набору відомостей у актуальному стані, долучення нових даних, оновлення алгоритмів, визначення повторюваності та наявності інформації в датасеті тощо.

Цікавим також залишається питання моделювання контексту оцінювання, перспективним виглядає використання принципів метаконтекстного обміну інформацією та адаптації зазначеного підходу до оброблення неструктурованих даних. Цей підхід також може бути запроваджений як інструмент усунення різнорідності тренувальних наборів даних.

Необхідно наголосити, що методи машиного та глибокого навчання чуттєві до якості інформації, яка зберігається безпосередньо в датасетах. Унаслідок цього перспективними є дослідження, пов'язані з теорією якості даних.

Список літератури

1. Lauesen, S. IT project failures, causes and cures. *IEEE Access*. 2020. Vol. 8. P. 72059–72067. DOI: https://doi.org/10.1109/ACCESS.2020.2986545
2. SaasList. The State of Project Management in 2023 [42 Statistics]. 2023. URL: https://saaslist.com/blog/project-management-statistics/ (дата звернення: 15.04.2023).
3. Gupta R. G., Dumka A., Mazumdar B. D. Software Cost Estimation: A Comparative Analysis. *2024 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*. 2024. P. 1–8. DOI: https://doi.org/10.1109/ICCECE58645.2024.10497286
4. Nesma. What is Function Point Analysis (FPA) and what are function points? 2015. URL: https://nesma.org/faq/function-point-analysis-fpa-function-points/ (дата звернення: 17.04.2024).
5. Brar P., Nandal D. A Systematic Literature Review of Machine Learning Techniques for Software Effort Estimation Models. *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)*. 2022. P. 494–499. DOI: https://doi.org/10.1109/CCiCT56684.2022.00093
6. Milošević D. *Project Management ToolBox. Tools and Techniques for the Practicing Project Manager*. Wiley, Hoboken, New Jersey, 2003. ISBN: 9780471208228. 584 p.

7. Wolverton R. W. The Cost of Developing Large-Scale Software. *IEEE Transactions on Computers*. 1974. Vol. C-23. No. 6. P. 615–636. DOI: https://doi.org/10.1109/T-C.1974.224002
8. APMP. Competitive Price To Win. 2023. URL: https://www.apmp.org/assets/BoK-PTW-M-v4.pdf (дата звернення: 19.04.2024).
9. Affenzeller M., Wagner S., Winkler S., Beham A. *Genetic Algorithms and Genetic Programming. Modern Concepts and Practical Applications*. CRC Press, Boca Raton, Florida. 2009. ISBN: 9781420011326. 379 p.
10. Kim A., Lee D. Dynamic Bayesian network-based situational awareness and course of action decision-making support model. *Expert Systems with Applications*. 2024. Vol. 252, Part A. 124093 p. DOI: https://doi.org/10.1016/j.eswa.2024.124093
11. Chong L. W., Rengasamy D., Wong Y. W., Rajkumar R. K. Load prediction using support vector regression. *TENCON 2017 – 2017 IEEE Region 10 Conference*. 2017. P. 1069–1074. DOI: https://doi.org/10.1109/TENCON.2017.8228016
12. Elish, M. O. Improved estimation of software project effort using multiple additive regression trees. *Expert Systems with Applications*. 2009. Vol. 36, No. 7. P. 10774–10778. DOI: https://doi.org/10.1016/j.eswa.2009.02.013
13. Yunning Z., Xixi S. Research on Improved PERT Model in Analysis of Schedule Risk of Project. *2010 International Conference on E-Business and E-Government*. 2010. P. 2768-2771. DOI: https://doi.org/10.1109/ICEE.2010.699
14. Cunnama L. (nee Shillington), Sinanovic E., Ramma L., Foster N., Berrie L., Stevens W., Molapo S., Marokane P., McCarthy K., Churchyard G., Vassall A. Using Top-down and Bottom-up Costing Approaches in LMICs: The Case for Using Both to Assess the Incremental Costs of New Technologies at Scale. *Health economics*. 2016. Vol. 25. P. 53–66. DOI: https://doi.org/10.1002/hec.3295
15. Biletskiy, Y., Campeanu, C., Dudar, Z., Vorochek, O. Meta-context mediation to attain semantic interoperability. *2004 2nd International IEEE Conference on 'Intelligent Systems'*. 2004. Vol. 1. P. 238–243. DOI: https://doi.org/10.1109/IS.2004.1344674

References

1. Lauesen, S. (2020), "IT project failures, causes and cures", *IEEE Access*, Vol. 8, P. 72059–72067. DOI: https://doi.org/10.1109/ACCESS.2020.2986545
2. SaasList (2023), "The State of Project Management in 2023 [42 Statistics]", available at: https://saaslist.com/blog/project-management-statistics/ (last accessed 15.04.2023).
3. Gupta, R. G., Dumka, A., Mazumdar, B. D. (2024), "Software Cost Estimation: A Comparative Analysis", 2024 *International Conference on Computer, Electrical & Communication Engineering (ICCECE),* P. 1–8, DOI: https://doi.org/10.1109/ICCECE58645.2024.10497286.
4. Nesma (2023), "What is Function Point Analysis (FPA) and what are function points?" available at: https://nesma.org/faq/function-point-analysis-fpa-function-points/ (last accessed 17.04.2024).
5. Brar, P., Nandal, D. (2022), "A Systematic Literature Review of Machine Learning Techniques for Software Effort Estimation Models", *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, P. 494–499, DOI: https://doi.org/10.1109/CCiCT56684.2022.00093
6. Milošević, D. (2003), *Project Management ToolBox. Tools and Techniques for the Practicing Project Manager*, Wiley, Hoboken, New Jersey, 584 p. ISBN: 9780471208228
7. Wolverton, R. W. (1974), "The Cost of Developing Large-Scale Software", *IEEE Transactions on Computers*, Vol. C-23, No. 6, P. 615–636. DOI: https://doi.org/10.1109/T-C.1974.224002
8. APMP (2023), "Competitive Price To Win", available at: https://www.apmp.org/assets/BoK-PTW-M-v4.pdf (last accessed 19.04.2024).
9. Affenzeller, M., Wagner, S., Winkler, S., Beham, A. (2009), *Genetic Algorithms and Genetic Programming. Modern Concepts and Practical Applications*, CRC Press, Boca Raton, Florida, 379 p. ISBN: 9781420011326.
10. Kim, A., Lee, D. (2024), "Dynamic Bayesian network-based situational awareness and course of action decision-making support model", *Expert Systems with Applications*, Vol. 252, Part A. 124093 p., DOI: https://doi.org/10.1016/j.eswa.2024.124093.
11. Chong, L. W., Rengasamy, D., Wong, Y. W., Rajkumar, R. K. (2017), "Load prediction using support vector regression", *TENCON 2017 – 2017 IEEE Region 10 Conference*, 1069–1074 p. DOI: https://doi.org/10.1109/TENCON.2017.8228016
12. Elish, M. O. (2009), "Improved estimation of software project effort using multiple additive regression trees", *Expert Systems with Applications*, Vol. 36, No. 7, P. 10774–10778, DOI: https://doi.org/10.1016/j.eswa.2009.02.013
13. Yunning, Z., Xixi, S. (2010), "Research on Improved PERT Model in Analysis of Schedule Risk of Project", *2010 International Conference on E-Business and E-Government*, P. 2768–2771. DOI: https://doi.org/10.1109/ICEE.2010.699
14. Cunnama, L. (nee Shillington), Sinanovic, E., Ramma, L., Foster, N., Berrie, L., Stevens, W., Molapo, S., Marokane, P., McCarthy, K., Churchyard, G., Vassall, A. (2016), "Using Top-down and Bottom-up Costing Approaches in LMICs: The Case for Using Both to Assess the Incremental Costs of New Technologies at Scale", *Health economics*, Vol. 25, P. 53–66. DOI: https://doi.org/10.1002/hec.3295

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

15. Biletskiy, Y., Campeanu, C., Dudar, Z., Vorochek, O. (2004), "Meta-context mediation to attain semantic interoperability", *2004 2nd International IEEE Conference on Intelligent Systems*, Vol. 1, P. 238–243, DOI: https://doi.org/10.1101/IS.2004.1344674

Надійшла (Received) 09.05.2024

Відомості про авторів / About the Authors

Соловей Ілля Владиславович – Харківський національний університет радіоелектроніки, здобувач вищої освіти факультету комп'ютерних наук, Харків, Україна; e-mail: illia.solovei@nure.ua; ORCID ID: https://orcid.org/0009-0005-5715-2755

Ворочек Ольга Григорівна – кандидат технічних наук, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії, Харків, Україна; e-mail: olga.vorochek@nure.ua; ORCID ID: https://orcid.org/0000-0002-9054-9894

Solovei Illia – Kharkiv National University of Radio Electronics, Higher Education Seeker at the Faculty of Computer Science, Kharkiv, Ukraine.

Vorochek Olga – PhD (Engineering Sciences), Kharkiv National University of Radio Electronics, Associate Professor at the Department of Software Engineering, Kharkiv, Ukraine.

IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE METHODS TO THE PROCESSES OF AUTOMATED METRICS FORECASTING FOR SOFTWARE SYSTEMS DEVELOPMENT PROJECTS

The **subject matter** of the article is the process of automated forecasting of project metrics for software development projects that are typically subject to evaluation. It also covers AI methods and models that can be used to generate basic roadmap templates and operational work lists, as well as alternative estimates depending on the context. The **goal** of the work is to study the foundations of creating a system for automated predicting of alternative evaluations of a software product. The following **tasks** were solved in the article: determining the stages of evaluation related to the assessment of alternatives in the life cycle of a software development project; investigating the problems of predicting and the main factors affecting the final indicators; exploring predicting methods that can be used to implement multivariate assessment of a software development project. The following **methods** are used: methods for evaluating and predicting labor costs in software development projects, machine and deep learning, and assessing their effectiveness for solving the prediction problem. The following **results** were obtained: the conceptual foundations for creating automated evaluation and prediction systems based on the analysis of the effectiveness of selected machine learning models were determined, the areas of application for artificial intelligence methods in the process of evaluating software development project indicators were identified, the performance indicators of various machine learning models were assessed based on certain model evaluation parameters that characterize prediction accuracy; a conceptual architecture of a project roadmap generation software tool based on the GPT language model was proposed. **Conclusions**: the use of machine and deep learning methods can improve the accuracy of predictions for key project indicators, provide the possibility of flexible generation of various alternative roadmap templates and operational work lists, making the planning and management process more efficient and transparent under conditions of high uncertainty of project requirements.

Keywords: project evaluation; software; machine learning; generative models.

Бібліографічні описи / Bibliographic descriptions

Соловей І. В., Ворочек О. Г. Упровадження методів штучного інтелекту в процеси автоматизованого прогнозування показників проєктів із розроблення програмних систем. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 153–1165. DOI: https://doi.org/10.30837/2522-9818.2024.2.153

Solovei, I., Vorochek, O. (2024), "Implementation of artificial intelligence methods to the processes of automated metrics forecasting for software systems development projects", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 153–165. DOI: https://doi.org/10.30837/2522-9818.2024.2.153

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

УДК 004.8 DOI: https://doi.org/10.30837/2522-9818.2024.2.166

А. Ховрат

ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ВИКОРИСТАННЯ ГІБРИДНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ВИЯВЛЕННЯ СФАЛЬСИФІКОВАНОЇ АУДІОІНФОРМАЦІЇ В СОЦІАЛЬНО ОРІЄНТОВАНИХ СИСТЕМАХ

Предметом дослідження є проблема виявлення фальсифікованої інформації, зокрема в аудіоформаті, у соціально орієнтованих системах. **Мета роботи** – розроблення ефективної моделі для визначення факту підроблення звукових даних, яка основана на рекурентних і згорткових нейронних мережах із використанням технології *MapReduce* для паралелізації. У статті розв'язуються такі **завдання**: визначення особливостей аудіо в соціально орієнтованих системах; аналіз алгоритмів для передоброблення аудіоінформації як у перетвореному на текст вигляді, так і у вигляді сигналу; формування переліку цільових архітектур нейромереж та розкриття особливостей їх імплементації; експериментальна перевірка ефективності обраних підходів. Упроваджуються такі **методи**: аналітичний та індуктивний – для визначення цільового набору архітектур нейронних мереж; експертне оцінювання – для формування найбільш впливових факторів ефективності; експериментальний, багатокритеріального оцінювання та статистичні методи аугментації інформації – для визначення найбільш ефективної моделі. **Досягнуті результати**. Сформовано алгоритм передоброблення аудіоінформації для можливості застосування рекурентних і згорткових мереж. Імплементовано декілька підходів до класифікації інформації з використанням аугментації, основаної на векторній авторегресії та технології паралелізації *MapReduce*. Визначено, що найбільш ефективною моделлю, за сформованою задачею багатокритеріального вибору, є поєднання двоспрямованої рекуретної нейромережі з підтримкою короткочасної та довгострокової пам'яті із декількома згортковими мережами. Показано переваги використання технології *MapReduce* для оптимізації часу навчання й передоброблення інформації та визначено набір відкритих питань для подальшого дослідження й прикладного впровадження. **Висновки**. Застосування аналітичного та індуктивного підходу з подальшим проведенням експериментальної перевірки дало змогу розробити ефективний (з точністю понад 96%) механізм виявлення сфабрикованої інформації як у вигляді сигналу, так і у текстовій формі. Досягнутий результат дає підстави стверджувати про доцільність упровадження запропонованого підходу, що зменшує вплив подібної інформації в соціально орієнтованих системах, особливо під час кризових явищ.

Ключові слова: аугментація сигналів; векторна авторегресія; класифікація; оброблення природних мов; фейкова інформація.

Вступ

Упродовж останніх десятиліть технології, спроможні сфальсифікувати інформацію, досягли того рівня, коли про необхідність виявлення підробок у соціально орієнтованих системах говорять на законодавчому рівні [1]. Варто зазначити, що ступінь гостроти проблеми диверсифікується залежно від виду відповідних даних. Зокрема щодо відеоінформації спотворення ще не змогло досягти необхідного рівня [2]. Якщо йдеться про текст і фото, то вже існують ґрунтовні дослідження та навіть певні їх імплементації для визначення факту підробки [3, 4]. Водночас фальсифікація аудіо лише нещодавно змогла перейти межу простої ідентифікації, тобто з використанням людського слуху. Подібний стан речей дає змогу пересічним громадянам сплутати фейковий запис із реальним.

У звичайних умовах проблема може породити локальні конфлікти в групі людей, особливо це помітно в соціальних мережах [5]. Гострішою ситуація стає в умовах військово-політичної нестабільності, коли будь-яка інформація сприймається крізь призму інтенсифікованих емоцій, що сповільнюють процес критичного мислення. У разі, коли підробки є частиною новинного інформаційного поля, вони можуть прискорити соціальні зрушення, викликані кризою, і таким чином посилити її наслідки [6]. Це може мати економічний характер, суто соціальний або навіть військовий, і зрештою негативно позначитися на настроях населення. Як приклад подібної ситуації можна згадати фейкові аудіо, пов'язані з пандемією COVID-19 [7], чи величезну кількість сфальсифікованих записів на початку вторгнення Російської Федерації на територію України [8], що використовувалися

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

для приховання фактів порушення законів ведення війни чи дискредитації Збройних сил України. Необхідно зауважити, що, хоча можливості для якісного підроблення аудіоінформації з'явилися відносно нещодавно, їх вже значно простіше реалізовувати, на відміну від фото, яке потребує достатньо часу й значного обсягу вихідного матеріалу. Це зі свого боку лише посилює можливі ризики для суспільства й держави.

Аналіз останніх досліджень і публікацій

У дослідженні фейкових текстових новин декілька груп іспанських учених [9, 10] показали, що машинне навчання за своєю сутністю потребує достатньої кількості інформації для досягнення позитивного (точність понад 95%) результату класифікації. До того ж таке навчання є доволі чутливим до інформаційних викидів. Однак, зважаючи на наявні у відкритому доступі бази даних, зазначені недоліки не є значними, що було доведено в праці українських дослідників [11]. Подібне виправдання можна застосувати й для інформації інших видів, зокрема аудіо. Це продемонстрували науковці з Массачусетського технологічного інституту в праці [12].

Щодо інших способів виявлення сфальсифікованої інформації, то ще одним доволі популярним методом є створення графових моделей, яке було широко досліджене вченими з Гарварду [13] на прикладі підроблених акаунтів. Зазначений метод гарантує швидкий результат за мінімальних базових даних. Однак у застосуванні для аудіо- чи текстової інформації метод вимагатиме значного передоброблення, і таким чином виграш у швидкодії нівелюється. Якщо ж розглядати питання виявлення неправдивої інформації, то варто згадати й про проблему виявлення спаму. Групою китайсько-американських учених доведена можливість ефективного застосування марковських мереж [14]. Однак, беручи до уваги особливість галузі, їх застосування є доволі громіздким і вимагатиме значних обчислювальних потужностей, що довів канадський дослідник з Монреалю [15].

Під час розгляду візуальної інформації, зокрема фото чи відео, можна застосовувати авторегресійні моделі для виявлення відхилень від базисних значень, інакше кажучи, для виявлення факту маніпуляції з даними. Цей підхід був успішно застосований ученими зі Стенфорду. Водночас його також можна використати для дослідження аудіоінформації у вигляді сигналу. Однак варто зауважити, що спосіб є найбільш ефективним, коли в оригінальний запис було поміщено фрагмент підробки або ж замінена послідовність сигналу.

Іншою можливістю є застосування авторегресії для виявлення факту синтезу інформації (лише за наявності оригінальних записів цільової особи). В умовах контекстуального викривлення подібні моделі не дадуть бажаного результату. Саме тому надалі вони не розглядатимуться як класифікаційні. Натомість, як показала група китайських дослідників, авторегресію можна застосувати для аугментації [16]. Це дасть змогу уникнути проблеми нестачі інформації. Хоча тут варто зауважити, що в процесі розгляду довгострокового проміжку, особливо пов'язаного із соціальними катастрофами та іншими виплесками соціальної активності, зазначений підхід потребуватиме суттєвого обсягу інформації про зовнішні показники. Подібну проблему описують представники Корнелівського університету [17], наголошуючи, що під час розгляду завдання з обмеженою зовнішньою інформацією точність цього підходу (як для прогнозування, так і генерації) суттєво падає. Тому в межах цієї роботи розглядатимуться лише нетривалі (до двох місяців) хронологічні межі.

