

Харківський національний
університет радіоелектроніки

Kharkiv National
University of Radio Electronics

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

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**

№ 3 (21), 2022

No. 3 (21), 2022

*Щоквартальний
науковий
журнал*

*Quarterly
scientific
journal*

Харків
2022

Kharkiv
2022

РЕДАКЦІЙНА КОЛЕГІЯ

Головний редактор
Бодянский Євгеній Володимирович,
д-р техн. наук, професор

Заступник головного редактора
Айзенберг Ігор Наумович,
канд. техн. наук, професор (США);
Шекер Серхат,
д-р техн. наук, професор (Туреччина)

Члени редколегії:

Артіух Роман Володимирович, канд. техн. наук;
Бабенко Віталіна Олексіївна, д-р екон. наук, канд. техн. наук, професор;
Безкорований Володимир Валентинович, д-р техн. наук, професор;
Гасімов Юсіф, д-р мат. наук, професор (Азербайджан);
Гопсєнко Віктор, д-р техн. наук, професор (Латвія);
Го Цян, д-р техн. наук, професор (КНР);
Джавад Хамісабаді, канд. техн. наук, доцент (Іран);
Зайцева Єлена, д-р техн. наук, професор (Словаччина);
Зачко Олег Богданович, д-р техн. наук, доцент;
Коваленко Андрій Анатолійович, д-р техн. наук, професор;
Костін Юрій Дмитрович, д-р екон. наук, професор;
Левашенко Віталій, д-р техн. наук, професор (Словаччина)
Лемешко Олександр Віталійович, д-р техн. наук, професор;
Малєєва Ольга Володимирівна, д-р техн. наук, професор;
Момот Тетяна Валеріївна, д-р екон. наук, професор;
Музика Катерина Миколаївна, д-р техн. наук, професор;
Назарова Галина Валентинівна, д-р екон. наук, професор;
Невлюдов Ігор Шакирович, д-р техн. наук, професор;
Опанасюк Анатолій Сергійович, д-р фіз.-мат. наук, професор;
Павлов Сергій Володимирович, д-р техн. наук, професор;
Перова Ірина Геннадіївна, д-р техн. наук, доцент;
Петленков Едуард, канд. техн. наук (Естонія);
Петришин Любомир Богданович, д-р техн. наук, професор (Польща);
Рубан Ігор Вікторович, д-р техн. наук, професор;
Семенець Валерій Васильович, д-р техн. наук, професор;
Сетлак Галина, д-р техн. наук, професор (Польща);
Терзіян Ваган Якович, д-р техн. наук, професор (Фінляндія);
Тєлєтов Олександр Сергійович, д-р екон. наук, професор;
Тимофєєв Володимир Олександрович, д-р техн. наук, професор;
Філатов Валентин Олександрович, д-р техн. наук, професор;
Чумаченко Ігор Володимирович, д-р техн. наук, професор;
Чухрай Наталія Іванівна, д-р екон. наук, професор;
Юн Джин, канд. фіз.-мат. наук, професор (КНР);
Ястремська Олена Миколаївна, д-р екон. наук, професор.

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>
E-mail редколегії: journal.itssi@gmail.com

EDITORIAL BOARD

Editor in Chief
Bodyanskiy Yevgeniy,
Dr. Sc. (Engineering), Professor, Ukraine

Deputy Chief Editor
Igor Aizenberg,
PhD (Computer Science), Professor (United States)
Serhat Seker,
Dr. Sc. (Engineering), Professor (Turkey)

Editorial Board Members:

Artiukh Roman, PhD (Engineering Sciences) (Ukraine);
Babenko Vitalina, Dr. Sc. (Economics); PhD (Engineering Sciences), Professor (Ukraine);
Bezkorovainyi Volodymyr, Dr. Sc. (Engineering), Professor (Ukraine);
Gasimov Yusif, Dr. Sc. (Mathematical), Professor (Azerbaijan);
Gopayenko Victor, Dr. Sc. (Engineering), Professor (Latvia);
Guo Qiang, Dr. Sc. (Engineering), Professor (P.R. of China);
Javad Khamisabadi, PhD (Industrial Management), Associate Professor (Iran);
Zaitseva Elena, Dr. Sc. (Engineering), Professor (Slovak Republic);
Zachko Oleh, Dr. Sc. (Engineering), Associate Professor (Ukraine);
Kovalenko Andrey, Dr. Sc. (Engineering), Professor, (Ukraine);
Kostin Yuri, Dr. Sc. (Economics), Professor (Ukraine);
Levashenko Vitaly, Dr. Sc. (Engineering), Professor (Slovakia);
Lemeshko Olexandr, Dr. Sc. (Engineering), Professor (Ukraine);
Malyeyeva Olga, Dr. Sc. (Engineering), Professor (Ukraine);
Momot Tetiana, Dr. Sc. (Economics), Professor, (Ukraine);
Muzyka Kateryna, Dr. Sc. (Engineering), Professor (Ukraine);
Nazarova Galina, Dr. Sc. (Economics), Professor (Ukraine);
Nevliudov Igor, Dr. Sc. (Engineering), Professor (Ukraine);
Opanasyuk Anatoliy, Dr. Sc. (Physical and Mathematical), Professor (Ukraine);
Pavlov Sergii, Dr. Sc. (Engineering), Professor (Ukraine);
Perova Iryna, Dr. Sc. (Engineering), Associate Professor (Ukraine);
Petlenkov Eduard, PhD (Engineering Sciences) (Poland);
Petryshyn Lubomyr, Dr. Sc. (Engineering), Professor (Poland);
Ruban Igor, Dr. Sc. (Engineering), Professor, (Ukraine);
Semenets Valery, Dr. Sc. (Engineering), Professor, (Ukraine);
Setlak Galina, Dr. Sc. (Engineering), Professor (Poland);
Terziyan Vagan, Dr. Sc. (Engineering), Professor, (Finland);
Teletov Aleksandr, Dr. Sc. (Economics), Professor (Ukraine);
Timofeyev Volodymyr, Dr. Sc. (Engineering), Professor (Ukraine);
Filatov Valentin, Dr. Sc. (Engineering), Professor (Ukraine);
Chumachenko Igor, Dr. Sc. (Engineering), Professor (Ukraine);
Chukhray Nataliya, Dr. Sc. (Economics), Professor (Ukraine);
Yu Zheng, PhD (Physico-Mathematical Sciences), Professor (P.R. of China);
Iastremaska Olena, Dr. Sc. (Economics), Professor (Ukraine).

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>
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).

Затверджений до друку Науково-технічною Радою Харківського національного університету радіоелектроніки (Протокол № 8 від 28 червня 2022 р.).

Свідectво про державну реєстрацію журналу Серія KB № 22696-12596P від 04.05.2017 р.

ЗМІСТ

Інформаційні технології

- 5 **Гобов Д. А.**
Залежності між техніками виявлення вимог: оглядове дослідження в українських компаніях (eng.)
- 16 **Есілевський В. С., Колядін А. В., Серета О. Г.**
Розробка модуля обробки відео для завдання розпізнавання повітряних об'єктів за їх контурами (eng.)
- 26 **Корабльов М. М., Луцький С. В.**
Системно-інформаційні моделі для інтелектуальної обробки інформації (eng.)
- 39 **Соловей О. Л.**
Нова організація процесу вибору ознак за допомогою фільтра на основі кореляції (eng.)
- 51 **Федорович О. Є., Косенко В. В., Лутай Л. М., Замірець Я. О.**
Методи та моделі дослідження інвестиційної привабливості та конкурентоспроможності проектно-орієнтованого підприємства при створенні інноваційних високотехнологічних виробів (eng.)
- 60 **Ювченко К. С., Есілевський В. С., Серета Е. Г.**
Система розпізнавання емоцій людини з використанням алгоритмів глибокого навчання (eng.)

Сучасні технології управління підприємством

- 70 **Андрусевич А. О., Омаров Ш. А., Стародубцев М. Г., Невлюдова В. В.**
Дослідження впливу рівня сучасної роботизації на виклики та ризики економічної безпеки промислового підприємства (eng.)
- 80 **Алфавітний покажчик**

CONTENTS

Information Technology

- 5 **Gobov D.**
Dependencies between requirements elicitation techniques: a survey study in ukrainian companies
- 16 **Yesilevskiy V., Koliadin A., Sereda O.**
Development of a video processing module for the task of air object recognition based on their contours
- 26 **Korablyov M., Lutskiy S.**
System-information models for intelligent information processing
- 39 **Solovei O.**
New organization process of feature selection by filter with correlation-based features selection method
- 51 **Fedorovich O., Kosenko V., Lutai L., Zamirets I.**
Methods and models of research of investment attractiveness and competitiveness of project-oriented enterprise in the process of creating innovative high-tech
- 60 **Yuvchenko K., Yesilevskiy V., Sereda O.**
Human emotion recognition system using deep learning algorithms

Modern Enterprise Management Technologies

- 70 **Andrusevich A., Omarov S., Starodubcev N., Nevliudova V.**
Study of the influence of the modern robotisation level on the challenges and risks of the economic security of an industrial enterprise
- 80 **Alphabetical index**

D. GOBOV

DEPENDENCIES BETWEEN REQUIREMENTS ELICITATION TECHNIQUES: A SURVEY STUDY IN UKRAINIAN COMPANIES

The **subject** of research in the article is requirements elicitation practices in IT projects. The **goal** of the work is to define how project context influences requirement elicitation technique selection and identify dependencies between requirement elicitation techniques. The following **tasks** are solved in the article: examine the industrial standards and experience of business analysts and requirements engineers in requirements acquisition activities, create and conduct a survey on practices in requirement elicitation activities in IT projects, define practitioners' preferences regarding elicitation techniques, and define how project context influences requirement elicitation technique selection, identify dependencies between requirement elicitation techniques. The following **methods** are used: a survey ($N = 324$) was performed among business analysts and requirement engineers in Ukraine regarding their use of requirement elicitation techniques and the context of using them. The Chi-Square test of independence and Cramer's V effect size measure were applied to define statistically significant dependencies between project context and elicitation techniques, as well as dependencies between techniques. The following **results** were obtained: Top elicitation techniques were identified and compared with other comprehensive studies. Twenty statistically significant associations for pairs "project context – elicitation technique" and "elicitation technique – elicitation technique" were found (based on the p-value and Cramer's V effect size measure). **Conclusion:** It is concluded that project context influences particular elicitation technique selection in IT projects. There are also dependencies between requirements elicitation techniques. These dependencies can guide the selection of an initial set of techniques or adjust a set of used elicitation techniques during business analysis planning and monitoring activities.

Keywords: requirements elicitation techniques; IT project; requirements engineering; Chi-Square test; Cramer's V.

Introduction

Business analysis is the practice of providing opportunities for change in the context of an enterprise's work by identifying needs and recommending solutions that bring value to stakeholders [1]. This discipline broadens the requirements engineering scope and application areas [2]. Depending on the project methodology and solution type, there are varying opinions on the set of business analysis tasks. Overall, all business analysis tasks can be grouped into six knowledge areas: Business Analysis Planning, Elicitation, Requirements Life Cycle Management, Strategy Analysis, Requirements Analysis and Design Definition, and Solution Evaluation [3]. Whereas the business analysis laid the groundwork for all future development and testing activities, the elicitation provides the baseline for subsequent requirements analysis, specification and modeling, verification and validation, prioritization, maintenance, monitoring [4], etc. Therefore, failure in the requirements acquisition leads to significant issues with project outcomes. According to [5], 39% of respondents identified errors in the requirements gathering phase as one of the most influential factors that caused software projects' failure. Elicitation is not an isolated act. Information is collected while performing any task involving

interaction with stakeholders and while the business analyst analyzes existing data. Elicitation may trigger additional elicitation for details to fill in gaps or increase understanding. Elicitation activities can be divided into three tasks: prepare for elicitation, conduct elicitation, and confirm elicitation results [6]. During the preparation, the scope of the elicitation should be understood, and an appropriate set of elicitation techniques selected. Choosing the proper techniques and ensuring that each technique is carried out correctly is extremely important for the elicitation activity's success. Best practices and recommendations in the field of elicitation techniques are defined by international standards [5], industrial bodies of knowledge [1], [7], [8], and international empirical studies [9], [10].

Many elicitation techniques have proven themselves in practice and are recommended in the abovementioned sources. Each has advantages and limitations and requires stakeholders' involvement or availability of study materials. As a part of the business analysis approach and business analysis activities plan, a business analyst must decide which techniques best suit a particular project. Usually, multiple methods are used for elicitation. A decision about the set of techniques depends on time and cost constraints, the types of business analysis information sources, their accessibility,

the company's culture, and the desired outcomes [1]. If elicitation is built based on the collaborative approach to the stakeholders' needs, their availability and location must be considered.

This study was conducted to analyze the current preferences of business analysts and requirement engineers regarding selecting elicitation techniques for software development projects. We also wanted to define the attributes of project contexts that influence the probability of choosing a specific elicitation technique. We studied the experience of practicing specialists from Ukrainian and international companies using a questionnaire, experts' judgment, and simple statistical analysis in conjunction with the Chi-Square test.