Визначення не розв'язаних раніше частин загальної проблеми. Мета роботи, завдання

Зазначений стан речей став підґрунтям для того, що у все більшої кількості країн обговорюють боротьбу зі сфабрикованою інформацією, хоча переважно й звертають увагу лише на її текстовий складник. Натомість питання виявлення видозміни аудіо є доволі відкритим, хоча й частково розглянутим, особливо коли йдеться про доповнення реальної інформації, а не синтетичну генерацію. З огляду на світову практику найбільш ефективними в цьому разі є нейронні мережі. Зважаючи на міжнародний досвід, вирішено зосередитися на згорткових та рекурентних нейронних мережах, однак розглядати не їх базові варіації, а більш просунуті та адаптовані для роботи зі значними обсягами текстової інформації – двоспрямовані рекурентні мережі з підтримкою довгострокової та короткочасної пам'яті (*BiLSTM*) і гібридні згортково-рекурентні мережі.

Однак необхідно зауважити, що вказані моделі потребують значних обсягів вихідної інформації,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

а також часу для навчання моделі. Для розв'язання зазначених проблем можна використовувати принципи аугментації аудіоматеріалів та паралелізму. Однак класичні форми розпаралелювання не гарантують значного виграшу у швидкості навчання моделі [18]. Як базову альтернативу зазначеним принципам використовують технологію *MapReduce*, що, окрім навчання, дає змогу пришвидшити й безпосередньо визначити факт фальсифікації. У межах цієї роботи розглянуто моделі класифікації фейкових аудіо на основі нейронних мереж і можливості модифікації цих алгоритмів і засобів їх розпаралелювання.

Мета роботи – розроблення ефективної моделі для визначення факту підроблення звукової інформації з використанням технології *MapReduce*. Для досягнення окресленої мети сформовано такий перелік *завдань*:

– визначити особливості аудіо в соціально орієнтованих системах (надалі для спрощення позначатимемо їх як новини);

– описати алгоритми, що дало б змогу передобробити аудіоінформацію як у перетвореному на текст вигляді, так і у вигляді сигналу;

– дослідити обрані архітектури нейронних мереж і визначити їх основні гіперпараметри;

– розкрити сутність імплементації технології *MapReduce*;

– сформувати план експерименту;

– проаналізувати результати дослідження та сформулювати відповідні висновки на основі розв'язання задачі багатокритеріального вибору.

Матеріали та методи

Сутність викривлення аудіоінформації

Почнемо розгляд з виявлення особливостей фейкової інформації. Насамперед необхідно зауважити, що фальсифікувати аудіо можна по-різному, зокрема такими способами:

– синтетичне створення аудіо за допомогою засобів штучного інтелекту: має місце за наявності значного обсягу вихідного матеріалу та необхідності отримати голос конкретної особи;

– компонування наявних звукових доріжок для викривлення сутності оригінальної інформації: на відміну від попереднього способу, тут можуть не застосовуватися генераційні алгоритми, однак необхідність отримання голосу конкретної особи зберігається;

– контекстуальне викривлення: у разі, коли власник голосу на аудіо не є важливим, на доріжках можуть записуватися неправдиві новини чи оголошення.

Незважаючи на природу фальсифікації, усі випадки зосереджуються на видозміні контексту з метою досягнення необхідного результату: погіршення настрою населення, шантаж тощо. З огляду на це можна сформувати такий перелік способів, як за наявною в повідомленні інформацією виявити та класифікувати неправдиві аудіо:

– використання неприродної кількості риторичних запитань, якщо йдеться про контекстуальне викривлення суспільно важливої інформації. У лінгвістичних дослідженнях зазначено, що в офіційно-діловому та публіцистичному стилях, призначених для ЗМІ, подібний тип мовленнєвих конструкцій майже відсутній [19]. Ця особливість властива як для текстових новин, так і аудіо, відео;

– відсутність заперечувальних конструкцій для зменшення когнітивного навантаження в поєднанні з песимістичним забарвленням обраних слів. Як приклад можна навести заміну слова "проблема" на "катастрофа". Однак необхідно зауважити, що в усному мовленні використання ненормативної лексики не дає змоги напевно визначити характер забарвлення, тому для подальшого аналізу оцінка подібних слів формуватиметься на основі контексту;

– вживання закликів і спонукань у недоречних формах. У разі контекстуального викривлення, мета якого – замінити реальні новини, подібні конструкції одразу вказують на неправдивість та некоректність інформації. Однак унаслідок розгляду мімікрійних записів ці особливості можуть визначити мету зловмисників;

– використання невиправданої кількості займенників. Цей фактор здебільшого необхідний для контекстуального викривлення, що наслідує публіцистичний стиль мовлення.

Зазначені характеристики не є вичерпними, однак маємо наголосити, що в разі оброблення трансформованих у текстовий вигляд аудіозаписів деякі особливості можуть не виявитися. Зокрема відомо, що для формування фейкових новин часто вживають короткі речення та слова, або ж у них наявна значна кількість різноманітних помилок [20]. Ці та подібні ним характеристики не беруться до уваги надалі, оскільки вони можуть виявитися через некоректність розпізнавання аудіо чи загалом бути особливою частиною процесу мовлення

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

людей, присутніх на аудіозаписі (наприклад, за умови змішування двох мов чи використання регіональних діалектизмів). Ці самі особливості можуть зумовити більшу кількість *false positive* випадків у виявленні фейків.

Аналіз аудіо як тексту

Унаслідок аналізу праці групи китайських учених визначено, що створення власного модуля *Speech to Text* супроводжується такими проблемами [16]:

– якість записів для тренування;

– нестача інформації для формування моделі (особливо гостро проблема постає для мов із невеликими корпусами);

– ігнорування дефектів вимови;

– коректність оброблення діалектизмів, неологізмів, скорочень тощо.

Зазначений перелік не є повним, тож аби уникнути вказаних проблем і досягти найліпшого результату перетворення аудіо на текст, вирішено використати *Google Speech to Text*, зокрема його відповідну обгортку для мови програмування *Python 3*. Додатково за допомогою голосу, записаного 20 людьми з різних регіонів України та різними вадами мовлення, а також 20 записами з українськомовних фільмів було встановлено:

– система має обмежені можливості в розпізнаванні скорочень;

– якщо паузи між словами є дуже тривалими, то модуль визначатиме окремі речення;

– якість записів не суттєво впливає на ефективність розпізнавання: наявність додаткового шуму нівелюється завдяки стадії передоброблення аудіо;

– без уваги до вищезазначеного точність розпізнавання сягала понад 95%, винятком стали сільські говори західних і східних областей.

Надалі зазначені твердження вважатимемо обмеженнями цієї статті.

Щоб перетворити добуту текстову інформацію в числове подання, скористаємося таким алгоритмом:

– розбиваємо текст на речення та окремі слова;

– вилучаємо слова без суттєвого лексико-графічного навантаження та ті, що не впливають на результат роботи алгоритмів (так звані стоп-слова). Наприклад: "однак", "це", "або", "тощо";

– формуємо на основі добутих лексем словник тексту;

– виокремлюємо основи кожного слова в словнику та прибираємо повтори (здійснюємо операцію стемінгу);

– визначаємо лему для кожної лексеми в словнику та знову прибираємо повтори (здійснюємо лематизацію);

– визначаємо частотну характеристку кожного слова (її опис буде здійснено нижче) та його емоційного забарвлення за допомогою засобів *Sentiment Analysis,* вбудованих у моделі *nltk* мови програмування *Python 3*;

– модифікуємо оцінку емоційного забарвлення у межах кожного окремого речення на основі правил, установлених раніше;

– агрегуємо й нормуємо частотно-емоційний показник для кожного речення, він слугуватиме цільовим індикатором для подальшого використання нейронних мереж;

– знаходимо показник підозрілості в нормованому вигляді на основі переліку слів, що часто вживаються у фейковій інформації.

Окрім указаних змінних, вхідними величинами також застосовуватимемо такі показники:

– частотно-емоційна характеристика 50 найбільш популярних новин за дату створення аудіозапису. Це дасть змогу взяти до уваги новинне зовнішнє середовище й, відповідно, скоригувати оцінку класифікатора;

– вага повідомлення. Вирішено створити набір даних, у якому аудіо поділятимуться на чотири групи: фрагменти домашнього діалогу, загальні новини, інформація з місця надзвичайних подій, новини особливої важливості. Маркування здійснюватиметься від 1 до 4 відповідно. Значення показника ваги також буде нормованим;

– ступінь надійності конвертації аудіо в текст. Визначатиметься в процесі порівняння реального тексту повідомлення з тим, що був оброблений *Google Speech to Text*, як відношення правильно розпізнаних слів.

Щодо частотної характеристики, то вирішено використати *BM25*, яка є певною модифікацією *TF-IDF*. Для кращого зрозуміння сутності модифікації детальніше розглянемо базову характеристику. Мета *TF-IDF* – зважати на важливість кожного слова в запиті та тексті з огляду на частоту вживання терміна як у певному документі, так і в корпусі загалом. Умовно слово "і" може бути одним із найбільш поширених у конкретному реченні,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

але воно часто вживається в корпусі загалом, тож матиме меншу значущість у пошуку за цим корпусом. Метод оснований лише на статистиці та рахується доволі швидко, тож і досі залишається популярним для задач, де не потрібні більш складні рішення. У разі *BM25* до *TF* додається насичена частота терміна. Тобто якщо термінологічна одиниця вже має високу частоту, то після певної познаки зростання частоти не матиме значного впливу на оцінку *TF. IDF* використовується так само. Додатково наявні два параметри – $k1$ і b, які можна налаштовувати під конкретний корпус. Параметр $k1$ відповідає за насичення частоти, а b – за міру впливу довжини документа на результати.

Вибір *TF-IDF* був зумовлений результатами попередніх досліджень, присвячених аналізу текстових новин [21]. Неточності та обмеження знизили ефективність запропонованих класифікаційних методів. Зокрема однією з проблем виокремлено перенасичення певними термінами, які не можна вважати стоп-словами, наприклад "катастрофа".

З'ясувавши особливості аудіо в текстовому вигляді, перейдемо до розгляду аудіо як сигналу.

Аналіз аудіо як сигналу

Першим етапом у підготовленні аудіо до його оброблення як сигналу є виокремлення вокалізованої частини від тиші. Подібна операція необхідна, адже в першому фрагменті присутні ключові елементи мовлення людей. Одним із загальновживаних способів маркування аудіосигналу є його розбиття на три стани:

– ділянка тиші (S), де відсутня вимова;

– невокалізована ділянка (U), де результуюча форма сигналу має аперіодичний або випадковий характер (має місце в разі, коли голосові зв'язки не вібрують);

– вокалізована ділянка (V), де результуюча форма сигналу є квазіперіодичною (має місце, коли голосові зв'язки суб'єкта мовлення напружені та, відповідно, вібрують).

Щодо поєднання двох перших ділянок, то зауважимо, що існують методи, які б дали змогу розмежувати тишу від невокалізації, однак вони вимагають постійного переналаштування для різного оточення, що в контексті аудіоновин є малоефективною процедурою. Подібна проблема наявна і для методів, які розмежовують вокалізовану ділянку від інших на основі малої енергії. Тому, зважаючи на зазначене твердження і той факт, що шумове оточення новин може різнитися, хоча є відносно стабільним, було вирішено застосувати методи, основані на розподілі. У цьому разі вважатимемо, що сигнал має гауссівську природу. Отже, щоб виокремити необхідну частину аудіо, можна використати функцію відстані Махаланобіса, яка є класифікатором лінійних шаблонів (*LPC*).

Щоб визначити параметри гауссівського розподілу, необхідно детермінувати базисне вікно. Для цього необхідне впровадження методу експертного оцінювання. Було опитано 30 фахівців з оброблення звуку з Харкова, Києва, Дніпра та Відня. Установлено, що оптимальний розмір вікна становить 200 мс. Тепер візьмемо до уваги формулу визначення гауссівського розподілу для одновимірного випадку:

$$g(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (1)$$

де μ – середнє значення;

σ – стандартне відхилення розподілу.

З огляду на це можна визначити такий набір правил:

$$\begin{aligned} P\left[|x-\mu|\le\sigma\right]&\approx 0,68,\\ P\left[|x-\mu|\le 2\sigma\right]&\approx 0,95,\\ P\left[|x-\mu|\le 3\sigma\right]&\approx 0,997. \end{aligned} \qquad (2)$$

Відстань Махаланобіса визначимо за допомогою формули

$$r=\frac{|x-\mu|}{\sigma}. \qquad (3)$$

Беручи до уваги (2) та (3), можна встановити, що з імовірністю 99,7% відстань становить менше ніж 3.

Процес передбачає такі кроки:

– алгоритм поступово аналізує аудіо у вікні 200 мс та визначає стандартне відхилення та середнє значення;

– для кожного наступного вікна обчислюється відстань Махаланобіса з використанням добутих раніше значень;

– якщо відстань перевищує 3, то вважаємо семпл вокалізованим, в іншому разі замінюємо його на порожні значення, фактично вилучаючи.

Нейронна мережа прийматиме на вхід перетворене аудіо, однак цього разу вікно для розбиття на семпли визначатиметься за допомогою крос-валідації. Як зазначалося вище, для уникнення проблем нестачі

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

аудіосигналів вирішено здійснити аугментацію за допомогою авторегресії.

$$\Phi_0 y_t = \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \Theta_0 u_t + \Theta_1 u_{t-1} + \cdots + \Theta_q u_{t-q} \ , \quad (4)$$

де y_t – K -вимірний часовий ряд;

Φ_i, Θ_j – матриці розмірності $K \times K$, $i = \overline{1, p}$, $j = \overline{1, q}$;

u_t – K -вимірний вектор білого шуму з нульовим середнім.

Тут варто зауважити, що, оскільки матриці коефіцієнтів не вироджені, їх легко нормалізувати

$$\Phi_0 \Delta y_t = \Pi y_{t-1} + \Psi_1 \Delta y_{t-1} + \cdots + \Psi_{p-1} y_{t-p+1} + \Theta_0 u_t + \Theta_1 u_{t-1} + \cdots + \Theta_q u_{t-q} \ , \quad (5)$$

де $\Pi = -\left(\Phi_0 - \Phi_1 - \cdots - \Phi_p\right)$;

$\Psi_i = -\left(\Phi_{i+1} + \cdots + \Phi_p\right)$, $i = \overline{1, p-1}$.

Добуті матриці коефіцієнтів за умови загальної кількості невідомих, що потрібно брати до уваги під час генерації, можуть змінюватися залежно від обраної моделі. У межах цього дослідження розглядатимуться такі варіації:

– проста авторегресія;
– сезонна авторегресія;
– авторегресія розподіленого лагу;
– авторегресія рухомого середнього;
– авторегресія інтегрованого рухомого середнього.

У цьому разі єдиним фактором ефективності можна вважати точність аугментації даних. Фактично необхідно гарантувати, що розподіл між вокалізованими та невокалізованими 200 мс семплами не зміниться.

Розглянувши основні засоби препроцесингу аудіо у вигляді тексту та сигналу, опишемо архітектури нейронних мереж, які плануємо використати.

Архітектури нейронних мереж

Обрані архітектури *BiLSTM* та *CNN-RNN* за своєю сутністю є нащадками класичних згорткових і рекурентних нейромереж. Тож щоб краще зрозуміти ці моделі, здійснемо поступовий огляд базових алгоритмів.

Почнемо з *RNN*. Вона містить декілька прихованих шарів, що працюють один за одним. Водночас кожен наступний шар на вхід отримує результат роботи попереднього. Зазначену особливість прийнято називати короткочасною пам'яттю за аналогією з людським мозком.

Усередині прихованого шару поступово обробляється вихідна інформація із використанням

Формально її можна подати таким чином:

в межах від 0 до 1. Модель (4) можна використовувати в процесі розгляду короткочасних періодів, до того ж потрібно гарантувати відсутність або ж несуттєвість зовнішнього впливу. Оскільки в межах цієї роботи вирішено розглянути середньострокову перспективу, необхідно знайти дельту між (4) та прогнозом на попередній період, тобто $\Phi_0 y_{t-1}$. Отже, маємо формулу:

градієнтів. Однак у разі, коли дані мають особливий характер та не є обмеженими за своєю сутністю (як згори, так і знизу), можуть виникати проблеми вибухового чи напівзниклого градієнтів. Це ситуації, коли значення градієнта починає прямувати до нескінченності та 0 відповідно. З огляду на міжнародний досвід під час аналізу текстової інформації (і сигналів також) ця проблема може мати місце. Щоб подолати вказаний недолік, вирішено використати нейронні мережі із підтримкою короткочасної та довгострокової пам'яті (*LSTM*).

Сутність математичного апарату в *LSTM* полягає в поступовому використанні декількох сигмоїд і гіперболічних тангенсів, що дають змогу коригувати значення таким чином, щоб уникнути спрямування як до 0, так і до нескінченності. Пропонуємо детальніше розглянути основні етапи роботи прихованих шарів цієї архітектури.

На першому етапі у *Forget Gate* додаються дві вхідні вагові інформації та здійснюється множення на сигма-функцію активації. Область значень цієї функції обмежує результат виконання вказаного етапу в межах від 0 до 1. Після завершення результат множиться на дані з каналу довгострокової пам'яті.

Другим етапом є *Input Gate*, що здійснює аналогічне множення на сигма-функцію. Однак цього разу результат коригується за допомогою застосування гіперболічного тангенса. Подібна операція дає змогу нівелювати проблему спрямування до нескінченності в процесі додавання до значення з каналу довгострокової пам'яті.

Унаслідок роботи двох зазначених етапів формується стан пам'яті, що разом із вхідною та попередньою вихідною інформацією слугує базисом для формування нового значення короткочасної пам'яті. Цей етап має назву *Output Gate* і є

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

завершальним кроком виконання одного прихованого шару. Сутність полягає в знаходженні гіперболічного тангенса від значення довгострокової пам'яті, який після цього множиться на сигма-функцію від попереднього значення в короткочасній пам'яті.