The paper is the extension and continuation of a paper [11] originally published in the Proceedings of the Federated Conference on Computer Science and Information Systems 2020. Research findings have been expanded by the result of the "technique-technique" pairs analysis and more detailed statistical analysis. The association nature's interpretation was performed for statistically significant dependencies based on the Standardized Pearson Residual values. The Chi-Square based measure of the effect size – Cramer's V – was used to define the strength of the found associations.

This article is organized as follows. Section II includes a review of related works describing elicitation activities and technique selection. Section III is devoted to the survey results, and section VI includes the result of statistical analysis. Section V concludes the paper with a discussion of the findings of our study and future work.

Related works

Most related works focus on analyzing elicitation activities and elicitation techniques in particular. Dieste and Juristo [12] performed a systematic review on requirements elicitation techniques based on 26 empirical studies published until 2005. They aggregated the results in terms of five guidelines for RE practitioners. Wong et al. [13] performed a systematic review on software requirement elicitation activities based on 35 articles. They defined that most contributions were focused on the "Identify Requirements" activity (91%). Still, other activities are poorly covered: "Acquire knowledge" (17%), "Identify sources" (4%), "Defining technique" (9%), "Document" (9%) and "Refine requirements" (4%). Pacheco and Garcia [14] performed a systematic review of stakeholder identification during requirements elicitation based on 47 primary

studies dated 1984 to 2011. They found that identified approaches cannot cover all aspects of stakeholder identification during requirements elicitation. In [15], authors noticed a need to replicate studies in different contexts wherein existing requirement engineers' interventions were evaluated and implemented in practice. It confirms that most case studies involve practitioners as participants; there is a need to work more closely with practitioners. Several studies assess the effectiveness of elicitation techniques in the context of a particular project. Hafsa et al. [16] performed a systematic study on elicitation techniques in a mobile application development project. Based on the analysis of 36 selected articles, 22 requirement gathering methods and eight categories of requirement gathering challenges for mobile applications were identified. In [17] authors defined several factors that can influence elicitation technique selection. This study selected five practitioners as informants from Yemen's companies and government agencies. Dieste and Juristo [18] proposed a framework to help requirements engineers select adequate elicitation techniques. The set of attributes is relevant to the elicitation process's context and influences the selection of one or other techniques that were discovered. Two groups of students were involved in the experiment. Practitioners did not take part in the experiment. The author noticed that their results were not generalizable and should be checked with larger samples. Wong and Mauricio [19] defined a set of factors that influenced each activity of the requirements elicitation process and, consequently, the quality: learning capacity, negotiation capacity, permanent staff, perceived utility, confidence, stress, and semi-autonomous. An empirical study was carried out on 182 respondents from software development companies in Peru. The empirical studies' main restrictions are the limited number of participants and low practitioners' involvement. Last year's dispersed team and outsourcing/outstaffing services model has become a rule rather than an exception. There is a series of studies devoted to the use of machine learning methods to select the optimal set of detection techniques [20], [21], [22]. The main limitation of these studies is a training sample, which may lose its relevance.

The following sources were used for the elicitation technique long list creation: "A Guide to the Business Analysis Body of Knowledge" (BABOK) from the International Institute of Business Analysis (IIBA), "The PMI Guide to Business Analysis" from the Project Management Institute (PMI), a study guide from the

International Requirement Engineering Board (IREB) "Requirements engineering fundamentals" and book "Business Analysis" from British Computer Society (BCS). The analysis of the contents of these sources gives us a set of 13 requirements elicitation: Benchmarking, Brainstorming, Business rules analysis, Collaborative games, Data mining, Document analysis, Interface analysis, Interview, Observation, Process analysis, Prototyping, Questionnaires or Survey, Workshop.

Based on feedback received from business analyst experts during survey review sessions and survey structure from [10], the following techniques were added:

- Design thinking / Lean startup [23].
- Reuse database and guidelines (as a subset of document analysis).
- Stakeholders list, map, or personas.

Survey study

A. Research questions

The primary research objective was to identify factors that influence elicitation techniques in IT projects. The first research question (RQ) makes the research objective explicit.

RQ1: Which elicitation techniques do practitioners use in their IT projects?

The answer of this RQ provides information about what specific elicitation techniques are used by practitioners. It allows to compare technique popularity in comparison with other researches result.

RQ2: Does the context of the IT project influence the choice of requirements elicitation techniques?

This RQ is aimed at defining factors that have to be taken into account during forming the start set of elicitation techniques in IT projects.

RQ3: Are there dependencies between elicitation techniques in IT projects?

The answer of this RQ allows us to form recommendation regarding tailoring a set of elicitation techniques for projects where start set of techniques has been already formed.

Questionnaire Design

The literature review has shown that many kinds of research have been conducted to identify common patterns and problems in IT business analysis and requirements elicitation. However, after studying the

existing questionnaires developed for international surveys, we realized the necessity of adjusting them to Ukrainian IT companies' specifics. It was decided to take the questions' basis from the NaPIRE initiative [9], [10] and rework it concerning mentioned above sources such as [1], [6], [7], [8]. Survey items were carefully written using the business analysis vocabulary, mostly from BABOK. The questionnaire's types of questions are open-ended, closed-ended (multiple and single choices), and Likert scale. Details regarding survey structure and target group are provided in [11]. The overall number of survey participants is 328. Four participants were filtered out because they were not involved in elicitation activities and were not in this research's target group. English and Ukrainian languages were used for questionnaires. The questionnaire was created using Google forms and links to it. It was shared in the local Business Analysis communities, professional and social networks, and via personal contacts in TOP 10 Ukrainian IT companies. The answers were collected in one month. After that, data were merged and coded for further analysis. The dataset of the survey results is published in the Mendeley Data repository [24].

B. Survey Results.

The most used elicitation techniques are shown in fig. 1. Participants were allowed to select multiple techniques.

Regardless of the context in which the Ukrainian business analyst is working, we may see the following most popular elicitation techniques:

- Interview (87.35%)
- Document analysis (85.49%)
- Interface analysis (71.3%)
- Brainstorming (69.44%)
- Prototyping (66.36%)
- Process analysis/Process modeling (66.36%)

The rare techniques are Collaborative games (1.25%) and Design thinking (13.89%).

In the table 1 we compared our results with the NaPIRE study, namely, their top 5 elicitation techniques versus ours. In general, the list of techniques in [8] was adopted from SWEBoK [25]. The results of NaPIRE surveys indicate that the most frequently used techniques are interviews (167 respondents or 73% of the respondents) and facilitated meetings (153 / 67 %), closely followed by prototyping (132 / 58 %) and scenarios (93 / 41 %). Observations were only quoted 62 times (29%). Additional answers for "others" included "Created

personas and presented them to our stakeholders", "Questionnaires" / "Surveys", "Analysis of existing system" and "It depends on the client" [12].

Our survey results demonstrate a slight difference from the NaPiRE study regarding the popularity of elicitation techniques. The sample size and uniformity

could explain it, i.e., the number of respondents in our study is more significant. It is worth mentioning that the list of proposed elicitation techniques in our study is not limited to SWEBoK only but refers to international standards [6] and industrial bodies of knowledge [1], [7], [8].

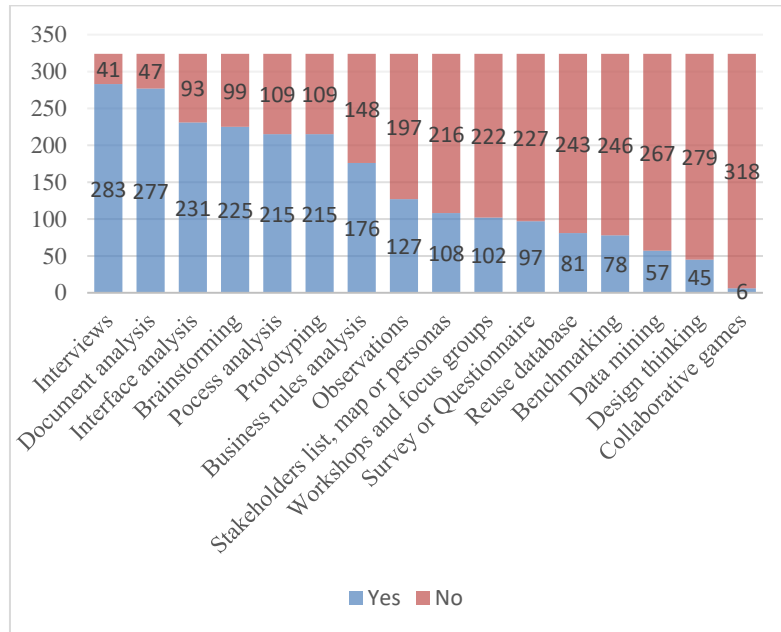


Fig. 1. Elicitation techniques popularity

Table 1. Survey-based comparison of top elicitation techniques

Elicitation technique	Gobov, Huchenko (2020)	Popularity	Wagner et al. (2019)
Interview	283 respondents or 87.35%	1	167 respondents or 73%
Document analysis	277 respondents or 85.49%	2	–
Interface analysis	231 respondents or 71.3%	3	–
Brainstorming	225 respondents or 69.44%	4	–
Prototyping	215 respondents or 66.36%	5	132 respondents or 58%
Workshop and focus group	102 respondents or 31.48%	10	153 respondents or 67%
Scenarios	–	–	93 respondents or 41%
Observation	127 respondents or 39.19%	8	62 respondents or 29%

Survey Results. Elicitation Techniques Usage Analysis with Chi-Square and Cramer's V

The questionnaire results analysis checked each "background factor-elicitation technique" and "technique-technique" pair for the association. The Chi-Square test of independence, commonly used for testing relationships between categorical variables, was applied to examine the differences within a single dependent sample (population). A set of hypotheses about the

association between techniques was developed. An example of the null and alternative hypothesis is:

H0: There is no association between technique A and elicitation technique B usage.

H1: There is an association between technique A and elicitation technique B usage.

After calculating P-Value, which should be less than 0.05 considering a 0.95 confidence level, the conclusion about statistical significance was made for the technique – technique pairs.

While the Chi-Square test is advantageous for testing a relationship, it has several weak points. One of the difficulties with the test is that it does not indicate the nature of the relationship. It is impossible to determine how one variable changes as the values of the other variable change. The only way to do this is to carefully assess the table to ascertain the relationship between the two variables. Standardized Pearson Residual (further SPR) was used to identify those specific cells that contributed most significantly to the Chi-square test results. According to [26], a cell-by-cell comparison of observed and estimated expected frequencies is used to assess the evidence's nature. SPR having an absolute value that exceeds ± 2 when there are few cells or ± 3 when there are many cells indicates a lack of fit of H_0 in that cell:

- If the residual is less than -2 or -3 , respectively, the observed frequency is less than the expected frequency.
- If the residual is greater than 2 or 3 , respectively, the observed frequency is greater than the expected frequency.

Considering mentioned above, SPRs were used to interpret the identified dependencies for project factor – elicitation technique pairs defined in [11] and elicitation technique – elicitation technique pairs.

The second issue with the Chi-Square independence test is that the chi-square statistics' value may vary based on the number of cells in the table. It may be misleading to compare the chi-square statistics for two tables of entirely different dimensions (i.e., different numbers of rows and columns in the table). Cramer's V – Chi-square based association measure – was used to adjust the Chi-Square test results and consider differences in table size. Different sources give a different interpretations of Cramer's V value [27], [28].

As we used the IBM SPSS tool [28] for analysis, we refer to their stricter definition of effect size, namely:

$V \leq 0.2$ – weak association.

$0.2 < V \leq 0.6$ – moderate association.

$V > 0.6$ – strong association.

SPR and Cramer's V were also used to adjust the Chi-Square test results for associations between project factors and techniques usage defined in [11].

The associations' analysis is presented below but only for some techniques where the effect size is at least more than 0.25.

C. Benchmarking

Twelve statistically significant associations have been found for benchmarking. Three of them are "project factor-technique", and nine are "technique-technique".

- "Project Category – Benchmarking": p -value=0.011, Cramer's $V = 0.186$. Benchmarking is used more frequently in greenfield engineering projects (SPR=2.9) and less frequently in product/platform customization projects (SPR=–2.4). SPRs for user interface engineering and reengineering projects are 0.6 and –1.4, respectively.

- "Experiment – Benchmarking": p -value=0, Cramer's $V = 0.209$. Benchmarking is more frequently used if experiment elicitation sources are used (SPR=3.8) and less frequently used in the contrary case (SPR = –3.8).

- "Research – Benchmarking": p -value=0, Cramer's $V = 0.274$. Benchmarking is more frequently used if research elicitation sources are used (SPR=4.9) and less frequently used in the contrary case (SPR= –4.9).

- Benchmarking is more frequently used if the following techniques are used as well: Brainstorming (SPR=2.2), Business rule analysis (SPR=2.3), Interface analysis (SPR=2.4), Observations (SPR=4.1), Process analysis (SPR=3.1), Prototyping (SPR=3.9), Stakeholders list, map, or Personas (SPR=4.7), Survey (SPR=4.7), Workshop (SPR=2.9).

D. Business rules analysis

Thirteen statistically significant associations have been found for Business rule analysis. Four of them are "project factor-technique", and nine are "technique-technique".

- "Company size – Business rule analysis": p -value=0.014, Cramer's $V = 0.162$. Brainstorming is more frequently used in companies with over 1500 specialists (SPR=2.4) and less frequently used in companies with up to 200 specialists (SPR= –2.8). SPR for a company with 201–1500 specialists is 0.2.