Хоча вказана архітектура дає змогу уникнути проблеми градієнтів, вона все ж має один істотний недолік під час оброблення природної мови – неможливість зважати на майбутній контекст. Для кращого розуміння наведемо приклад.

Нехай маємо початок речення "*Apple is something that…*". Звичайна *LSTM*-архітектура не зможе визначити, що саме мається на увазі під "*Apple*" – фрукт чи компанія, бо немає інформації про кінець наведеного речення. Для цієї архітектури варіанти "*Apple is something that competitors simply cannot reproduce*" та "*Apple is something that I like to eat*" є ідемпотентними. Для уникнення зазначеної проблеми вирішено використати двоспрямовану рекурентну нейромережу з підтримкою довгострокової та короткочасної пам'яті (*BiLSTM*). Фактично сутність цієї мережі полягає в поєднанні двох *LSTM*, спрямованих у різні напрямки. У цьому разі "допоміжна мережа" дає змогу зважати на контекст для початку речень.

Після відпрацювання двох підмереж результат обох рівнів поєднується, спочатку способом простої конкатенації, а після цього за допомогою лінійних трансформацій. Для того щоб визначити відповідні операції, вирішено провести крос-валідацію, під час якої встановлено, що найкращий результат досягається за умови використання усереднених значень. Аналогічний висновок було зроблено в процесі експертного оцінювання серед 10 осіб, що займаються обробленням природної мови.

Розглянувши першу із запропонованих архітектур, перейдемо до *CNN*-архітектури.

Порівняно з попередньою, ця модель не має ні короткочасної, ні довгострокової пам'яті. Натомість вона використовує шар згортки, що дає змогу суттєво зменшити розмірність вихідної інформації. Це особливо ефективно в розпізнаванні образів і визначенні факту фальсифікації зображень чи відео.

Щоб мати змогу побудувати якомога ефективнішу *CNN*-архітектуру, необхідно задати низку гіперпараметрів моделі. Одним із найбільш важливих серед них є розмір фільтра. Це елемент прихованого шару, що здійснює прохід поміж інформацією та виконує згортку. Після проведення крос-валідації встановлено, що для обраного випадку найкращим буде фільтр розмірністю $5\times5\times5$.

Тут варто зауважити, що останнє значення розмірності в нашому дослідженні відповідатиме кількості дескрипторів, що утворюють цільову змінну. Саме тому для аналізу аудіо як тексту розмірність становитиме $5\times5\times5$, однак у разі аналізу аудіо як сигналу розглядатимемо як дескриптори лише стандартне відхилення та середнє значення, тож розмірність фільтра становитиме $5\times5\times2$.

У процесі його проходження поміж даними розташований скалярний добуток між записами фільтра та вхідною інформацією. Це дозволить сформувати активаційну карту, розмірність якої дорівнюватиме кількості використаних фільтрів, інакше – глибині нейромережі. З огляду на кількість факторів, що беруться до уваги для класифікації, було вирішено зупинитися на глибині, рівній 5 та 2 відповідно.

Окрім зазначеного гіперпараметра для *CNN*-архітектури, розглядаються такі характеристики:

– розмір ядра (у процесі крос-валідації було використано ядро в межах від 2 до 5 і встановлено оптимальне значення, що дорівнює 4);

– розмір кроку під час розгляду. Зважаючи на рекомендації, що вказують на небажаність використання кроку понад 3 для тексту, було визначено оптимальне значення, що дорівнює 1;

– беручи до уваги встановлений крок, параметр додавання неістотних нулів не застосовуватиметься;

– з огляду на особливість предметної галузі вирішено не застосовувати параметр зміщення.

Варто зауважити, що кількість шарів згорткової мережі для аналізу текстової інформації має дорівнювати 1.

Проблемою зазначеної архітектури за умови її використання для оброблення природних мов є обмеженість щодо уваги до контексту. Звичайно, проходження фільтра дає змогу зважати на окіл кожного зі слів, однак особливість української мови полягає у великих реченнях. Отже, визначений контекст може розташовуватися поза фільтром *CNN*-моделі. Для уникнення цієї проблеми було вирішено поєднати *RNN* та *CNN*.

Хоча способів подібного поєднання існує декілька, у межах цієї роботи розглядатиметься лише *RCNN*-архітектура, що послідовно використовує дві нейронні мережі. Інакше кажучи, після здійснення згортки результат не лише конкатенується, а надсилається до шару з рекурентною нейронною мережею.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Щоб уникнути окреслених проблем під час розгляду тексту, вирішено застосувати не класичну *RNN*-архітектуру, а згадану вище двоспрямовану рекурентну нейромережу з підтримкою довгострокової та короткочасної пам'яті. Отже, архітектуру *RCNN* можна подати у вигляді, зображеному на рис. 1.

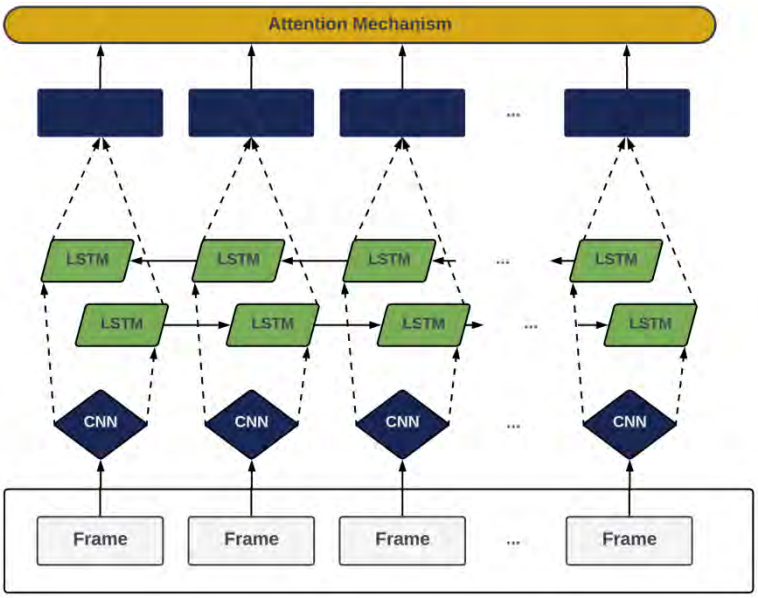


Рис. 1. Схематичне зображення *RCNN*-архітектури

Як зазначалося вище, у використанні складних нейронних мереж є суттєвий недолік – час їх навчання та оброблення. Крім цього, проблемою також є час передоброблення інформації. З метою зменшення впливу вказаних недоліків було вирішено впровадити технологію *MapReduce.*

Сутність технології MapReduce

Технологія *MapReduce* полягає в розподілі вихідного набору інформації таким чином, щоб вона оброблялася на окремих вузлах.

Ключовими операціями є застосування функцій мепінгу та редукції. Перша дає змогу розподілити інформацію між вузлами, на яких здійснюється бажане оброблення, а друга функція натомість збирає дані з усіх вузлів і уніфікує їх.

Варто зауважити, що технологія *MapReduce* визначає лише особливості реалізації відповідних модулів у межах певних фреймворків. Отже, ця реалізація може суттєво відрізнятися. Для цієї роботи обрано технологію *MapReduce* на основі *Hadoop*. Графічно запропоноване рішення можна подати так, як показано нижче (рис. 2).

У цьому разі особливу увагу варто приділити функціям розподілу та комбінування. Вони необхідні для того, щоб здійснити додаткову паралелізацію в кожному з вузлів, застосувавши різні регіони пам'яті. Щоб краще зрозуміти сутність цього підходу, можна вважати базові вузли процесами, а вказані регіони пам'яті – потоками.

Окрім цих двох функцій, важливою особливістю є сортування інформації перед редукцією. У статті розглядається інформація, що суттєво залежить від порядку й не має додаткових часових міток, таких як у часових рядах. Щоб уникнути проблеми внаслідок редукції, вирішено додати поле з порядковим номером кожного фрагменту тексту / сигналу. За ним і здійснюватиметься сортування.

MapReduce використовуватиметься незалежно в процесі препроцесингу вихідної інформації та навчання нейронних мереж.

Для здійснення передоброблення у разі сигналів важливим є лише здійснення редукції в правильному порядку. Для оброблення аудіо як тексту необхідно брати до уваги важливість формування якомога більшого словника. Для цього вирішено створити окрему нереляційну базу даних із підтримкою багатопотоковості, куди після базового оброблення (вилучення стоп-слів, лематизації, стемінгу) записуватиметься весь наявний словник. Отже, що більше матеріалу буде оброблено, то вищою буде точність формування відповідної частотної характеристики.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online)
Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

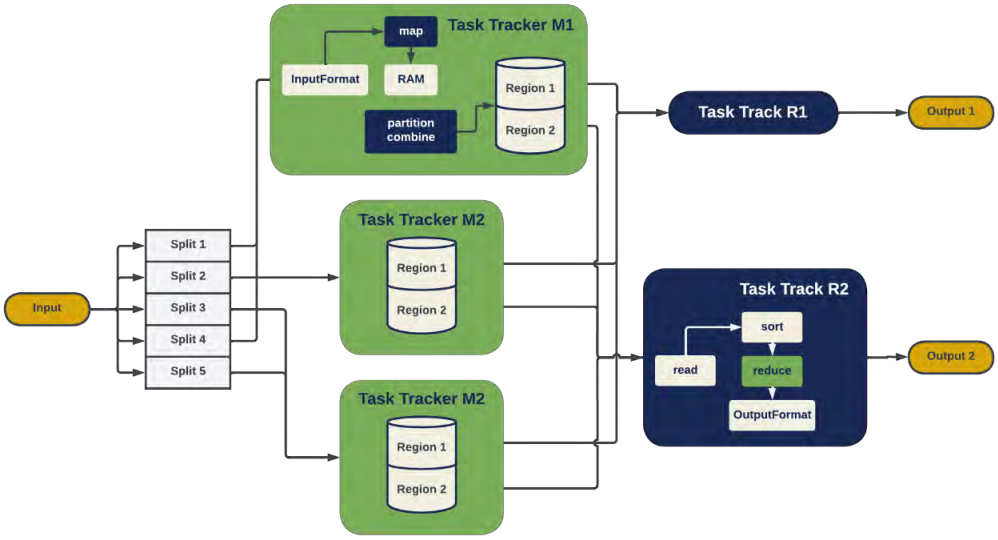


Рис. 2. Схематичне зображення *MapReduce* на основі *Hadoop*

Для *RCNN*-архітектури першим кроком є шар із *CNN*. У ньому ітеративно регулюються вагові коефіцієнти, обчислюючи їх часткові градієнти після того, як кожен набір навчальної інформації поширюється з допомогою мережі.

Так, розпаралелювання під час фази навчання може бути здійснено способом розподілу даних на декілька фрагментів. Потім кожен фрагмент передається в кілька *CNN*, і всі *CNN* навчаються незалежно. Після цього результати агрегуються за допомогою редуктора для отримання остаточної інформації, що потім використовується для оновлення вагових коефіцієнтів для подальшої ітерації.

Після завершення роботи шару *CNN* агрегована інформація передається у *BiLSTM*. Щоб пришвидшити двоспрямовану нейромережу, можна розподілити роботу двох нейронних мереж між двома вузлами. У цьому разі функція редукції фактично слугуватиме функцією агрегації результатів двох мереж.

Серед загальних переваг запропонованого підходу можна виокремити його масштабованість, відносну дешевизну, простоту у використанні та можливість моніторингу виконання за допомогою відповідних засобів *Hadoop* (якщо потрібен внутрішній моніторинг, використовуються базові методи мови програмування *Python*). Серед недоліків можна наголосити на необхідності створення великого обсягу програмного коду, прихованості оброблення (хоча і з можливістю перегляду лог-файлів) та потребі в тривалому налаштуванні конфігурації.

Експериментальне середовище

Зважаючи на особливість запропонованого дослідження, обрано метод контрольованого експерименту. Базове середовище виконання має такий набір характеристик:

- CPU: Intel Core i5-1135G7;
- RAM: 16 Гб;
- VRAM: 4 Гб;
- ОС: Ubuntu 21.04.

Перелічені характеристики майже в повному обсязі продубльовані на віртуальних вузлах, на яких планується здійснюватися часткове обчислення (було зменшено *RAM* з 16 до 8). Їх кількість становитиме від 3 до 4 (2 – у разі паралелізації двоспрямованих нейронних мереж).

Засобом обчислення часу виконання обрано бібліотеку *datetime* для *Python* 3, що має точність до наносекунди. Щоб базові обчислення не сповільнювали роботу програми, застосовано бібліотеки *numpy* та *polars*. Для оброблення природних мов (зважаючи на лематизацію, токенізацію та інші необхідні функції) обрано *python*-версію бібліотеки *nltk*. З метою реалізації нейронних мереж використано *tensorflow* із інструментарієм, що надає *pipeline* субмодуль.

У процесі розгляду аудіо у вигляді тексту, як уже зазначалося, було вирішено застосувати *Google Cloud Platform* для уникнення впливу оточення на швидкість роботи. Оптимізація цього інструменту дала змогу зменшити час затримки в разі отримання розшифровки текстової інформації з 10 с (на оброблення двохвилинного аудіо) до 2 с.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

Перше значення отримане за допомогою власноруч побудованого інструменту розпізнавання мови.

Інформацією для перевірки використано власноруч сформовані набори аудіо, згенеровані на основі текстових новин та частково видозмінених різними способами, описаними раніше.

Перший набір інформації стосується повномасштабного вторгнення Росії на територію України та містить як власне новини, так і реакції на певні події користувачів соціальних мереж, експертів із телебачення тощо. Другий набір інформації присвячений виборчому процесу 2019 р., що супроводжувався появою значної кількості неправдивих відомостей. Кожна з цих вибірок ділитиметься у співвідношенні 80 до 20 на навчальну та тестову підвибірки відповідно.

Щоб порівняти різні види нейронних мереж з та без використання *MapReduce,* необхідно визначитися з основними критеріями вибору. З огляду на те, що розглядається завдання класифікації для соціально гострих процесів, найбільш важливими критеріями є економія часу й точність класифікації.

Загалом було обрано такий перелік факторів:

– показник точності;

– економія часу навчання моделі за однакових потужностей;

– економія часу передоброблення інформації;

– економія мінімально допустимого обсягу інформації для досягнення *Accuracy* = 90%;

– можливість уваги до контексту.

Тут варто зауважити, що показник економії часу передоброблення інформації є важливим лише для визначення авторегресійного алгоритму та виграшу у швидкості, що надає технологія *MapReduce*. Тож для оцінювання ефективності нейронних мереж використовуватиметься лише чотири метрики з наведених вище.

Визначившись із критеріями, опишемо відповідні шкали оцінювання.

Економія часу навчання моделі вимірюватиметься в секундах за допомогою зазначеної вище бібліотеки. Сам показник у цьому разі не обмежуємо. Щоб зменшити вплив можливого вимірювання, викликаного проблемами з точністю роботи часових модулів чи оточення, вирішено проводити по п’ять замірів для показників часу та перевірити точність прогнозування на двох вибірках інформації.

Точність класифікації визначена за допомогою комбінації *F1-score* та *Precision*, нормалізована в межах від 0 до 1. Замір точності здійснюватимемо для двох вибірок і братимемо усереднене значення.

Як це було продемонстровано на прикладі *LSTM*- та *BiLSTM*-архітектур, на контекст можна зважати в різному обсязі:

– повною мірою в обох напрямках – 5 балів;

– лише в одному напрямку – 4 бали;

– лише в околі в обох напрямках – 3 бали;

– лише в околі в одному напрямку – 2 бали;

– не береться до уваги – 1 бал.

Варто зауважити, що в класифікації сигналів цей показник не використовуватиметься.

Щоб визначити, яка з моделей є найбільш ефективною за поданими вище критеріями, застосовано принцип лінійної адитивної згортки з ваговими коефіцієнтами. Для визначення вагових коефіцієнтів проведено експертне оцінювання серед журналістів та аналітиків (кількість рецензентів становила 50 осіб). Отже, перейдемо до визначення вагових коефіцієнтів. У питанні класифікації як сигналів, так і тексту найбільш важливим є показник точності. На другому місці – можливість зважати на контекст, на третьому – показник часу. Отже, можемо призначити:

– для точності – 16 очок;

– для можливості зважати на контекст – 10 очок;

– для економії часу навчання моделі за однакових потужностей – 2 очки;

– для економії мінімально допустимого обсягу інформації з метою досягнення потрібного рівня *Accuracy* – 2 очки.

З огляду на це отримуємо такі вагові коефіцієнти для кожного критерію:

– для точності: $16/30 = 8/15$ у разі аудіо як тексту; $16/20 = 0.8$, якщо аудіо у вигляді сигналу;

– для можливості зважати на контекст: $10/30 = 5/15$, якщо аудіо – текст;

– для економії часу навчання моделі за однакових потужностей: $2/30 = 1/15$ у разі аудіо як тексту, $2/20 = 0.1$, якщо аудіо у вигляді сигналу;

– для економії мінімального допустимого обсягу інформації: $2/30 = 1/15$ у разі аудіо у вигляді тексту, $2/20 = 0.1$, якщо аудіо як сигнал.

Наступним важливим елементом експериментального середовища є визначення можливих похибок. З огляду на описаний план можна виокремити такі фактори, що здатні вплинути на результат:

– під час перевірки економії часу – людський фактор та інструментальна похибка;

– під час перевірки точності – проблема даних.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Щоб пом'якшити зазначені невизначеності показники вимірюватимуться декілька разів.

Результати дослідження

Спочатку подамо значення можливості уваги до контексту для кожної з наведених вище моделей:

– *CNN* – 3 бали, контекст береться до уваги лише в околі в обох напрямках за допомогою функції згортки;

– *RNN* – 2 бали, контекст береться до уваги лише в околі в одному напрямку за допомогою короткочасної пам'яті;

– *LSTM* – 4 бали, контекст береться до уваги лише в одному напрямку;

– *BiLSTM* – 5 балів, контекст береться до уваги повною мірою в обох напрямках унаслідок двоспрямованості мережі;

– *RCNN* – 5 балів, контекст береться до уваги повною мірою в обох напрямках за допомогою функцій згортки та двоспрямованої мережі з довгостроковою пам'яттю.

Почнемо з показника економії часу передоброблення інформації в разі використання сигналів (з аугментацією за допомогою авторегресійних моделей).

Результати наведені в табл. 1. Усі значення економії пораховані відповідно до найповільнішого алгоритму – послідовної версії векторної авторегресії інтегрованого рухомого середнього.

Значення критеріїв для кожного методу налаштування моделі подано в табл. 1. Для спрощення викладок запропоновано такі познаки:

– *R* – класична векторна авторегресія;

– *RS* – сезонна векторна авторегресія;

– *RL* – векторна авторегресія розподіленого лагу;

– *RMA* – векторна авторегресія рухомого середнього;

– *RIMA* – векторна авторегресія інтегрованого рухомого середнього.

Таблиця 1. *Збереження часу передоброблення для сигналу (у мілісекундах)*

Послідовний підхід					MapReduce підхід				
R	**RS**	**RL**	**RMA**	**RIMA**	**R**	**RS**	**RL**	**RMA**	**RIMA**
55	47	36	12	0	169	138	105	49	4
58	45	30	13	0	173	135	101	52	5
61	48	36	20	0	165	130	99	47	3
57	42	35	22	0	168	132	102	49	4
59	48	38	18	0	166	132	101	50	5

Знайдемо середнє значення для кожного випадку за умови послідовних версій. Для *R* маємо 0.058 с; для *RS* – 0.046 с; для *RL* – 0.035 с; для *RMA* – 0.017 с; для *RIMA* – 0 с. Як бачимо, в середньому алгоритми рухомого середнього та інтегрованого рухомого середнього є значно повільнішими. Це пояснюється тим, що вони беруть до уваги екзогенні змінні в повному обсязі і врегульовують шуми. Оскільки для нашого дослідження точність аугментації не є найсуттєвішим показником, з огляду на результати вирішено скористатися класичною векторною авторегресією.

Якщо ж порівнювати з версіями *MapReduce*, то виграш у швидкості становить ~ 2.9 для кожної моделі. Якщо покращити конфігурацію для *MapReduce* і збільшити кількість вузлів до 4, то виграш становитиме ~ 3.74.

Аналіз аудіо як тексту показав різницю лише між паралелізованою версією та послідовною. Для трьох *MapReduce* вузлів прискорення становило ~ 3.1, у разі чотирьох ~ 4.3. Кількість вузлів менша за прискорення через додаткову оптимізацію запитів до *Google Speech to Text API*.

Перейдемо до результатів замірів економії часу навчання та почнемо з аналізу аудіо як сигналу (табл. 2).

Цього разу застосування паралелізації наявне лише для *BiLSTM*- і *RCNN*-архітектур для визначення загального рівня прискорення. Найбільш повільною є модель, побудована на *RCNN*-архітектурі.

Маємо такі середні значення показників: *CNN* – 51 с; *RNN* – 46 с; *LSTM* – 30 с; *BiLSTM* – 16 с; *RCNN* – 0 с. Як бачимо, що складніша за своєю сутністю архітектура, то повільнішим є навчання відповідної моделі.

Прискорення, досягнуте за допомогою *MapReduce* для *BiLSTM,* становило 2 (пояснюється простотою паралелізації та обмеженістю двома вузлами); для RCNN ~ 3.52 (для чотирьох вузлів результат збільшився до ~ 4.68).

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Таблиця 2. *Збереження часу тренування мереж для сигналу (у секундах)*

Послідовний підхід					MapReduce підхід	
CNN	**RNN**	**LSTM**	**BiLSTM**	**RCNN**	**BiLSTM**	**RCNN**
50	48	29	15	0	30	25
49	44	30	14	0	29	24
52	47	28	17	0	31	27
54	45	33	16	0	28	24
50	46	30	18	0	30	26

Унаслідок замірів для навчання обраних нейронних мереж із текстовою інформацією досягнуті такі результати середнього значення економії часу: *CNN* – 45 с; *RNN* – 41 с; *LSTM* – 24 с; *BiLSTM* – 12 с; *RCNN* – 0 с; *BiLSTM based MapReduce* – 24 с; *RCNN based MapReduct* – 22 с.

Як бачимо, в середньому економія часу менша, що пояснюється особливістю оброблення природної мови. Цього разу прискорення, досягнуте за допомогою *MapReduce* для *BiLSTM*, становило 2 (ситуація аналогічна попередній); для *RCNN* ~ 3.2 (для чотирьох вузлів результат збільшився до ~ 4.41).

Перейдемо до результатів точності класифікації для аналізу аудіо як сигналу (табл. 3).

Таблиця 3. *Точність класифікації для сигналу*

Інформація	**CNN**	**RNN**	**LSTM**	**BiLSTM**	**RCNN**
Вибори	0.89	0.91	0.93	0.95	0.97
Війна	0.91	0.89	0.92	0.97	0.97

Треба зауважити, що використання *MapReduce* не вплинуло на точність класифікації для обох наборів інформації, тому відповідні викладки було відкинуто.

З огляду на досягнутий результат *RCNN* гарантує найвищу точність класифікації (хоча різниця з *BiLSTM* не є суттєвою). Розгляд аудіо як тексту показав, що ситуація майже не змінюється (табл. 4).

Таблиця 4. *Точність класифікації для тексту*

Інформація	**CNN**	**RNN**	**LSTM**	**BiLSTM**	**RCNN**
Вибори	0.92	0.91	0.91	0.96	0.96
Війна	0.90	0.91	0.94	0.95	0.96

Останньою метрикою є розмір навчальної вибірки, необхідний для досягнення точності щонайменше 90%. Для цього проведено кілька ітерацій з поступовим збільшенням кількості записів від 5000 до 10000 (у разі аудіосигналів переважна більшість була результатом аугментації). Виявлено, що задана точність досягається за умов: 7000 записів для *CNN*; 7500 записів для *RNN*; 6600 записів для *LSTM*; 6100 записів для *BiLSTM*; 5800 записів для *RCNN*. Отже, економія мінімального допустимого обсягу інформації становить: 1700 для *RCNN*; 1400 для *BiLSTM*; 900 для *LSTM*; 500 для *CNN*; 0 для *RNN*.

Тепер можемо систематизувати добуті значення метрик та визначимо альтернативи, оптимальні за Парето, для оброблення аудіо як сигналу (табл. 5). Усі ненормалізовані значення були нормовані та округлені до сотих.

Таблиця 5. *Значення критеріїв, оптимальних за Парето, унаслідок аналізу аудіо*

Модель	**Збереження часу**	**Точність**	**Збереження обсягу інформації**
CNN	1.00	0.90	0.29
LSTM	0.59	0.93	0.53
BiLSTM	0.31	0.96	0.82
RCNN	0.00	0.97	1.00

На основі результатів можна обчислити значення лінійної адитивної згортки з ваговими коефіцієнтами. Для *CNN* маємо 0.849, для *LSTM* – 0.856, для *BiLSTM* – 0.881, а для *RCNN* – 0.876. Можна зауважити, що найбільш ефективною моделлю у виявленні факту фальсифікації для аудіо як сигналу є двоспрямована рекурентна нейромережа з підтримкою довгострокової та короткочасної пам'яті.

Однак різниця між *BiLSTM* та *RCNN* мало відчутна й може вважатися похибкою. Крім цього, суттєвий виграш у швидкості для *BiLSTM* частково нейтралізується за допомогою застосування технології *MapReduce*.

Перейдемо до систематизації результатів, досягнутих унаслідок класифікації аудіо як текстової інформації. Відповідні нормалізовані значення, оптимальні за Парето, наведені нижче (табл. 6).

На основі результатів можна обчислити значення лінійної адитивної згортки з ваговими коефіцієнтами. Для *CNN* маємо 0.771, для *LSTM* – 0.833, для *BiLSTM* – 0.918, а для *RCNN* – 0.917. Як і в попередньому разі, найефективнішою моделлю є *BiLSTM*, однак виграш у швидкодії зменшується з допомогою паралелізації.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Таблиця 6. *Значення критеріїв, оптимальних за Парето, унаслідок аналізу тексту*

Модель	Збереження часу	Точність	Збереження обсягу інформації	Контекст
CNN	1.00	0.91	0.29	0.60
LSTM	0.53	0.93	0.53	0.80
BiLSTM	0.27	0.96	0.82	1.00
RCNN	0.00	0.96	1.00	1.00

Беручи до уваги наведене вище, зазначимо, що найбільш ефективними моделями є *BiLSTM* та *RCNN*, а результати застосування технології *MapReduce* доводять доцільність її використання для класифікації фейкових аудіозаписів різного виду.

Висновки

Метою статті було розроблення ефективної моделі для визначення факту підробки звукової інформації з використанням технології *MapReduce*. Для цього проаналізовано особливості фальсифікації аудіоінформації у вигляді сигналу й тексту. Крім цього, досліджені сучасні наукові публікації, присвячені обраній темі, і низка експертних опитувань дали змогу сформувати набір алгоритмів для створення власної моделі визначення фейкових аудіо. Перша стадія цієї моделі передбачає:

– якщо аудіо – це текст: передоброблення інформації за допомогою *Google Speech to Text* та подальшу конвертацію тексту в числове подання, зважаючи на частотно-емоційній характеристики (знайдені за допомогою алгоритму BM-25) самого повідомлення та останніх перевірених новин, ступені надійності конвертації, вагу повідомлення;

– якщо аудіо – це сигнал: очищення сигналу від шуму та невокалізованих ділянок із подальшою аугментацією за допомогою векторної авторегресії, паралелізованої за допомогою *MapReduce* (результати експерименту дали змогу обгрунтувати вибір класичної векторної авторегресії).

Наступною стадією є застосування нейронної мережі. З огляду на проаналізовані дослідження обрано рекурентні та згорткові нейронні мережі, зокрема:

– класична згорткова нейромережа;

– класична рекурентна нейромережа;

– рекурентна нейромережа з довгостроковою пам’яттю;

– двоспрямована рекурентна нейромережа з довгостроковою пам’яттю;

– гібридна нейромережа, що поєднує декілька згорткових мереж із двоспрямованою рекурентною мережею з довгостроковою пам’яттю.

Щоб подолати проблему, пов’язну з часом навчання моделі, використано технологію *MapReduce*. Для визначення найбільш ефективної нейронної мережі та доцільності застосування запропонованого способу паралелізації сформовано набір критеріїв, що дав змогу використати принцип лінійної адитивної згортки з ваговими коефіцієнтами. На основі цих критеріїв, зазначених модифікацій та імплементації обраних моделей за допомогою бібліотек *Python 3* проведено серії експериментів з інформацією щодо виборчого процесу в Україні 2019 р. та повномасштабного вторгнення Російської Федерації на територію нашої країни.

У процесі експериментів виявлено, що двоспрямована рекурентна нейромережа з довгостроковою пам’яттю є найбільш ефективною, хоча вона й поступається у швидкості менш складним моделям. Водночас різниця в ефективності між нею та гібридною нейромережею не суттєва. З’ясовано, що виграш у економії часу передоброблення внаслідок застосування технології *MapReduce* може становити 4.3 – для тексту і 4 – для сигналу. Перевага, пов’язана з економією часу навчання нейромереж може досягати 4.71 – для тексту і 4.68 – для сигналу, нівелюючи розрив у ефективності між *BiLSTM* та *RCNN*.

Отже, використання побудованої моделі на основі *BiLSTM* (або *RCNN*) є високоефективним для визначення факту підробки аудіо як у вигляді тексту, так і сигналу. Упровадження передоброблення інформації та навчання нейромереж за допомогою *MapReduce* є доцільним. Відкритими залишаються проблеми розширення результатів на зображення та відеоматеріали, а також можливості застосування інших підходів для класифікації та аугментації інформації.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

Список літератури

1. Anders M. Fake News Detection. European Data Protection Supervisor. URL: https://edps.europa.eu/press-publications/publications/techsonar/fake-news-detection_en (дата звернення: 27.05.2024).
2. Real-Time Advanced Computational Intelligence for Deep Fake Video Detection / N. Bansal та ін. *Applied Science*. 2023. Vol. 13 (5). 3095 p. DOI: 10.3390/app13053095
3. A Signal Detection Approach to Understanding the Identification of Fake News / C. Batailler та ін. *Perspectives on Psychological Science*. 2023. Vol. 17 (1). P. 78–98. DOI: 10.1177/1745691620986135
4. Reis J. Supervised Learning for Fake News Detection / J. C. S. Reis та ін. *IEEE Explore*. 2019. Vol. 34 (2). P. 76–81. DOI: 10.1109/MIS.2019.2899143
5. Giandomenico D. D. Fake news, social media and marketing: A systematic review / D. D. Giandomenico та ін. *Journal of Business Research*. 2021. Vol. 124. P. 329–341. DOI: 10.1016/j.jbusres.2020.11.037
6. Yuan L. Sustainable Development of Information Dissemination: A Review of Current Fake News Detection Research and Practice / L. Yuan та ін. *Systems*. 2023. Vol. 11 (9). 458 p. DOI: 10.3390/systems11090458
7. The impact of fake news on social media and its influence on health during the COVID-19 pandemic / Y. M. Rocha та ін. *Journal of Public Health.* 2023. Vol. 31. P. 1007–1016. DOI: 10.1007/s10389-021-01658-z
8. Alonso M. Dataset for multimodal fake news detection and verification tasks / A. Bondielli та ін. *Data in Brief*. 2024. Vol. 54. 110440 p. DOI: 10.1016/j.dib.2024.110440.
9. Tolosana R. Sentiment Analysis for Fake News Detection / Tolosana R. та ін. *Electronics.* 2021. Vol. 10 (11). 1348 p. DOI: 10.3390/electronics10111348
10. Afanasieva I. Deepfakes and beyond: A Survey of face manipulation and fake detection / Afanasieva I. та ін. *Information Fusion*. 2020. Vol. 64. P. 131–148. DOI: 10.1016/j.inffus.2020.06.014
11. Afanasieva Nataliia Application of Neural Networks to Identify of Fake News / N. Afanasieva та ін. *Computational Linguistics and Intelligent Systems*, Kharkiv, 20–21 квітня 2023 р. 2023. 3396 p. URL: https://ceur-ws.org/Vol-3396/paper28.pdf (дата звернення: 27.05.2024)
12. Bhatia T. Using transfer learning, spectrogram audio classification, and MIT app inventor to facilitate machine learning understanding. *Massachusetts Institute of Technology*. URL: https://dspace.mit.edu/handle/1721.1/127379 (дата звернення: 27.05.2024).
13. Breuer A., Eilat R., Weinsberg U. Friend or Faux: Graph-Based Early Detection of Fake Accounts on Social Networks. *Web Conference*, Taipei, 20–24 квіт. 2023. P. 1287–1297. DOI: 10.1145/3366423.3380204
14. Xia T., Chen X. A. Discrete Hidden Markov Model for SMS Spam Detection. *Applied Science*. 2020. Vol. 10 (14). 5011 p. DOI: 10.3390/app10145011
15. Najar F., Zamzami N., Bouguila S. Fake News Detection Using Bayesian Inference. *Information Reuse and Integration for Data Science*, Los Angeles, 30 черв. – 1 серп. 2019. P. 389–394. DOI: 10.1109/IRI.2019.00066
16. Montserrat D. Generative Autoregressive Ensembles for Satellite Imagery Manipulation Detection / D. M. Montserrat та ін. *Workshop on Information Forensics and Security*, New York, 6–11 груд. 2020. P. 1–6. DOI: 10.1109/WIFS49906.2020.9360909
17. Ning C., You F. Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers & Chemical Engineering*. 2019. Vol. 125. P. 434–448. DOI: 10.1016/j.compchemeng.2019.03.034
18. Sardar T. H., Ansari Z. An Analysis of Distributed Document Clustering Using MapReduce Based K-Means Algorithm. *Journal of The Institution of Engineers (India):Series B*. 2020. Vol. 101. P. 641–650. DOI: 10.1007/s40031-020-00485-2
19. Deng R., Duzhin, F. Topological Data Analysis Helps to Improve Accuracy of Deep Learning Models for Fake News Detection Trained on Very Small Training Sets. *Big Data and Cognitive Computing*. 2022. Vol. 6 (3). 74 p. DOI: 10.3390/bdcc6030074
20. Choudhary A., Arora A. Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications*. 2021. Vol. 169. 114171 p. DOI: 10.1016/j.eswa.2020.114171
21. Khovrat A. Parallelization of the VAR Algorithm Family to Increase the Efficiency of Forecasting Market Indicators During Social Disaster / A. Khovrat та ін. *Information Technology and Implementation*, м. Kyiv, 30 лист. – 2 груд. 2022. P. 222–233. URL: https://ceur-ws.org/Vol-3347/Paper_19.pdf (дата звернення: 27.05.2024).