- "Project Category – Business rule analysis": p -value=0.024, Cramer's $V = 0.171$. Business rule analysis is more frequently used in greenfield engineering projects (SPR=1.7) and less frequently used in product/platform customization projects (SPR= –3). SPRs for user interface engineering and reengineering projects are 0.7 and 0.4, respectively.

- "Experiment – Business rule analysis": p -value=0.001, Cramer's $V = 0.177$. Business rule analysis is more frequently used if experiment elicitation sources are used (SPR=3.2) and less frequently used in the contrary case (SPR = –3.2).

- "Research – Business rule analysis": p -value=0, Cramer's $V = 0.246$. Business rule analysis is more frequently used if research elicitation sources are used (SPR=4.4) and less frequently used in the contrary case (SPR= –4.4).

- Business rule analysis is more frequently used if the following techniques are used as well: Benchmarking and Market analysis (SPR=2.3), Data mining (SPR=2.1), Document analysis (SPR=2.1), Interface analysis (SPR=2.8), Observations (SPR=2.3), Process analysis (SPR=4.5), Prototyping (SPR=2.4), Stakeholders list, map, and Personas (SPR=3.6), Workshop (SPR=5.4).

E. Collaboration games

Two statistically significant associations have been found:

- "Industrial Sector – Collaboration games": p-value=0.001, Cramer's $V = 0.404$.

- "Design Thinking – Collaboration games": p-value=0.01, Cramer's $V = 0.143$.

Unfortunately, too many cells (70% and 25%, respectively) have an expected count of less than 5.

F. Design thinking

Ten statistically significant associations have been found for Design thinking. Two of them are "project factor-technique", and eight are "technique-technique".

- "Way of working – Design thinking": p-value=0.001, Cramer's $V = 0.201$. Design thinking is more frequently used in agile (SPR=3.5) and less frequently used in hybrid projects (SPR= -3.2). SPR for a plan-driven project is 0.2.

- "Experiment – Design thinking": p-value=0.007, Cramer's $V = 0.149$. Design thinking is more frequently used if experiment elicitation sources are used (SPR=2.7) and less frequently used in the contrary case (SPR = -2.7).

- Design thinking is more frequently used if the following techniques are used as well: Brainstorming (SPR=3.1), Collaboration games (SPR=2.4), Interface analysis (SPR=2.1), Prototyping (SPR=2.8), Reuse database (SPR=2.5), Stakeholders list, map, and Personas (SPR=4.8), Survey (SPR=2.6), Workshop (SPR=3.4).

G. Document analysis

Nine statistically significant associations have been found for Document analysis. Two of them are "project factor-technique", and seven are "technique-technique".

- "Industrial sector – Document analysis": p-value=0.006, Cramer's $V = 0.372$. Unfortunately, 64% of cells have an expected count of less than 5.

- "Research – Document analysis": p-value=0, Cramer's $V = 0.241$. Document analysis is more frequently used if experiment elicitation sources are used (SPR=4.3) and less frequently used in the contrary case (SPR = -4.3).

- Document analysis is more frequently used if the following techniques are used as well: Business rules analysis (SPR = 2.1), Interface analysis (SPR = 4.4), Interviews (SPR = 3.3), Process analysis (SPR = 4.1), Prototyping (SPR = 2.1), Reuse database (SPR = 3.2), Workshop (SPR = 2.6).

H. Interview

Eleven statistically significant associations have been found for the Interview. Two of them are "project factor-technique", and nine are "technique-technique".

- "System service class – Interview": p-value=0.015, Cramer's $V = 0.18$. Unfortunately, 50% of cells have an expected count of less than 5.

- "Experiment – Interview": p-value=0, Cramer's $V = 0.182$.

- The interview is more frequently used by specialists with experience between 5 and 10 years (SPR=2.1) and less frequently used by specialists with experience up to 3 years (SPR = -2.2).

- Interview is more frequently used if the following techniques are used as well: Brainstorming (SPR=2.7), Document analysis (SPR=3.3), Observations (SPR=2.8), Process analysis (SPR=3.3), Prototyping (SPR=4.7), Reuse database (SPR=2), Stakeholders list, map, and Personas (SPR=2.7), Survey (SPR=2.7), Workshop (SPR=2.1).

I. Process analysis

Sixteen statistically significant associations have been found for Process analysis. Six of them are "project factor-technique", and ten are "technique-technique".

- "Company type – Process analysis": p-value=0.005, Cramer's $V = 0.198$. Process analysis is more frequently used in Non-IT, In-house development (SPR=3.1) and less frequently in outstaff companies (SPR = -2.3). SPRs for IT Outsource and IT product companies are -0.6 and 0, respectively.

- "Experience – Process analysis": p-value=0.014, Cramer's $V = 0.181$. Process analysis is less frequently used by specialists with experience up to 3 years (SPR= -3.1). SPRs for 3–5 years, 5–10 years, and over ten years are 1.8, 1.6, and 0.1, respectively.

- "Project Category – Process analysis": p-value=0.001, Cramer's V = 0.227. Process analysis is more frequently used in greenfield engineering projects (SPR=3.7) and less frequently used in product/platform customization projects (SPR= -2.4) and user interface engineering projects (SPR= -2.3). SPR for reengineering projects is -0.6.

- "Experiment – Process analysis": p-value=0, Cramer's V = 0.202. Process analysis is more frequently used if experiment elicitation sources are used (SPR=3.6) and less frequently used in the contrary case (SPR = -3.6).

- "Research – Process analysis": p-value=0, Cramer's V = 0.256. Process analysis is more frequently used if research elicitation sources are used (SPR=4.6) and less frequently used in the contrary case (SPR= -4.6).

- "Elicitation responsibility – Process analysis": p-value=0.013, Cramer's V = 0.164. Process analysis is more frequently used if the business analyst/requirement engineer has the primary responsibility for the solution requirements (FRs/NFRs) elicitation (SPR=2.6) and less frequently used if nor business analyst/requirement engineer nor Product Owner/Business analyst have the primary responsibility (SPR= -2.3). SPR for the case "Product owner/Business analyst has primary responsibility for the solution requirements (FRs/NFRs) elicitation" = -0.9.

- Process analysis is more frequently used if the following techniques are used as well: Benchmarking and Market analysis (SPR=3.1), Business rules analysis (SPR=4.5), Document analysis (SPR=4.1), Interface analysis (SPR=2.8), Interviews (SPR=3.3), Observations (SPR=3.3), Prototyping (SPR=2.1), Reuse database (SPR=3.1), Stakeholders list, map, and Personas (SPR=2.8), Survey (SPR=3).

J. Prototyping

Sixteen statistically significant associations have been found for Prototyping. Four of them are "project factor-technique", and twelve are "technique-technique".