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

References

1. Anders M. "Fake News Detection. European Data Protection Supervisor", available at: https://edps.europa.eu/press-publications/publications/techsonar/fake-news-detection_en (last accessed 27.05.2024).
2. Bansal, N., Aljrees, T., Yadav, D. P., Singh, K. U., Kumar, A., Verma, G. K., Singh, T. (2023), "Real-Time Advanced Computational Intelligence for Deep Fake Video Detection", *Applied Science*, No. 13(5), 3095 p. DOI: 10.3390/app13053095
3. Batailler, C., Brannon, S. M., Teas, P. E., Gawronski, B. (2023), "A Signal Detection Approach to Understanding the Identification of Fake News", *Perspectives on Psychological Science*, No. 17(1), P. 78–98. DOI: 10.1177/1745691620986135
4. Reis, J. C. S., Correia, A., Murai, F., Veloso, A., Benevenuto, F. (2019), "Supervised Learning for Fake News Detection", *IEEE Intelligent Systems*, No. 34(2), P. 76–81. DOI: 10.1109/MIS.2019.2899143
5. Giandomenico, D. D., Sit, J., Ishizaka, A., Nunan, D. (2021), "Fake news, social media and marketing: A systematic review", *Journal of Business Research*, Vol. 124, P. 329–341. DOI: 10.1016/j.jbusres.2020.11.037
6. Yuan, L., Jiang, H., Shen, H., Shi, L., Cheng, N. (2023), "Sustainable Development of Information Dissemination: A Review of Current Fake News Detection Research and Practice", *Systems*, No. 11(9), 458 p. DOI: 10.3390/systems11090458
7. Rocha, Y. M., de Moura, G. A., Desiderio, G. A., de Oliveira, C. H., Lourenço, F. D., de Figueiredo Nicolete, L. D. (2023), "The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review", *Journal of Public Health*, Vol. 31, P. 1007–1016. DOI: 10.1007/s10389-021-01658-z
8. Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., Vilares, J. (2021), "Sentiment Analysis for Fake News Detection", *Electronics*, No. 10(11), 1348 p. DOI: 10.3390/electronics10111348
9. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., Ortega-Garcia, J. (2020), "Deepfakes and beyond: A Survey of face manipulation and fake detection", *Information Fusion*, Vol. 64, P. 131–148. DOI: 10.1016/j.inffus.2020.06.014
10. Afanasieva, I., Golian, N., Golian, V., Khovrat, A., Onyshchenko, K. (2023), "Application of Neural Networks to Identify of Fake News". *Computational Linguistics and Intelligent Systems (COLINS 2023): 7th International Conference, Kharkiv, 20 April – 21 April 2023: CEUR workshop proceedings*, No. 3396, P. 346–358, available at: https://ceur-ws.org/Vol-3396/paper28.pdf (last accessed: 27.05.2023).
11. Afanasieva Nataliia Application of Neural Networks to Identify of Fake News (2023), / N. Afanasieva et al. *Computational Linguistics and Intelligent Systems*, Kharkiv, 20–21.04.2023. 3396 p. available at: https://ceur-ws.org/Vol-3396/paper28.pdf (last accessed: 27.05.2024)
12. Bhatia, N. (2020), "Using transfer learning, spectrogram audio classification, and MIT app inventor to facilitate machine learning understanding", *Massachusetts Institute of Technology*, available at: https://dspace.mit.edu/handle/1721.1/127379 (last accessed 27.05.2024)
13. Breuer, A., Eilat, R., Weinsberg, U. (2023), "Friend or Faux: Graph-Based Early Detection of Fake Accounts on Social Networks", *Web Conference*, *20–24 April 2023*, *Taipei*, P. 1287–1297. DOI: 10.1145/3366423.3380204
14. Xia, T., Chen, X. A. (2020), "Discrete Hidden Markov Model for SMS Spam Detection", *Applied Science*, Vol. 10 (14), 5011 p. DOI: 10.3390/app10145011
15. Najar, F., Zamzami, N., Bouguila, S. (2019), "Fake News Detection Using Bayesian Inference", *Information Reuse and Integration for Data Science*, *30 July – 1 August 2019*, *Los Angeles*, P. 389–394. DOI: 10.1109/IRI.2019.00066
16. Montserrat, D. M., Horváth, J., Yarlagadda, S. K., Zhu, F., Delp, E. J. (2020), "Generative Autoregressive Ensembles for Satellite Imagery Manipulation Detection". *Workshop on Information Forensics and Security (WIFS 2020): 12th IEEE International Workshop, New York, 6 December – 11 December 2020: IEEE,* P. 1–6. DOI: 10.1109/WIFS49906.2020.9360909
17. Ning, C., You, F. (2019), "Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming", *Computers & Chemical Engineering*, Vol. 125, P. 434–448. DOI: 10.1016/j.compchemeng.2019.03.034
18. Sardar, T. H., Ansari, Z. (2020), "An Analysis of Distributed Document Clustering Using MapReduce Based K-Means Algorithm", *Journal of The Institution of Engineers (India): Series B*, Vol. 101, P. 641–650. DOI: 10.1007/s40031-020-00485-2
19. Deng, R., Duzhin, F. (2022), "Topological Data Analysis Helps to Improve Accuracy of Deep Learning Models for Fake News Detection Trained on Very Small Training Sets", *Big Data and Cognitive Computing*, Vol. 6 (3), 74 p. DOI: 10.3390/bdcc6030074
20. Choudhary, A., Arora, A. (2021), "Linguistic feature based learning model for fake news detection and classification", *Expert Systems with Applications*, Vol. 169, Article 114171. DOI: 10.1016/j.eswa.2020.114171

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

21. Khovrat, A., Kobziev, V., Nazarov, A., Yakovlev, S. (2022), "Parallelization of the VAR Algorithm Family to Increase the Efficiency of Forecasting Market Indicators During Social Disaster". *Information Technology and Implementation (IT&I 2022): 9th Internaional Conference, Kyiv, 30 November – 2 December 2022: CEUR Workshop Proceedings.* No. 3347, P. 222–233, available at: https://ceur-ws.org/Vol-3347/Paper_19.pdf (last accessed: 27.05.2024).

Надійшла (Received) 30.05.2024

Відомості про авторів / About the Authors

Ховрат Артем Вячеславович – Харківський національний університет радіоелектроніки, аспірант, Харків, Україна; e-mail: artem.khovrat@nure.ua; ORCID ID: https://orcid.org/0000-0002-1753-8929

Khovrat Artem – Kharkiv National University of Radio Electronics, Postgraduate Student at the Department of Software Engineering, Kharkiv, Ukraine.

THE EFFICIENCY ASSESSMENT OF USING HYBRID NEURAL NETWORKS FOR THE DETECTION OF FORGED AUDIO DATA IN SOCIALLY ORIENTED SYSTEMS

The **subject** of the research is the problem of detecting falsified data, in particular in audio format, in socially oriented systems. The **goal** of the work is to develop an effective model based on recurrent and convolutional neural networks for determining the fact of forgery of sound data, using MapReduce technology for parallelization. The article addresses the following **tasks**: determining the features of audio in socially-oriented systems, conducting an analysis of algorithms for processing audio information both in the form of text and in the form of a signal, forming a list of target architectures of neural networks and revealing the features of their implementation, conducting an experimental test of effectiveness selected approaches. The following **methods** used are – analytical and inductive method for determining the target set of neural network architectures; expert assessment for the formation of the most influential efficiency factors; experimental, multi-criteria evaluation and statistical methods of data augmentation to determine the most effective model. The following **results** were obtained: an audio data reprocessing algorithm was developed for the possibility of using recurrent and convolutional networks. Several approaches to data classification using augmentation based on vector autoregression and MapReduce parallelization technology have been implemented. It was determined that the most effective model for the multi-criteria selection problem is a combination of a bidirectional recurrent neural network with support for short- and long-term memory with several convolutional networks. The advantages of using MapReduce technology to optimize training time and data processing are shown, and a set of open questions for further research and applied implementation is defined. **Conclusions**: he application of an analytical and inductive approach followed by experimental verification made it possible to develop an effective (with an accuracy of more than 96%) a mechanism for detecting fabricated data both in the form of a signal and in text form. The obtained result makes it possible to assert the feasibility of implementing the proposed approach, and, accordingly, makes it possible to reduce the influence of such information in socially oriented systems, especially during crisis events.

Keywords: signal augmentation; vector autoregression; classification; natural language processing; fake information.

Бібліографічні описи / Bibliographic descriptions

Ховрат А. В. Оцінювання ефективності використання гібридних нейронних мереж для виявлення сфальсифікованої аудіоінформації в соціально орієнтованих системах. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 166–181. DOI: https://doi.org/10.30837/2522-9818.2024.2.166

Khovrat, A. (2024), "The efficiency assessment of using hybrid neural networks for the detection of forged audio data in socially oriented systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 166–181. DOI: https://doi.org/10.30837/2522-9818.2024.2.166

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

УДК 004.056.5 DOI: https://doi.org/10.30837/2522-9818.2024.2.182

К. Шуліка, Д. Балагура, А. Смірнов, Д. Непокритов, А. Литвин

МЕТОД ВИКОРИСТАННЯ СУЧАСНИХ СИСТЕМ ЗАХИСТУ КІНЦЕВИХ ТОЧОК (EDR) ДЛЯ УБЕЗПЕЧЕННЯ ВІД КОМПЛЕКСНИХ АТАК

Предметом дослідження в статті є архітектура систем захисту кінцевих точок (EDR) та агентів EDR як їх базового складника з погляду механізмів виявлення комплексних атак на інформаційно-комунікаційні системи (ІКС) та протидії загрозам. **Мета роботи** – розроблення методу підвищення ефективності використання систем захисту кінцевих точок для зниження ризиків компрометації ІКС інформаційних, промислових та інфраструктурних об’єктів щодо ефективного перерозподілу та використання механізмів EDR, команди з кібербезпеки та інших ресурсів для здійснення заходів з організації безпеки на підприємстві, в установі чи організації. У статті розв’язуються такі **завдання**: огляд та аналіз систем EDR; дослідження архітектури EDR-рішень та агентів EDR, особливостей їх використання, логіки побудови методів і механізмів виявлення загроз для системи з боку зловмисників та зловмисного коду; надання рекомендацій щодо організації ІКС для її захисту загалом та окремих елементів, а також з огляду на наявні сили (команда із кіберзахисту, її кваліфікація та рівень обізнаності в архітектурі EDR-рішень) та засоби (елементи EDR-систем) для організації захисту. Упроваджуються такі **методи**: моделювання механізмів атак, моделювання поведінки зловмисника. **Досягнуті результати**: сформульовано загальні та конкретні рекомендації щодо оптимізації роботи EDR-систем та забезпечення ефективного використання елементів EDR-систем у інформаційно-комунікаційних мережах підприємств чи організацій різного типу та спрямованості залежно від ресурсів і наявної інформації з погляду необхідності її захисту. **Висновки**: запропоновані рекомендації щодо застосування EDR-механізмів для захисту інформаційних систем і мереж дають змогу оптимізувати витрати на створення інфраструктури захисту та здійснення відповідних заходів з огляду на особливості наявного інструментарію, навченості та обізнаності команди з кібербезпеки як щодо часу реакцій на загрози, так і з погляду складності та вартості виконання завдань із захисту.

Ключові слова: інформаційно-комунікаційні системи (ІКС); EDR-система; операційний центр безпеки SOC; EDR-агент; аналіз загроз; політика EDR; виявлення атак.

Вступ

Комп’ютерна мережа будь-якого промислового підприємства завжди є під загрозою проникнення з боку зловмисників для отримання чи знищення конфіденційної інформації, руйнування мережі чи інших зловмисних дій, зокрема вимагання коштів за збереження даних недоторканими. Наприклад, відповідно до *Statista* [1], незважаючи на певне зниження темпів зростання за останні два роки, кількість організацій, що постраждала від атак програм-вимагачів, продовжує впевнено зростати впродовж останніх шести років (рис. 1). Зауважимо, що це тільки офіційно зареєстровані атаки, тобто ті, які вдалося виявити. Кількість атак, що встановити не вдалося, підрахувати неможливо.

Зазначений графік відтворює тільки один з безлічі варіантів, які зловмисники можуть використовувати для власного збагачення або знищення інформації, що належить якій-небудь компанії, і, відповідно, самої компанії.

Тому нині неможливо уявити функціонування будь-якої інформаційно-комунікаційної системи організації чи промислового підприємства без систем захисту інформації.

Водночас захист даних у будь-яких мережах, починаючи від відкритих інтернет-мереж та IoT [2] і завершуючи мережами промислових об’єктів, без застосування комплексних антивірусних рішень є достатньо складним, оскільки кількість загроз збільшується щодня й тільки проактивний підхід до безпеки дає змогу ефективно виявляти та протистояти сучасним кіберзагрозам [3]. Рішення EDR наразі є лідерами ринку для протидії сучасним кіберзагрозам. На їх основі було створено XDR- та XSOAR-системи, що доповнюють та автоматизують захист кінцевих точок мереж.

Використання EDR-систем потребує достатнього рівня знань у сфері кібербезпеки, навичок розслідування та реагування на інциденти [4], а також високого рівня обізнаності про архітектуру рішення та розуміння "сліпих зон", особливостей

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

та недоліків типового EDR-рішення для попередження комплексних атак на інфраструктуру, зокрема від DOS/DDOS-атак і до XSS-атак [5]. Аналіз архітектури агента EDR необхідний, щоб зрозуміти особливість типового рішення та висунути низку рекомендацій, що дадуть змогу ефективно виявляти й протидіяти комплексним атакам на ІКС у майбутньому.

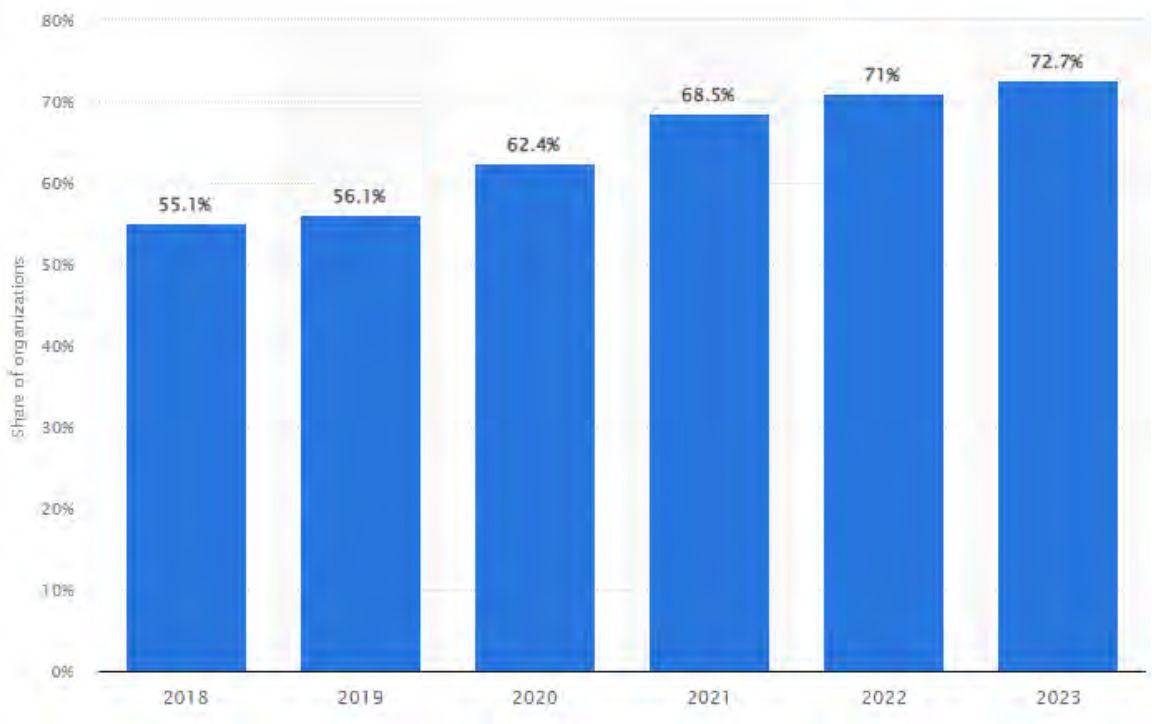


Рис. 1. Річна частка організацій, постраждалих від атак програм-вимагачів у всьому світі впродовж 2018–2023 рр.

Постановка завдання

Метою статті є розроблення методу підвищення ефективності використання EDR для зниження ризиків компрометації ІКС інформаційних, промислових та інфраструктурних об’єктів з погляду ефективного перерозподілу та використання наявних механізмів систем захисту кінцевих точок EDR, команди з кібербезпеки та інших ресурсів, призначених для здійснення заходів з організації безпеки на підприємстві, в установі чи організації.

Для реалізації окресленої мети виконуються такі завдання: огляд та аналіз наявних систем EDR, аналіз архітектури EDR-рішень та агентів EDR, особливостей їх використання, логіки побудови методів і механізмів виявлення загроз для системи з боку зловмисників і зловмисного коду. Окремо надаються рекомендацій щодо організації ІКС для її захисту загалом та окремих елементів, а також з огляду на наявні сили (команда із кіберзахисту, її кваліфікація та рівень обізнаності в архітектурі EDR-рішень) та засоби (елементи EDR-систем) для організації захисту.

Аналіз публікацій

Рішення EDR – це достатньо нова технологія, основана на механізмах так званих "класичних антивірусів" NGA (*Next Generation Antivirus*), що отримав таку назву з 2013 р. [7]. Оскільки цей клас рішень надавав небачену до цього можливість у реальному часі реагувати на загрози (*Prevention*), EDR швидко стали популярними в різних сегментах економіки.

Технологія показала свою ефективність завдяки застосуванню сенсорів – програмних давачів, що EDR розподіляє в операційній системі та надалі використовує для моніторингу активності [7].

У цьому разі впровадження технології за умови ігнорування рекомендацій призводить до того, що зловмисники можуть здійснювати атаки різних типів для обходу EDR: обхід конфігурації, обхід сприйняття, логічний обхід та обхід класифікації [3, 7]. Додаткові рекомендації до експлуатації EDR зазвичай подаються у вигляді звітів про обхід конкретного рішення та містять інструкції для команд з кібербезпеки про виправлення поданої

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

вразливості [4, 6]. Недолік такого підходу полягає в тому, що рекомендації в цьому разі точкові й не допоможуть захистити себе всебічно. Для значного посилення безпеки використання EDR недостатньо убезпечити себе від одного типу обходу EDR – необхідно починати посилювати безпеку з моменту встановлення рішення в межах корпоративної інфраструктури.

Системи захисту кінцевих точок

Системи захисту кінцевих точок EDR (*Endpoint Detection and Response*) [6–8] – це корпоративні антивірусні рішення, що забезпечують багаторівневий підхід до захисту кінцевих точок у межах корпоративної інфраструктури та набули значного поширення в останні роки. Вони використовуються для виявлення та реагування на загрози на кінцевих точках, таких як настільні ПК, ноутбуки, сервери. Також для забезпечення безпеки хмарних сховищ і мобільних пристроїв було створено XDR-рішення, що фактично є розширеннями класичних EDR.

Особливістю, що відрізняє EDR-рішення від EPP (*Endpoint Protection Platform*) – гілки розвитку класичних антивірусів, – є здатність до реагування на інцидент безпеки: блокування процесів та ізоляції зараженого пристрою. Ізоляція означає, що EDR забороняє хід трафіку по всіх портах, окрім виділеного, зазвичай 443, що спілкується з вебконсоллю EDR. Фактично в момент ізоляції доступ зловмисника до пристрою переривається, як і будь-які інші під'єднання, і обмежений віддалений доступ до хоста може мати тільки оператор консолі EDR.

За останнє десятиліття рішення EDR значно вдосконалилися. Сучасні EDR зазвичай інтегровані з іншими рішеннями безпеки, такими як SIEM і платформи аналізу загроз (*threat intelligence*), щоб забезпечити більш повне покриття для корпоративної інфраструктури.

Архітектура EDR

EDR-рішення містять серверну частину та агентів – невеликих за розміром програм, що встановлюються на кінцеві точки (ПК та сервери). Серверна частина рішення, що забезпечує доступ до менеджменту всіх агентів у форматі вебсторінки, може бути розгорнута в хмарному рішенні або на окремому сервері (*on-prem*). Друге рішення є дорожчим і здебільшого підходить для компаній закритого типу або державних установ та об'єктів критичної інфраструктури.

Наразі EDR конкурують з XDR, що фактично є їх розширеною версією, але щодо бізнесу вони ефективні для різних типів компаній. Якщо EDR більш призначені для захисту фізичних і віртуальних кінцевих точок (робочих станцій фахівців, серверів, віртуальних станцій у контейнерах), то XDR розширюється ще на хмарне середовище, смартфони тощо.

Також варто зауважити, що 2024 р. такі інструменти, як EDR, містять достатню кількість додаткових модулів, що дають змогу розширити це базове корпоративне антивірусне рішення в бік, необхідний бізнесу. Для прикладу, додатково до базового функціоналу *Crowdstrike*, призначеного саме для захисту кінцевих точок, можна докупити модулі моніторингу внесення критичних змін, моніторингу USB-пристроїв, пошуку загроз, менеджменту застосунків тощо. Сучасні EDR дають змогу легко масштабувати рівень видимості в умовах корпоративної інфраструктури, будучи певним "конструктором" інструментарію для команди з кібербезпеки.

EDR є класом корпоративних рішень, що активно використовуються в сучасних операційних центрах безпеки (SOC) [9]. SOC є пунктом, де оцінюється кожна подія, що стосується безпеки кінцевих точок і збирається агентом EDR, в якому аналітики з кібербезпеки всебічно розглядають зібрану інформацію та приймають остаточне рішення про інцидент безпеки. Для кожного із спрацювань, що було сформовано з цих подій, аналітики SOC мають вирішити, як їх класифікувати та як діяти надалі.

Архітектура агентів EDR як базових елементів EDR-систем

Агент EDR – це невелика програма, що є базовим складником системи. Вона контролює та споживає дані з компонентів сенсора, виконує базовий аналіз і визначає, чи відповідає активність або серія подій поведінці зловмисника. Далі агент EDR пересилає телеметрію на головний сервер, який аналітик надалі бачитиме як хмарний складник EDR-рішення, далі – вебконсоль, що аналізує події від усіх агентів, розгорнутих в інфраструктурі

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

бізнесу. Якість інформації, що отримується агентами, багато в чому визначає якість захисту за допомогою EDR, тому їх налаштування, параметри та особливості використання є базовими для ефективного застосування EDR загалом.

Більшість EDR-рішень спрямована на те, щоб навантажувати кінцеві точки менше, ніж на 1–5% CPU. Також варто зауважити, що розгортання рішення типу EDR можливе тільки після аналізу та оптимізації локальної мережі компанії, впорядкування всіх наявних активів, укладання списків використовуваних програм і створення моделі загроз, моделі порушника та написання внутрішніх політик. Основна мета створення переліченої документації – не завадити бізнес-процесам компанії, оскільки, як побачимо далі, агент має здатність блокувати процеси без можливості легко обійти блокування.

Якщо агент вважає, що певна активність є аномальною і варта уваги, він може виконати одну з таких дій [10]:

– зареєструвати зловмисну активність у вигляді сповіщення про інцидент, надісланого до консолі – інформаційної панелі EDR, або перенаправити сповіщення в центр агрегації даних SIEM;

– заблокувати виконання зловмисної операції, повернувши програмі, яка виконує дію, значення, що вказують на збій;

– увести в оману зловмисника, повернувши йому неправильні значення, такі як неправильні адреси пам’яті або модифіковані маски доступу, що змусить зловмисне програмне забезпечення вважати, що операція завершилася успішно, навіть якщо подальші дії не вдається виконати.

Кожен давач, що є складником агента EDR на хості, слугує загальній меті: збору телеметричних показників. Простіше, телеметрія – це необроблені дані, що генеруються компонентом агента або самим хостом, та фахівці SOC-центру можуть аналізувати їх як вручну, так і за допомогою вбудованих аналітичних механізмів EDR (як машинне навчання та динамічний аналіз), щоб визначити, чи мала місце зловмисна активність. Кожна дія в системі – від відкриття файлу до створення нового процесу – генерує певну форму даних для поповнення телеметрії. Ця інформація є відправною точкою у внутрішній логіці сповіщення про інциденти безпеки.

Можна порівняти телеметрію з показниками, що збирає радіолокаційна система: радари використовують електромагнітні хвилі для виявлення присутності, курсу та швидкості об’єктів у певному діапазоні. Коли радіохвиля відбивається від об’єкта й повертається до радіолокаційної системи, вона створює точку на дисплеї радару, яка вказує на те, що там щось є [11]. Використовуючи ці точки даних, процесор радарної системи може визначити такі параметри, як швидкість, місце розташування та висоту об’єкта, а потім обробляти кожен випадок по-різному. Наприклад, система може реагувати на об’єкт, що летить на низькій швидкості на малій висоті, інакше, ніж на об’єкт, що летить на значній швидкості на великій висоті. Це дуже схоже на те, як EDR обробляє телеметрію, зібрану його сенсорами. Сама по собі інформація про те, як було створено процес або отримано доступ до файлу, рідко забезпечує достатній контекст для прийняття обґрунтованого рішення щодо подальших дій. Крім того, процес, виявлений радаром, може припинитися будь-якої миті. Тому важливо, щоб телеметрія, яка надходить до EDR, була якомога повнішою.

EDR передає інформацію телеметрії до агента, а потім, якщо необхідно, до хмарного сховища, у компонентах яких прописана комплексна логіка виявлення загроз. Алгоритми логіки виявлення аналізують всю доступну телеметрію й за допомогою внутрішніх методів, зокрема евристики навколишнього середовища або бібліотеки статичних сигнатур, установлює, чи була активність зловмисною і чи досягла вона порогу критичності для створення повідомлення про атаку для аналітиків або блокування процесу для запобігання комплексній атаці.

Якщо телеметрія – це виявлені об’єкти на радарі, то сенсори – це передавач, дуплексор і приймач, тобто компоненти, що відповідають за виявлення об’єктів і формування їх у повідомлення для аналітиків, що працюють з консоллю EDR. Тоді як радіолокаційні системи постійно відправляють повторні сигнали до об’єктів, щоб відстежувати їх переміщення, давачі EDR працюють трохи пасивніше, перехоплюючи дані, що проходять крізь внутрішній процес, витягуючи інформацію та пересилаючи її центральному агенту. Оскільки ці давачі мають бути вбудовані в якийсь системний процес, також необхідно, щоб вони працювали неймовірно швидко. Середньостатистичний давач, який відстежує запити до реєстру, виконує свою роботу за 5 мс, перш ніж операція з реєстром буде дозволена та зможе продовжуватись. Це не здається значною проблемою, доки не буде взято до уваги, що в сучасних системах за секунду можуть відбуватися тисячі запитів до реєстру.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

Маленька затримка у 5 мс, що виникне в процесі оброблення 1 тис. подій, призведе до п'ятисекундної затримки в роботі системи. Більшість користувачів вважатимуть це неприйнятним, що відштовхне клієнтів від використання EDR взагалі. Хоча *Windows* має численні джерела телеметрії, продукти EDR, як правило, використовують лише деякі з них. Це пов'язано з тим, що багатьом джерелам бракує якості або кількості даних, вони можуть не відповідати вимогам безпеки комп'ютера або бути важкодоступними.

Деякі компоненти сенсора вбудовані в операційну систему, наприклад, у журнал подій ОС. EDR також можуть впроваджувати в систему драйвери, DLL, що перехоплюють функції, і мініфільтри, що будуть компонентами сенсорів. Фахівці атакуючих команд *(red team*), що превентивно виконують пошук вразливостей організації для того, щоб зменшити ризик успішної кібератаки, здебільшого дбають про запобігання, обмеження або нормалізацію (наприклад, змішування з потоком) зібраної сенсором телеметрії. Метою цієї тактики є зменшення кількості "точок на радарі", тобто показників, які алгоритми EDR можуть зіставити та використати для створення детального сповіщення про атаку для аналітика, що міститиме всю інформацію про активність зловмисника, а також попередить виконання зловмисних дій, блокуючи їх запуск. Власне, ми намагаємося згенерувати *False Negative* алерт, що згадувався раніше як тип алертів у SOC-центрі. Розуміючи кожен компонент давача EDR і телеметричні показники, які він збирає, можемо приймати обґрунтовані рішення щодо реагування на підтверджені інциденти безпеки та запобігання обходу корпоративного EDR.

Виявлення інцидентів безпеки – це логіка, що пов'язує окремі фрагменти телеметрії з певною поведінкою, поміченою в системі. Механізм виявлення може перевіряти окрему умову (наприклад, наявність файлу, геш якого збігається з гешем відомого шкідливого програмного забезпечення) або складну послідовність подій, що надходять із багатьох різних джерел (наприклад, що був створений дочірній процес *chrome.exe*, який потім зв'язався через TCP-порт 88 з контролером домену).

Як правило, інженер, що проєктує механізм виявлення, пише правила на основі наявних сенсорів. Вони мають ретельно зважати на масштаб, оскільки виявлення, вірогідніше за все, вплине на значну кількість організацій. З іншого боку, інженери з виявлення, які працюють в організації клієнта, який замовляє EDR для своєї компанії, найчастіше члени команди захисників (*Blue Team*), можуть створювати правила, що розширюють можливості EDR за межами тих, що надає постачальник ПЗ, щоб пристосувати виявлення інцидентів безпеки до потреб інфраструктури (створення списків, дозволених і заборонених для виконання програм, створення списку власних індикаторів компрометації, написання специфічних для організації правил тощо).

Логіка виявлення EDR зазвичай існує в агенті та підпорядкованих йому давачах або у внутрішній системі збору даних (системі, якій підпорядковуються всі агенти організації), до якої аналітики мають доступ за допомогою вебконсолі. Іноді вона функціонує в певній комбінації цих двох систем. У кожного підходу є переваги й недоліки. Виявлення, реалізоване в агенті або його давачах, може дозволити EDR вжити негайних превентивних заходів (блокування, ізоляція тощо), але не дасть йому змоги проаналізувати складну ситуацію, коли дій зловмисника багато й наявна значна кількість індикаторів компрометації. І навпаки, виявлення, реалізоване у внутрішній системі збору даних, може підтримувати величезний набір правил виявлення, але призводить до затримок у вжитті будь-яких попереджувальних заходів.

Усі EDR-продукти, що є на ринку, побудовані за однією логікою, яка відрізняється від платформи до платформи, на якій встановлено рішення [12, 13]. У цій роботі розглядаємо алгоритми й методи, що використовуються на платформі *Windows*, оскільки наразі вона залишається найбільш популярною операційною системою у світі (69% всіх користувачів; для порівняння – *macOS* застосовують лише 21% користувачів), а це приблизно 1,4 більйона активних пристроїв.

"Крихкі" та "надійні" методи виявлення спрацювань

Одним із способів задовольнити потреби клієнтів є використання комбінації так званих "крихких" і "надійних" методів виявлення інцидентів безпеки.

Крихкі засоби призначені для виявлення певного артефакту, наприклад простого рядка або геш-підпису, який зазвичай асоціюється з відомим шкідливим програмним забезпеченням. Надійні

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

методи спрямовані на виявлення поведінки й можуть підтримуватися моделями машинного навчання, навченими для певного середовища. Обидва типи виявлення мають місце в сучасних системах сканування, оскільки вони допомагають збалансувати хибні спрацювання (*False Positive*) та хибні негативні спрацювання (*False Negative*).

Наприклад, виявлення, побудоване на основі гешу шкідливого файлу, дуже ефективно визначає певну версію цього файлу, але будь-яка незначна зміна файлу змінить його геш, що призведе до збою в роботі правила виявлення. Ось чому такі правила називають "крихкими" – вони дуже специфічні, часто спрямовані на один артефакт. Це означає, що ймовірність хибнопозитивного спрацювання майже відсутня, тоді як ймовірність хибнонегативного спрацювання дуже висока.

Незважаючи на недоліки, ці системи виявлення пропонують явні переваги для команд кібербезпеки. Їх легко розробляти та підтримувати, тому інженери можуть швидко змінювати їх відповідно до потреб організації. Вони також можуть ефективно виявляти деякі поширені атаки. Наприклад, єдине правило для виявлення немодифікованої версії інструменту експлуатації *Mimikatz* має величезну користь, оскільки його рівень помилкових спрацювань майже нульовий, а ймовірність зловмисного використання інструменту висока.

Попри це інженер з виявлення має ретельно продумати, які дані застосовувати для створення правил для "крихких" спрацювань. Якщо зловмисник може простими способами змінити індикатор, уникнути виявлення стає набагато легше. Наприклад, якщо програма перевіряє наявність файлу *mimikatz.exe*, зловмисник може просто змінити ім’я файлу на *mimicats.exe* та обійти логіку правила. З цієї причини найкращі правила "крихких" виявлень спрямовані на атрибути, які або незмінні, або їх важко модифікувати.

З іншого боку, надійний набір правил, підкріплений моделлю машинного навчання, може позначити змінений файл як підозрілий, оскільки він є унікальним для середовища або містить певний атрибут, якому алгоритм класифікації надає велике значення. Більшість надійних засобів виявлення – це просто правила, які ширше намагаються бути спрямованими на метод. Ці типи виявлень обмінюють свою особливість на здатність виявляти атаку в більш загальному вигляді, зменшуючи ймовірність хибнонегативних результатів унаслідок збільшення ймовірності хибнопозитивних.

Хоча індустрія схильна надавати перевагу "надійним" методам виявлення, вони мають недоліки. Якщо порівняти з "крихкими" сигнатурами, "надійні" правила набагато важче розробити через їх складність. Крім того, інженер з виявлення має брати до уваги терпимість організації до хибнопозитивних спрацювань і задатися питанням: яку кількість хибнопозитивних спрацювань може обробити внутрішній SOC-центр, аби не знизити свою продуктивність і не заробити так звану *alert fatigue*, тобто нездатність аналітика, що постійно витрачає час на закриття неінформативних спрацювань, відреагувати на справді важливу аномалію в системі. Через це більшість EDR застосовують гібридний підхід, упроваджуючи "крихкі" методи для виявлення очевидних загроз і "надійні" – для виявлення тактик і технік зловмисників у більш загальному плані.

Одним із небагатьох постачальників EDR, який публічно розкриває свої правила виявлення, є *Elastic Stack*. Правила SIEM публікуються в репозиторії *GitHub*, і ці правила містять чудові приклади як крихких, так і надійних виявлень.

Наприклад, розглянемо правило *Elastic* для виявлення спроб *Kerberoasting*, які використовують *Bifrost*, інструмент *macOS* для взаємодії з *Kerberos*. *Kerberoasting* – це метод отримання квитків *Kerberos* і злому їх для розкриття даних службових облікових записів.

Це правило перевіряє наявність певних аргументів командного рядка, які підтримує *Bifrost*. Зловмисник може банально обійти це виявлення, перейменувавши аргументи у вихідному коді (наприклад, змінивши *-action* на *-dothis*), а потім перекомпілювавши інструмент. Крім того, хибне спрацювання може статися, якщо не пов’язаний з ним інструмент підтримує аргументи, перелічені в правилі.

Із зазначених причин правило може здатися поганим детектором. Але варто пам’ятати, що не всі зловмисники діють на одному рівні й чимало груп загроз продовжують використовувати готові інструменти, доступні в популярних фреймворках, як *Metasploit*. Це правило слугує для виявлення тих, хто застосовує базову версію *Bifrost* і не більше.

Через вузьку спрямованість правила *Elastic* має доповнити його більш надійним виявленням, яке закриває очевидні прогалини. Розв’язанням проблеми

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

в цьому разі стає додаткове правило, створене розробником, яке закриває сліпі місця першого.

Це правило спрямоване на нетипові процеси, що створюють вихідні з'єднання з TCP-портом 88, стандартним портом *Kerberos*. Хоча це правило містить деякі прогалини для усунення помилкових спрацювань, загалом воно більш надійне, ніж крихке правило запуску *Bifrost*.

Більшість агентів EDR прагнуть до балансу між крихким і надійним виявленням, але роблять це непрозоро, тому організаціям може бути дуже складно забезпечити покриття, особливо з рішеннями, що не підтримують створення окремих налаштованих користувацьких правил. Із цієї причини інженери команди безпеки мають тестувати та перевіряти виявлення за допомогою таких інструментів, як *Atomic Test Harnesses* від *Red Canary*.

Типи агентів EDR

Як зловмисники, так і інженери команди безпеки мають приділяти пильну увагу типу агенту EDR, розгорнутому на кінцевих точках. Розглянемо типи (або ж побудови) агентів.