- "Experience – Prototyping": p-value=0.005, Cramer's V = 0.2. Prototyping is less frequently used by specialists with experience up to 3 years (SPR= -3.6). SPRs for 3–5 years, 5–10 years, and over 10 years are 1.6, 1.3, and 1.3, respectively.

- "Experiment – Prototyping": p-value=0, Cramer's V = 0.294. Prototyping is more frequently used if experiment elicitation sources are used (SPR=2.2) and less frequently used in the contrary case (SPR = -2.2).

- "Research – Prototyping": p-value=0.03, Cramer's V = 0.121. Prototyping is more frequently if research elicitation sources are used (SPR=2.2) and less frequently used in the contrary case (SPR= -2.2).

- "Elicitation responsibility – Prototyping": p-value=0.008, Cramer's V = 0.171. Prototyping is more frequently used if the business analyst/requirement engineer has the primary responsibility for the solution requirements (FRs/NFRs) elicitation (SPR=2.6) and less frequently used if neither the business analyst/requirement engineer nor Product Owner/Business analyst have the primary responsibility (SPR= -2.6). SPR for the case "Product owner/Business analyst has primary responsibility for the solution requirements (FRs/NFRs) elicitation" = -0.6.

- Prototyping is more frequently used if the following techniques are used as well: Benchmarking and Market analysis (SPR=3.9), Brainstorming (SPR=3.5), Business rules analysis (SPR=2.4), Design thinking (SPR=2.8), Document analysis (SPR=2.8), Interface analysis (SPR=3.6), Interviews (SPR=4.7), Observations (SPR=3.1), Process analysis (SPR=2.1), Stakeholders list, map, and Personas (SPR=4.1), Survey (SPR=2), Workshop (SPR=4.1).

K. Stakeholders list, map, and Personas

Sixteen statistically significant associations have been found for the Stakeholders list, map, and personas. Four of them are "project factor-technique", and twelve are "technique-technique".

- "Company size – Stakeholders list, map, and personas": p-value=0.017, Cramer's V = 0.158. Techniques are more frequently used in companies with over 1500 specialists (SPR=2.8) and less frequently in companies with up to 200 specialists (SPR= -2.1). SPR for a company with 201–1500 specialists is -1.2.

- "Team distribution – Stakeholders list, map, and personas": p-value=0.019, Cramer's V = 0.131. Techniques are more frequently used in distributed teams (SPR=2.4) and less frequently used in co-located teams (SPR= -2.4).

- "Experiment – Stakeholders list, map, and personas": p-value=0.005, Cramer's V = 0.157. Techniques are more frequently used if experiment elicitation sources are used (SPR=2.8) and less frequently in the contrary case (SPR = -2.8).

- "Research – Stakeholders list, map, and personas": p-value=0, Cramer's V = 0.212. Stakeholders list, map, and personas are more frequently used

if research elicitation sources are used (SPR=2.2) and less frequently used in the contrary case (SPR= -2.2).

- Stakeholders list, map, and Personas are more frequently used if the following techniques are used as well: Benchmarking and Market analysis (SPR=4.7), Brainstorming (SPR=2.3), Business rule analysis (SPR=3.6), Data Mining (SPR=3.1), Design Thinking (SPR=3.6), Interviews (SPR=2.7), Observations (SPR=3.1), Process Analysis (SPR=2.8), Prototyping (SPR=4.1), Reuse Database (SPR=3).

L. Survey

Thirteen statistically significant associations have been found for Survey. Four of them are «project factor-technique», and nine are «technique-technique».

- "Project Category – Survey": p-value=0.015, Cramer's V = 0.179. A survey is more frequently used in greenfield engineering projects (SPR=2.6). SPRs for user interface engineering projects, reengineering projects, and product/platform customization projects are -1.5, -0.6, and -1.6, respectively.

- "Experiment – Survey": p-value=0.024, Cramer's V = 0.125. The survey is more frequently used if experiment elicitation sources are used (SPR=2.3) and less frequently in the contrary case (SPR = -2.3).

- "Research – Survey": p-value=0.001, Cramer's V = 0.188. The survey is more frequently used if research elicitation sources are used (SPR=3.4) and less frequently used in the contrary case (SPR= -3.4).

- Survey is more frequently used if the following techniques are used as well: Benchmarking and Market analysis (SPR=4.7), Brainstorming (SPR=2), Design thinking (SPR=2.6), Interviews (SPR=2.7), Observations (SPR=3.2), Process analysis (SPR=3), Prototyping (SPR=2), Stakeholders list, map, and Personas (SPR=3), Workshop (SPR=2.7), Reuse Database (SPR=3).

M. Workshop and Focus group

Fifteen statistically significant associations have been found for the Workshop and Focus group. Five of them are "project factor-technique", and ten are "technique-technique".

- "Company size – Workshop and focus group": p-value=0.005, Cramer's V = 0.181. Workshop and focus group is more frequently used in companies with over 1500 specialists (SPR=3.1) and less frequently used in companies with up to 200 specialists (SPR= -2.7). SPR for a company with 201–1500 specialists is -0.9.

- "Team distribution – Workshop and focus group": p-value=0.019, Cramer's V = 0.156. Workshop and focus group are more frequently used in distributed teams (SPR=2.8) and less frequently in co-located teams (SPR= -2.8).

- "Experience – Workshop and focus group": p-value=0, Cramer's V = 0.265. Workshop and focus group are more frequently used by specialists with experience over 10 years (SPR=2.9) and less frequently used by specialists with experience up to 3 years (SPR= -4.4). SPRs for 3–5 years and 5–10 years are 0.6 and 1.9, respectively.

- "Experiment – Workshop and focus group": p-value=0.004, Cramer's V = 0.158. Workshop and focus group are more frequently used if experiment elicitation sources are used (SPR=2.8) and less frequently used in the contrary case (SPR = -2.8).

- "Research – Workshop and focus group": p-value=0.038, Cramer's V = 0.115. Workshop and Focus groups are more frequently used if research elicitation sources are used (SPR=2.1) and less frequently used in the contrary case (SPR= -2.1).

- Workshop and focus group are more frequently used if the following techniques are used as well: Benchmarking and Market analysis (SPR=2.9), Brainstorming (SPR=3.9), Business rules analysis (SPR=5.4), Design thinking (SPR=3.4), Document analysis (SPR=2.6), Interface analysis (SPR=2.2), Interviews (SPR=2.1), Prototyping (SPR=4.1), Stakeholders list, map, and Personas (SPR=5.6), Survey (SPR=2.7).