1. Базовий

Агенти містять окремі частини, кожна з яких має свою мету й тип телеметрії, яку вона може збирати. Найчастіше агенти мають такі компоненти:

– статичний сканер: застосунок або компонент самого агента, що виконує статичний аналіз зображень, таких як *Portable Executable* (PE) файли або довільні діапазони віртуальної пам'яті, щоб визначити, чи є їх вміст шкідливим. Статичні сканери зазвичай становлять основу антивірусних сервісів;

– DLL-функція перехоплення: DLL, що відповідає за перехоплення викликів певних функцій інтерфейсу прикладного програмування (API);

– драйвер ядра: драйвер режиму ядра, що відповідає за впровадження перехоплюючої DLL у цільові процеси та збір специфічної для ядра телеметрії;

– служба агента: ПЗ, відповідальне за агрегування телеметрії, створеної двома попередніми компонентами. Воно корелює дані або генерує сповіщення, щоб потім передати зібрану інформацію на централізований сервер EDR.

На рис. 2 зображено найпростішу архітектуру агентів, яку нині застосовують комерційні продукти.

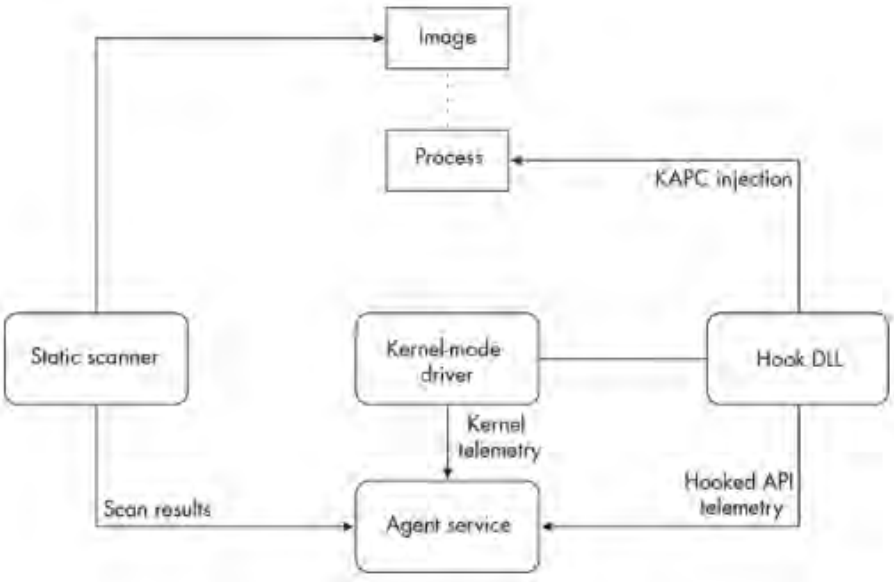


Рис. 2. Базова архітектура агента

Як бачимо, ця базова побудова має небагато джерел телеметрії. Три давачі (сканер, драйвер і DLL-функція перехоплення функцій) надають агенту дані про події створення процесів, виклики функцій, які вважаються чутливими до атак (наприклад, k*ernel32! CreateRemoteThread*), сигнатури файлів і часто віртуальну пам'ять процесу. Така схема може забезпечити достатнє покриття для деяких випадків використання, але більшість комерційних продуктів EDR нині виходять далеко за межі цих можливостей. Наприклад, цей базовий EDR не зможе виявити файли, що створюються, вилучаються або шифруються на хості.

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

2. *Проміжний*

Хоча базовий агент може збирати значну кількість цінної інформації, на основі якої можна створювати виявлення, ці дані можуть не давати повної картини дій, що виконуються на комп'ютері [15]. Програмні продукти для захисту кінцевих точок, що нині розгортаються в корпоративних середовищах, вже істотно розширили свої можливості для збору додаткової телеметрії.

Більшість агентів EDR наразі належать до середнього рівня складності [16]. Ці агенти не лише впроваджують нові давачі, але й використовують джерела телеметрії, властиві операційній системі. Доповнення на цьому рівні можуть містити:

– драйвери мережних фільтрів: виконують аналіз мережного трафіку для виявлення ознак зловмисної активності;

– драйвери фільтрів файлової системи: спеціальний тип драйверів, що можуть відстежувати операції у файловій системі комп'ютера;

– споживачі ETW: компоненти агента, які можуть слідкувати за подіями, створеними операційною системою хоста або сторонніми програмами;

– компоненти раннього запуску антивірусного програмного забезпечення (ELAM): функції, що надають підтримуваний *Microsoft* механізм завантаження драйвера антивірусного програмного забезпечення перед іншими службами запуску завантаження, щоб контролювати ініціалізацію інших драйверів завантаження. Ці компоненти також дають змогу отримувати *Secure ETW* події, спеціальний тип подій, що генеруються групою захищених постачальників подій.

Хоча сучасні EDR можуть не реалізовувати всі перелічені компоненти, зазвичай використовується драйвер ELAM, розгорнутий разом з основним драйвером ядра.

На рис. 3 показано, як може виглядати більш сучасна архітектура агента.

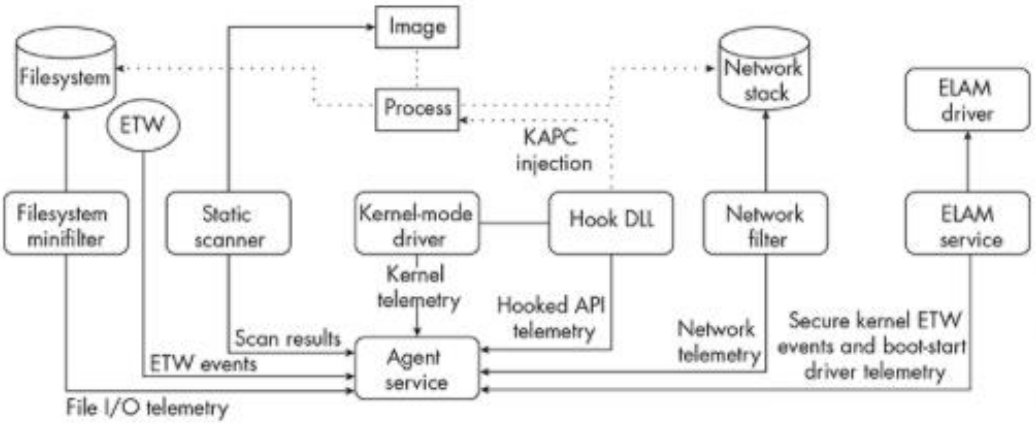


Рис. 3. Архітектура проміжного агента

Ця побудова основана на базовій архітектурі та додає багато нових давачів, з яких можна збирати телеметричні показники. Наприклад, EDR з проміжною архітектурою може відстежувати події файлової системи, зокрема створення файлів, отримувати інформацію від провайдерів ETW, що надають дані, які агент інакше не зміг би зібрати, і спостерігати за мережним зв'язком на хості крізь драйвер фільтра, що потенційно дає змогу агенту виявляти активність маячків командного рядка й команд, що запускаються через неї. Це також додає рівень відмовостійкості, щоб у разі виходу з ладу одного давача інший міг замінити його.

3. *Розширений*

Деякі рішення EDR реалізують більш просунуті функції для моніторингу ділянок системи, які їх цікавлять. Ось два приклади таких функцій:

– гіпервізори: забезпечують перехоплення системних викликів, віртуалізацію певних компонентів системи та ізольоване виконання коду. Вони також дають агенту змогу відстежувати переходи у виконанні між гостьовою машиною та хостом. Зазвичай вони використовуються як компонент захисту від програм-вимагачів та експлойтів;

– обман зловмисника: надає неправдиві дані замість того, щоб запобігти виконанню шкідливого коду. Це може призвести до того, що зловмисник зосередиться на налагодженні свого інструментарію,

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

не усвідомлюючи, що інформація, яку він отримав від системи, була підроблена.

Це досить специфічні для конкретного продукту доповнення і наразі вони не є загальновживаними. Крім того, багато компонентів цієї категорії більше пов'язані зі стратегіями запобігання, ніж із виявленням. Однак з часом деякі розширені функції можуть стати більш поширеними, а нові, ймовірно, будуть винайдені.

Недоліки сучасних EDR

Говорячи про слабкі місця сучасних EDR, не можна обійти теми їх довіри репутації файлу та наскільки структура файлу впливає на процес. Якщо ми розуміємо логіку роботи сенсорів EDR, то легко зможемо проаналізувати, наскільки репутація файлу, сформована іншими вендорами на сервісах як *VirusTotal*, впливає на кінцевий результат аналізу конкретного постачальника ПЗ. І навпаки, якщо файл ще ніде не був проаналізований і його геш не відомий, це означає що він не має поганої репутації, то яка вірогідність його блокування навіть просунутими рішеннями EDR. Нині навіть розрекламовані EDR, такі як *SentinelOne, Trellix, ESET* тощо, що продаються за тисячі доларів, підлягають обходу, якщо зловмисник добре знається на логіці їх сенсорів.

Чимало сучасних рішень EDR не можуть належним чином аналізувати бінарні файли, розроблені за допомогою нового покоління розробки, або, якщо перефразувати, – нетипових мов програмування. Оскільки антивірусні модулі EDR зазвичай розробляються для статичного аналізу зловмисного програмного забезпечення, написаного мовами C, C++ і C#, вони, коли стикаються з файлом зі структурою та виразами, що не є типовими, не можуть застосувати для нього звичні методи аналізу та пропускають далі, що б цей файл не робив далі. У кращому випадку будуть заблоковані дочірні процеси, такі як запуск командного рядка, в гіршому – вони будуть зчитані як нормальна поведінка, оскільки EDR не виявив достатньо індикаторів, що вказували б на хід атаки.

Друга річ, на яку необхідно звернути увагу, це те, наскільки нормально виглядає програма та дії, які вона виконує, для конкретного середовища [17]. Нормалізація, частково пов'язана з репутацією файлів, посідає дуже важливе місце серед методів обходу EDR. Для прикладу візьмемо кейс із простою програмою виведення *MsgBox* за допомогою скриптингової мови *AutoIt*, що часто використовується в сценаріях скриптингу, і спробу перевірити її за допомогою *VirusTotal*, як зробили дослідники сервісу *Secunnix* – отримано результат, що в багатьох рішеннях EDR показники виявлення різняться залежно від того, 32-розрядний чи 64-розрядний файл, а також, чи має файл піктограму, чи ні. Програмне забезпечення, що має значок і скомпільоване як 64-розрядне, рідше спостерігалося *VirusTotal* як зловмисне, а 32-розрядне програмне забезпечення, скомпільоване без піктограм, вважалося більш ризикованим і отримало більше спрацювань.

Тож, найпростіші речі, які можна зробити, щоб зменшити рівень виявлення ПЗ антивірусом, це:

– компіляція зловмисного ПЗ до x64;

– підготовка піктограми;

– якщо можливо, створення ПЗ, що буде застосунком GUI замість CLI;

– уникнення високої ентропії;

– використання дуже популярних, відомих вебсайтів, як C2 для крадіжки даних.

Кожен з перелічених пунктів необхідно брати до уваги під час експлуатації EDR, аби не припустити обходу рішення зловмисником [18].

Метод підвищення ефективності роботи EDR та посилення безпеки ІКС

Тепер, знаючи про особливості побудови агентів EDR та про слабкі місця типового EDR-рішення, можемо сформувати метод використання EDR. Це дасть змогу підвищити ефективності застосування цього рішення.

Для найкращих показників працеспроможності EDR важливо підготуватись до його розгортання в корпоративному середовищі заздалегідь. Перед закупівлею та початком розгортання важливо виконати певні дії.

1. Оптимізувати ресурси локальної мережі та сегментувати мережу, якщо це не було виконано раніше.

2. Ввести стандарт іменування хостів у мережі.

3. Оцінити масштаби й навички фахівців наявного SOC-центру.

4. Провести інвентаризацію всіх активів і програмного забезпечення.

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) ISSN 2524-2296 (online)

5. Дослідити доцільність обраного рішення та його інтеграцію з уже присутніми інструментами.

6. Протестувати рішення в межах тест-драйву.

Безпосередньо метод, що дасть змогу покращити ефективність використання EDR-систем, передбачає послідовність дій, наведених нижче. У певних аспектах він перегукується із заходами, що мають відбутися до початку впровадження запропонованого методу.

1. ***Сегментація мережі*** дасть змогу розмежувати критично важливі відділи компанії від відділів, що підпадають під ризик зараження шкідливим ПЗ (всі відділи, що спілкуються із зовнішнім світом: технічна підтримка, рекрутери, маркетологи тощо). Також критично важливо створити окрему мережу для будь-яких пристроїв, що не є корпоративними, – смартфонів, особистих ПК тощо.

2. ***Оптимізація ресурсів локальної мережі*** дасть змогу їй працювати без затримок і значно покращить працездатність EDR, тому що, як говорили раніше, хмарна частина рішення є не менш важливою за частину агента на хості. Оскільки EDR має відповідати на запити в реальному часі, будь-яка затримка може бути критичною – від часу передачі телеметрії, повернення результату аналізу від хмарного середовища до часу блокування процесу та ізоляції кінцевої точки. Також якщо агент EDR стикнувся зі збоєм, необхідно, щоб інформація про несправність агента була якомога швидше надана аналітикам у вебконсолі.

3. ***Деактивація інших антивірусних рішень.*** Під час встановлення агента EDR на хости необхідно переконатися, що на них деактивовані всі інші антивірусні рішення, наприклад *Windows Defender*. Через особливості роботи сенсорів агента EDR вони будуть сприйняті як зловмисні іншим антивірусним ПЗ. Також таке сусідство може викликати колізію та призвести до несправності ПК.

4. ***Розроблення політик EDR.*** Розробленню політик EDR (виявлення та реагування окремо) приділяється особлива увага на початку розгортання рішення.

Зазвичай створюються три політики:

– легка – створена для високочутливих хостів, на яких небажано блокувати процеси;

– середня – створена за найкращими практиками, що пропонує постачальник ПЗ; зазвичай всі характеристики усереднені та видають найкращий баланс ефективності та продуктивності агентів;

– строга – політика, яку часто іменують "режимом атаки на компанію"; вона вмикається, коли компрометація хостів підтверджена та аналітики хочуть бачити всі аномалії, що можуть доповнити картину руху зловмисника по середовищу. Ця політика також ефективна під час тестування самого рішення та його чутливості до спроб його обходу.

Крім того, компанії часто утворюють окремі політики для різних відділів, щоб наголосити на особливості налаштувань безпеки для критичних відділів і структур. Можна створити полегшену політику, виняток, білий список, якщо EDR "заважає" працювати, але не уникати встановлення агента на пристрій.

5. ***Створення білого та чорного списків ПЗ*** необхідно для того, щоб під час такої глобальної дії, як розгортання рішення для захисту кінцевих точок, не заважати нормальним процесам бізнесу. ПЗ, дозволене для використання в компанії, має бути додано у виняток, якщо створює хибнопозитивні спрацювання. Варто зауважити, що білий список має застосовуватися з підвищеною обережністю. Використання чорного списку не обмежене. Білий та чорний списки основані на класичних сигнатурах, таких як геш, IP-адреса, домен. Винятки зі свого боку відрізняються від них тим, що прив'язуються до конкретного шляху або поведінки файлу, що робить їх більш застосовуваними й не такими критичними для загальної видимості агента. Також деякі вендори пропонують обмежити видимість певних файлів для агента на хості, але такі винятки вважаються занадто ризикованими й можуть використовуватися лише в окремих випадках, затверджених керівництвом компанії.

6. ***Застосування останніх версій програмного забезпечення*** агентів EDR. Упровадження практики оновлення програмного забезпечення є класичним підходом для будь-якого ПЗ. На жаль, часто ця практика не застосовується в багатьох системах захисту. Опція автоматичного оновлення має бути додана для всіх складників EDR.

7. ***Розроблення плейбуків з реагування на інциденти.*** Плейбуки з реагування на інциденти мають бути розроблені в межах керівництва SOC-центру й підлягають щорічному переоцінюванню їх ефективності, а також впроваджуються з огляду на нові можливості, що надає EDR, і нових методів реагування. Плейбуки мають брати до уваги особливість конкретного обраного EDR-рішення та час реагування на інцидент безпеки, а також

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

можливості активного блокування зловмисних дій з вебконсолі EDR.

8. ***Інвентаризація встановлених агентів і встановлення додаткових інструментів*** проводиться після завершення розгортання рішення та виконується для того, щоб отримати статистику з покриття рішенням усіх кінцевих точок компанії, та оцінити поточний рівень видимості та безпеки корпоративного середовища, і підвищити цей рівень за допомогою додаткових інструментів.

9. ***Постійне тестування та оновлення правил і конфігурацій EDR***. Виконання цього правила дасть змогу пришвидшити впровадження додаткових механізмів виявлення та реагування на найновіші методи, засоби та механізми, що використовуються зловмисниками для атак на ІКС.

Під час роботи з EDR у межах корпоративної інфраструктури важливо зважати на описані вище методи обходу EDR та особливості їх архітектури.

Розглянемо, як метод підвищення ефективності роботи EDR дає змогу попередити різні варіанти обходу EDR зловмисниками.

Обхід конфігурації можна попередити, якщо інженери SOC-центру коректно налаштують політики відповідно до потреб компанії та кращих практик індустрії. На цьому етапі важливо дати бізнесу розуміння, що безпека інформації – це завжди битва між зручністю та швидкістю й безпекою. Варто переоцінювати політики щоразу, коли постачальник ПЗ випускає відповідні оновлення, і тестувати нові опції для підтвердження їх ефективності. Також варто тримати рівень підозрілості агента до процесів на середньому або високому рівні для виявлення, і низькому або середньому – для блокування.

Обхід сприйняття можна попередити використанням додаткових інструментів, окрім EDR, наприклад зовнішніх сканерів вразливостей, IDS/IPS-систем, застосуванням SIEM тощо. Якщо один інструмент не відстежує певні процеси (виявити їх можна під час первинного тестування й далі під час експлуатації рішення), то варто переконатися, що ці процеси відстежуються іншим інструментом, який може генерувати спрацювання для SOC-команди.

Обхід логіки EDR можна попередити регулярним тестуванням правил EDR і кастомних правил за допомогою взаємодії з командою пентестерів в організації. Також обов'язковою практикою є оновлення сенсорів для автоматичного закриття відомих прогалин. Хорошою практикою є моніторинг даркнету для пошуку відомих методів обходу конкретного рішення та створення правил, що закриють для потенційного зловмисника шлях експлуатації цих методів.