Conclusion

A survey study has been undertaken to analyze the current state and requirements elicitation techniques in different software project contexts. The survey structure was built based on the worldwide known industrial standards. Attributes of project context were established to analyze their influence on the requirement elicitation techniques selection. The survey was conducted among practitioners from Ukrainian IT and non-IT companies, 328 specialists (mainly business analysts and product owners) took part in it. The most used elicitation techniques were identified and compared with top techniques from the NaPiRE study [10]. Further analysis was conducted based on the Chi-Square test of independence to examine the "project context-elicitation technique" dependencies and possible "technique-technique" associations. One hundred seventy-nine statistically significant associations were found (based

on the p-value). This result was adjusted using the SPR value for the dependency interpretation and Cramer’s V effect size measure to define the strength of association.

The pairs for which moderate association has been found are shown in fig. 2 in yellow ($0.25 \leq V < 0.3$) and green ($0.3 \leq V < 0.6$) colors. No strong associations have been found.

These dependencies can guide the selection of supportive techniques or adjust a set of core elicitation techniques.

Our study had several limitations. The list of techniques included in the survey is not exhaustive, and elicitation techniques may be applied alternatively or in conjunction with other techniques. Due to specific project context, business analysts are encouraged to modify techniques or create new ones. The survey

result gathering was done via a google survey engine and was intended to be anonymous (requiring personal data is problematic on legal and ethical grounds). Therefore, we cannot prove that respondents provided accurate information about the project context and used elicitation techniques. Considering that the survey was limited to one country only, its results cannot be extrapolated to the worldwide software industry (even though the IT industry in Ukraine is integrated into international environments, especially outsourcing and outstaffing companies, whose employees were the majority of respondents (65%). Several directions for future research can be considered. Other business analysis’ tasks can be analyzed to define dependencies and recommendations regarding selection techniques for requirement specification and modeling, validation, and verification.

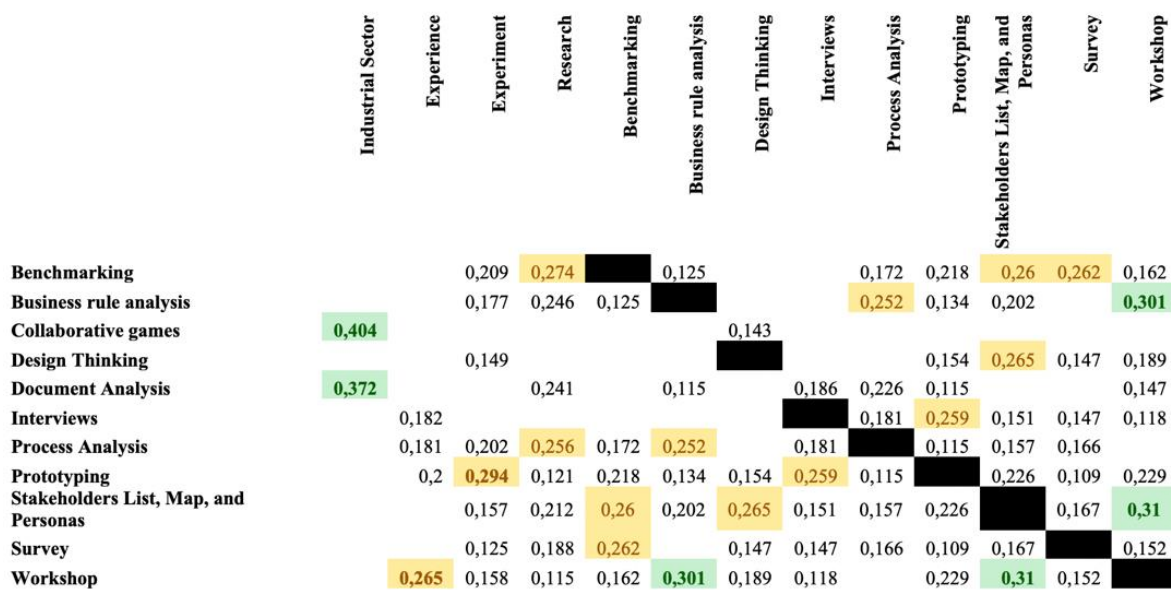


Fig. 2. Strength of association (Cramer’s V)

References

1. International Institute of Business Analysis (2015), *A guide to the business analysis body of knowledge BABOK Guide*, Ver. 3, IIBA, Toronto, Ontario, Canada, 514 p.
2. Rubens, J. (2007), "Business analysis and requirements engineering: the same, only different?", *Requirements Engineering*, Vol. 12 (2), P. 121–123. DOI: <http://dx.doi.org/10.1007/s00766-007-0043-3>
3. Gobov, D., Maliarcuk, C., Kunanets, N., Oliinyk, Y. (2020), "Approaches for the Concept "Business Analysis" Definition in IT Projects and Frameworks", *CEUR Workshop Proceedings*, 2711, P. 321–332.
4. Martynenko, O., Husieva, Y., Chumachenko, I. (2017), "The method of earned requirements for project monitoring", *Innovative technologies and scientific solutions for industries*, Vol. 1 (1), P. 58–63.
5. Clancy, T. (2014), *The chaos report*, The Standish Group.
6. ISO/IEC/IEEE 29148. (2011), *Systems and Software Engineering—Life Cycle Processes—Requirements Engineering*, ISO/IEC/IEE, Standard 29148-2011.
7. Project Management Institute (2017), *The PMI Guide to BUSINESS ANALYSIS*. PMI, Newtown Square, Pennsylvania, 488 p.
8. Pohl, K. (2010), *Requirements engineering: fundamentals, principles, and techniques*, Springer Publishing Company, 182 p.