Обхід класифікації можна попередити, налаштовуючи політики EDR-рішення на достатньому рівні чутливості для того, щоб більше подій ставилися під сумнів. Також у цьому разі доцільним буде участь інших інструментів безпеки, навіть якщо корпоративний EDR не побачить аномалії в трафіку, то IDS-система або DLP відправить повідомлення про перевищення ліміту надсилання інформації для користувача.

Висновки

У статті описано побудову типових рішень EDR та схеми їх агентів різної складності; проаналізовано методи обходу EDR та висунуто пропозиції з підходу до експлуатації рішення з огляду на сучасні архітектурні особливості EDR-рішень та на основі добутої інформації; сформовано відомості про специфіку цих рішень та наведено рекомендації щодо кращих практик, які можуть бути застосовані в командах фахівців із кібербезпеки для оптимізації та покращення роботи з EDR у мережах організацій різного типу, починаючи від державних органів і завершуючи промисловими об'єктами.

Подані рекомендації можуть бути застосовані під час формування процесів у команді з кібербезпеки, для покращення роботи з наявним рішенням EDR, а також у проведенні менеджменту ризиків та оцінюванні профілю інформаційної безпеки організації, зважаючи на особливості наявного інструментарію команди з кібербезпеки. Це дасть змогу ефективно використовувати наявні програмні, апаратні та людські ресурси, а також здійснювати ефективне планування подальшого розвитку системи кібербезпеки підприємств.

Список літератури

1. Annual share of organizations affected by ransomware attacks worldwide from 2018 to 2023 URL: https://www.statista.com/statistics/204457/businesses-ransomware-attack-rate/ (дата звернення 24.05.2024).

Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

2. Журило О., Ляшенко О. Архітектура та системи безпеки IoT на основі туманних обчислень, *Сучасний стан наукових досліджень та технологій в промисловості,* 2024, Вип. (1(27), С. 54–66. DOI: 10.30837/ITSSI.2024.27.054
3. Когут Ю. Кібервійна та безпека об'єктів критичної інфраструктури. Сідкон, 2021. 336 с.
4. Matt Hand. Evading EDR: The Definitive Guide to Defeating Endpoint Detection Systems. No Starch Press. 2023. 312 p.
5. Мерзлікін Є., Бабешко Є. Аналіз кібербезпеки веборієнтованих індустріальних IOT-систем. *Сучасний стан наукових досліджень та технологій в промисловості.* 2023. Вип. 2(24). С. 131–144. DOI: 10.30837/ITSSI.2023.24.131
6. Forrester Wave October 2023, URL: https://www.forrester.com/ (дата звернення 24.05.2024).
7. Баклан Я. А., Сєвєрінов О. В. Аналіз систем захисту кінцевих точок від складних загроз EDR (Endpoint Detection and Response). Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління: матеріали дванадцятої міжнар. наук.-практ. конф. 2022. Баку–Харків–Жиліна. 141 р. URL: https://openarchive.nure.ua/handle/document/24142
8. ISO/IEC 27035:2011 Information technology. Security techniques. Information security incident management, 2011.
9. Crowdstrike October 2023, URL: https://www.crowdstrike.com/ (дата звернення 24.05.2024).
10. Arfeen A., Ahmed S., Khan M. A., Jafri, S. F. A. Endpoint Detection and Response: A Malware Identification Solution. *International Conference on Cyber Warfare and Security (ICCWS).* 2021. DOI: 10.1109/ICCWS53234.2021.9703010
11. Сєвєрінов О. В., Хрєнов А. Г., Поляков А. О. Аналіз сучасних методів атак на автоматизовані системи управління військами та інформаційні мережі. *Системи обробки інформації*. 2015. Вип. 9. С. 101–104. URL: http://nbuv.gov.ua/UJRN/soi_2015_9_24
12. Exploring the History of Antivirus: Fusion Computing. URL: https://fusioncomputing.ca/history-of-antivirus/ (дата звернення 21.03.2024).
13. Сєвєрінов О. В., Шевцов В. О., Сокол-Кутиловська А. С. Аналіз сучасних методів атак на електронні ресурси органів управління. *Системи озброєння і військова техніка*. 2017. Вип. 1. С. 65–67. URL: http://nbuv.gov.ua/UJRN/soivt_2017_1_13 (дата звернення 21.03.2024).
14. Ушатов В., Сєвєрінов О. В. Проблеми оперативного виявлення і реагування на інциденти інформаційної безпеки *Global Cyber Security Forum: матеріали Першого міжнародного науково-практичного форуму*, 2019 С. 104–105. URL: https://openarchive.nure.ua/bitstreams/c2575d95-c877-47e6-aef8-2c19e286d900/download (дата звернення 21.03.2024).
15. FZE B. B. History of antivirus software. UKEssays. 2023. URL: https://us.ukessays.com/essays/information-technology/history-of-antivirus-software.php
16. Zhuravchak D., Dudykevych, V., Tolkachova, A. Дослідження структури системи виявлення та протидії атакам вірусів-вимагачів на базі endpoint detection and response. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка». 2023.* Вип. 3(19), С 69–82. DOI: https://doi.org/10.28925/2663-4023.2023.19.6982
17. Зубок В. Ю., Гончар С. Ф., Єрмошин В. В., Карасюк Г. О. Архітектурно-функціональне порівняння відомих платформ та систем кіберзахисту промислових об'єктів. *Електронне моделювання*, 2022, Вип. 44. Том 3. 65 с. DOI: 10.15407/emodel.44.03.065
18. Коробейнікова Т., Федорченко В. Системний моніторинг мережевої безпеки в тріаді SIEM-EDR-NDR. *Grail of Science*. 2023 Вип. 27. С. 354–360. DOI: https://doi.org/10.36074/grail-of-science.12.05.2023.055

References

1. "Annual share of organizations affected by ransomware attacks worldwide from 2018 to 2023", available at https://www.statista.com/statistics/204457/businesses-ransomware-attack-rate/ (last accessed 24.05.2024).
2. Zhurilo, O. and Lyashenko, O. (2024), "Architecture and security systems of IoT based on fog computing", ["Arkhitektura ta systemy bezpeky IoT na osnovi tumannykh obchyslen"], *Modern State of Scientific Research and Technologies in Industry*, No 1(27), P. 54–66. DOI: 10.30837/ITSSI.2024.27.054
3. Kogut, Y. (2021), *Cyber warfare and security of critical infrastructure objects, [Kiberviina ta bezpeka obiektiv krytychnoi infrastruktury],* Sidkon, 336 p.
4. Hand, M. (2023), *Evading EDR: The Definitive Guide to Defeating Endpoint Detection Systems*, No Starch Press, 312 p.
5. Merzlikin, Y., Babeshko, Y. (2023), "Cybersecurity analysis of web-oriented industrial IoT systems" ["Analiz kiberbezpeky veboriientovanykh industrialnykh iot-system"], *Modern State of Scientific Research and Technologies in Industry,* No. 2(24), P. 131–144. DOI: 10.30837/ITSSI.2023.24.131
6. "Forrester Wave October 2023", available at: https://www.forrester.com/ (last accessed: 24.05.2024).

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) *Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)*

7. Baklan, Y. and Severinov, O. (2022), "Analysis of endpoint protection systems against complex threats EDR (Endpoint Detection and Response)" ["Analiz system zakhystu kintsevykh tochok vid skladnykh zahroz EDR (Endpoint Detection and Response)"], Modern Trends in the Development of Information and Communication Technologies and Management Tools: materials of the twelfth international scientific-practical conference 2022, Baku Kharkiv Zhilina, 141 p., available at: https://openarchive.nure.ua/handle/document/24142
8. "ISO/IEC 27035:2011 Information technology. Security techniques. Information security incident management", 2011.
9. "Crowdstrike October 2023", available at: https://www.crowdstrike.com/ (last accessed: 24.05. 2024)-
10. Arfeen, A., Ahmed, S., Khan, M., Jafri, S. (2021), "Endpoint Detection and Response: A Malware Identification Solution". *International Conference on Cyber Warfare and Security (ICCWS).* DOI: 10.1109/ICCWS53234.2021.9703010
11. Severinov, O., Khrenov, A. and Polyakov, A. (2015), "Analysis of modern attack methods on automated control systems and information networks", ["Analiz suchasnykh metodiv atak na avtomatyzovani systemy upravlinnia viiskamy ta informatsiini merezhi"]*, Information Processing Systems,* No. 9, P. 101–104. available at: http://nbuv.gov.ua/UJRN/soi_2015_9_24
12. "Fusion Computing 'Exploring the History of Antivirus: Fusion Computing", available at: https://fusioncomputing.ca/history-of-antivirus/ (last accessed: 21.03.2024).
13. Severinov, O., Shevtsov, V., Sokol-Kutilovska, A. (2017), "Analysis of modern attack methods on electronic resources of management bodies" ["Analiz suchasnykh metodiv atak na elektronni resursy orhaniv upravlinnia"], *Weapons and Military Equipment Systems*, No 1, P. 65–67. available at: http://nbuv.gov.ua/UJRN/soivt_2017_1_13 (last accessed 21.03.2024).
14. Ushatov, V. and Severinov, O. V. (2019), "Problems of prompt detection and response to information security incidents" ["Problemy operatyvnoho vyiavlennia i reahuvannia na intsydenty informatsiinoi bezpeky"], Global Cyber Security Forum: materials of the First International Scientific and Practical Forum, P. 104–105. available at: https://openarchive.nure.ua/bitstreams/c2575d95-c877-47e6-aef8-2c19e286d900/download (last accessed 21.03.2024).
15. FZE, B. B. "History of antivirus software, UKEssays". 2023, available at: https://us.ukessays.com/essays/information-technology/history-of-antivirus-software.php
16. Zhuravchak, D., Dudykevych, V. and Tolkachova, A. (2023), "Research on the structure of the system for detecting and countering ransomware attacks based on endpoint detection and response", ["Doslidzhennia struktury systemy vyiavlennia ta protydii atakam virusiv-vymahachiv na bazi endpoint detection and response"], *Electronic Professional Scientific Publication "Cybersecurity: Education, Science, Technology"*, No 3(19), P. 69–82. DOI: 10.28925/2663-4023.2023.19.6982
17. Zubok, V., Honchar, S., Yermoshyn, V. and Karasyuk, H. (2022), "Architectural and functional comparison of known platforms and industrial cybersecurity systems", ["Arkhitekturno-funktsionalne porivniannia vidomykh platform ta system kiberzakhystu promyslovykh obiektiv"], *Electronic Modeling,* No 44, Vol. 3, 65 p. DOI: 10.15407/emodel.44.03.065
18. Korobeinikova, T., Fedorchenko, V. (2023), "System network security monitoring in the triad SIEM-EDR-NDR", ["Systemnyi monitorynh merezhevoi bezpeky v triadi SIEM-EDR-NDR"], *Grail of Science*, No. 27, P. 354–360. DOI: 10.36074/grail-of-science.12.05.2023.055

Надійшла 05.06.2024

Відомості про авторів / About the Authors

Шуліка Катерина Максимівна – Харківський національний університет радіоелектроніки, магістр кафедри безпеки інформаційних технологій, Харків, Україна; e-mail: kateryna.shulika@nure.ua; ORCID ID: https://orcid.org/0000–0003–2560–7426

Балагура Дмитро Сергійович – кандидат технічних наук, Харківський національний університет радіоелектроніки, доцент кафедри безпеки інформаційних технологій, Харків, Україна; e-mail: dmytro.balahura@nure.ua; ORCID ID: https://orcid.org/0009-0006-9839-3317

Смірнов Антон Олександрович – кандидат технічних наук, Харківський національний університет радіоелектроніки, доцент кафедри безпеки інформаційних технологій, Харків, Україна; e-mail: anton.smirnov@nure.ua; ORCID ID: https://orcid.org/0000-0003-4121-3902

Непокритов Дмитро Миколайович – Харківський національний університет Повітряних Сил імені Івана Кожедуба, доцент кафедри радіоелектронних систем пунктів управління Повітряних Сил, Харків, Україна; e-mail: ndn_ndn@ukr.net; ORCID ID: https://orcid.org/ 0000-0003-1752-8496

Литвин Андрій Володимирович – Харківський національний університет Повітряних Сил імені Івана Кожедуба, старший викладач кафедри радіоелектронних систем пунктів управління Повітряних Сил, Харків, Україна; e-mail: ravshan73@ukr.net; ORCID ID: https://orcid.org/0000-0003-1962-6356

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2024. №2 (28) *ISSN 2524-2296 (online)*

Shulika Kateryna – Kharkiv National University of Radio Electronics, M.Sc. at the Department of Information Technology Security, Kharkiv, Ukraine.

Balagura Dmytro – PhD (Engineering Sciences), Kharkiv National University of Radio Electronics, Associate Professor at the Department of Information Technology Security, Kharkiv, Ukraine.

Smirnov Anton – PhD (Engineering Sciences), Kharkiv National University of Radio Electronics, Associate Professor at the Department of Information Technology Security, Kharkiv, Ukraine.

Nepokrytov Dmytro – Ivan Kozhedub Kharkiv National Air Force University, Associate Professor at the Department of Radioelectronic Systems of Control Points of Air Forces, Kharkiv, Ukraine.

Lytvyn Andrii – Ivan Kozhedub Kharkiv National Air Force University, Senior Instructor at the Department of Radioelectronic Systems of Control Points of Air Forces, Kharkiv, Ukraine.

A METHOD OF USING MODERN ENDPOINT DETECTION AND RESPONSE (EDR) SYSTEMS TO PROTECT AGAINST COMPLEX ATTACKS

The **subject** of the research in this article is the architecture of Endpoint Detection and Response and the EDR agent as their base parts in terms of mechanisms for detecting and countering complex attacks on information and communication systems (ICS). The **aim** of the work is to develop of method for improving the efficiency of using Endpoint Detection and Response (EDR) to reduce the risks of compromising ICS information, industrial, and infrastructure objects by effectively redistributing and utilizing the available EDR mechanisms, the cybersecurity team, and other resources available for implementing security measures in an enterprise, institution, or organization. The article addresses the following **tasks**: reviewing and analyzing existing EDR systems, analyzing the architecture of EDR solutions and EDR agents, the features of their use, the logic behind the construction of methods and mechanisms for detecting threats to the system from malicious actors and malicious code. The task of providing recommendations for the organization of ICS is also separately addressed in terms of the need to protect the entire ICS and its individual elements, as well as in terms of the available resources (the cybersecurity team, their qualifications and level of awareness of the architecture of EDR solutions) and means (available EDR system elements) for organizing protection. The following **methods** are used: modeling attack mechanisms, modeling attacker behavior. The following **results** were obtained: general and specific recommendations were formulated for optimizing the operation of EDR systems and ensuring the effective use of EDR system elements in the information and communication networks of enterprises, organizations, and institutions of various types and orientations depending on the available resources and the information requiring protection. **Conclusions**: The identified recommendations for the application of EDR mechanisms for protecting information systems and networks allow optimizing the costs of creating a protection infrastructure and implementing security measures, taking into account the characteristics of the available tools and the training and awareness of the cybersecurity team both in terms of response time to threats and the complexity and cost of performing protection tasks.

Keywords: information and communication systems (ICS); EDR system; Security Opertion Center; EDR agent; threat intelligence; EDR policy; detection of vulnerabilities.

Бібліографічні описи / Bibliographic descriptions

Шуліка К. М., Балагура Д. С., Смірнов А. О., Непокритов Д. М., Литвин А. В. Метод використання сучасних систем захисту кінцевих точок (EDR) для убезпечення від комплексних атак. *Сучасний стан наукових досліджень та технологій в промисловості*. 2024. № 2 (28). С. 182–195. DOI: https://doi.org/10.30837/2522-9818.2024.2.182

Shulika, K., Balagura, D., Smirnov, A., Nepokrytov, D., Lytvyn, A. (2024), "A method of using modern endpoint detection and response (EDR) systems to protect against complex attacks", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (28), P. 182–195. DOI: https://doi.org/10.30837/2522-9818.2024.2.182

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2024. No. 2 (28)

АЛФАВІТНИЙ ПОКАЖЧИК

ALPHABETICAL INDEX

НАУКОВЕ ВИДАННЯ

СУЧАСНИЙ СТАН НАУКОВИХ ДОСЛІДЖЕНЬ ТА ТЕХНОЛОГІЙ В ПРОМИСЛОВОСТІ

Щоквартальний науковий журнал

№ 2 (28), 2024

SCIENTIFIC PUBLICATION

INNOVATIVE TECHNOLOGIES AND SCIENTIFIC SOLUTIONS FOR INDUSTRIES

Quarterly scientific journal

№ 2 (28), 2024

Відповідальний секретар журналу *І.Г. Перова*
Відповідальний за випуск *А.А. Коваленко*
Відповідальний за ліцензування *В.В. Косенко*
Редактор *Л.В. Кузьміна*
Комп'ютерна верстка *Л.Ю. Свєтайло*

Responsible secretary of journal *I. Perova*
Responsible for release *A. Kovalenko*
Responsible for licensing *V. Kosenko*
Editor *L. Kuzmina*
Computer layout *L. Svietailo*

Формат 60×84/8. Умов. друк. арк. 23,0.
Тираж 150 прим.

Format 60×84/8. Conventional printed sheets 23,0.
Edition of 150 copies.

Віддруковано з готових оригінал-макетів
в типографії ФОП Андреєв К.В.
Єдиний державний реєстр юридичних осіб
та фізичних осіб-підприємців.
Запис №24800170000045020 від 30.05.2003.

Printed from ready-made original layouts
in the printing house
of Individual Entrepreneur Andreev K.V.
Unified State Register of Legal Entities
and Individual Entrepreneurs.
Entry No. 24800170000045020 of 30.05.2003.

61157, Харків, вул. Акад. Богомольця, 9, кв. 50,
тел. +38 (063) 993-62-73
e-mail: ep.zakaz@gmail.com

fl. 50, 9, Acad. Bogomolets Str., Kharkiv, 61157,
тел. +38 (063) 993-62-73
e-mail: ep.zakaz@gmail.com