9. Wagner, S., et al. (2019), "Status Quo in Requirements Engineering: A Theory and a Global Family of Surveys", *ACM Trans. Softw. Eng. Methodol.*, No. 28 (2), Article 9. DOI: <http://dx.doi.org/10.1145/3306607>
10. Fernández, D. M., Wagner, S. (2015), "Naming the pain in requirements engineering: A design for a global family of surveys and first results from Germany", *Information and Software Technology*, Vol. 57, P. 616–643. DOI: <http://dx.doi.org/10.1016/j.infsof.2014.05.008>
11. Gobov, D., Huchenko, I. (2020), "Requirement Elicitation Techniques for Software Projects in Ukrainian IT: An Exploratory Study", *Proc. of the Federated Conference on Computer Science and Information Systems*, P. 673–681.
12. Dieste, O., Juristo, N. (2011), "Systematic review and aggregation of empirical studies on elicitation techniques", *IEEE Transactions on Software Engineering*, No. 37 (2), P. 283–304. DOI: <http://dx.doi.org/10.1109/TSE.2010.33>
13. Wong, L. R., Mauricio, D. S., Rodriguez, G. D. (2017), "A systematic literature review about software requirements elicitation", *Journal of Engineering Science and Technology*, No. 12 (2), P. 296–317.
14. Pacheco, C., Garcia, I. (2012), "A systematic literature review of stakeholder identification methods in requirements elicitation", *Journal of Systems and Software*, No. 85 (9), P. 2171–2181. DOI: <http://dx.doi.org/10.1016/j.jss.2012.04.075>
15. Ambreen, T., Ikram, N., Usman, M., Niazi, M. (2018), "Empirical research in requirements engineering: trends and opportunities", *Requirements Engineering*, Vol. 23 (1), P. 63–95. DOI: <http://dx.doi.org/10.1007/s00766-016-0258-2>
16. Dar, H., Lali, M. I., Ashraf, H., Ramzan, M., Amjad, T., Shahzad, B. (2018), "A systematic study on software requirements elicitation techniques and its challenges in mobile application development", *IEEE Access*, No. 6, P. 63859–63867. DOI: <https://dx.doi.org/10.1109/ACCESS.2018.2874981>
17. Anwar, F., Razali, R. (2012), "A practical guide to requirements elicitation techniques selection-an empirical study", *Middle-East Journal of Scientific Research*, No. 11 (8), P. 1059–1067.
18. Carrizo, D., Dieste, O., Juristo, N. (2014), "Systematizing requirements elicitation technique selection", *Information and Software Technology*, Vol. 56 (6), P. 644 – 669. DOI: <http://dx.doi.org/10.1016/j.infsof.2014.01.009>
19. Wong, L., Mauricio, D. (2018), "New factors that affect the activities of the requirements elicitation process", *Journal of Engineering Science and Technology*, No. 13 (7), P. 1992–2015.
20. Gobov, D., Huchenko, I. (2021), "Influence of the Software Development Project Context on the Requirements Elicitation Techniques Selection", *In Advances in Computer Science for Engineering and Education IV. ICCSEEA, Lecture Notes on Data Engineering and Communications Technologies*, Vol. 83, Springer, Cham. DOI: http://dx.doi.org/10.1007/978-3-030-80472-5_18
21. Darwish, N., Mohamed, A., Abdelghany, A. (2016), "A hybrid machine learning model for selecting suitable requirements elicitation techniques", *International Journal of Computer Science and Information Security*, Vol. 14 (6), P. 1–12.
22. Gobov, D., Huchenko, I. (2021), "Software requirements elicitation techniques selection method for the project scope management", *CEUR Workshop Proceedings*, 2851, P. 1–10.
23. Martins, F., et al. (2019), "Design thinking: Challenges for software requirements elicitation", *Information*, No. 10 (12), P. 371. DOI: <https://doi.org/10.3390/info10120371>
24. Gobov, D. (2021), "Project context and Elicitation Techniques. Ukrainian IT companies. Survey result 2020", available at: <https://data.mendeley.com/drafts/svzv7rs279> (last access 15.08.2022)
25. Fairley, P., Keppler, J. (2014), *Guide to the software engineering body of knowledge (SWEBOK): Version 3.0.*, IEEE Computer Society Press, Washington, DC, USA, 335 p.
26. Agresti, A. (2022), *Categorical Data Analysis (2nd Ed.)*, New York, Wiley, 734 p.
27. Rea, L. M., Parker, R. A. (2014), *Designing and conducting survey research: A comprehensive guide*, John Wiley & Sons, 352 p.
28. IBM statistical terms, available at: https://www.ibm.com/support/knowledgecenter/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_ca_dshb.doc/cramersv.html (last access 15.08.2022)

Received 16.08.2022

Відомості про авторів / Сведения об авторах / About the Authors

Гобов Денис Андрійович – кандидат технічних наук, Національний технічний університет України "Київський політехнічний інститут ім. Ігоря Сікорського", доцент кафедри інформатики та програмної інженерії факультету інформатики та обчислювальної техніки, Київ, Україна; e-mail: d.gobov@kpi.ua; ORCID ID: <https://orcid.org/0000-0001-9964-0339>

Гобов Денис Андреевич – кандидат технических наук, Национальный технический университет Украины "Киевский политехнический институт им. Игоря Сикорского", доцент кафедры информатики и программной инженерии факультета информатики и вычислительной техники, Киев, Украина.

Gobov Denys – PhD (Engineering), National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Senior lecturer of the Department of Computer Science and Software Engineering of the Faculty of Informatics and Computer Science, Kyiv, Ukraine.

ЗАЛЕЖНОСТІ МІЖ ТЕХНІКАМИ ВИЯВЛЕННЯ ВИМОГ: ОГЛЯДОВЕ ДОСЛІДЖЕННЯ В УКРАЇНСЬКИХ КОМПАНІЯХ

Предметом дослідження в статті є практики виявлення вимог у IT-проектах. **Мета роботи** – визначити, як контекст проекту впливає на вибір технік виявлення вимог, а також визначити залежності між техніками виявлення вимог. У роботі вирішуються такі **завдання**: вивчення галузевих стандартів і досвіду бізнес-аналітиків та інженерів з вимог у діяльності щодо визначення вимог у IT-проектах; створення та проведення опитування з практик виявлення вимог у IT-проектах; визначення вподобання фахівців-практиків щодо технік виявлення та з'ясування, як контекст проекту впливає на вибір технік виявлення вимог; визначення залежності між техніками виявлення вимог. Використовувались такі **методи**: опитування ($N = 324$) бізнес-аналітиків та інженерів з вимог, що працюють в Україні, щодо використання ними технік виявлення вимог та контексту їхнього застосування. Для визначення статистично значущих залежностей між контекстом проекту та техніками виявлення, а також залежностей між техніками виявлення застосовувались критерій незалежності χ^2 -квадрат та критерій розміру V -ефекту Крамера. Були отримані такі **результати**: визначено найпопулярніші техніки виявлення вимог, проведено порівняння цих результатів з іншими всебічними дослідженнями; виявлено 20 статистично значущих асоціацій для пар "контекст проекту – техніка виявлення" та "техніка виявлення – техніка виявлення" (на основі p -значення та міри розміру V -ефекту Крамера). **Висновок**: для низки технік виявлення вимог контекст проекту істотно впливає на їхнє використання для отримання інформації в процесі IT-проектів. Також існують статистично значущі залежності між техніками виявлення вимог. Ці залежності можуть бути застосовані для формування початкового набору технік виявлення або зміни набору використовуваних технік виявлення під час планування й моніторингу ефективності робіт з бізнес-аналізу.

Ключові слова: техніки виявлення вимог; IT-проекти; інженерія вимог; критерій χ^2 -квадрат; V Крамера.

ЗАВИСИМОСТІ МЕЖДУ ТЕХНИКАМИ ВЫЯВЛЕНИЯ ТРЕБОВАНИЙ: ОБЗОРНОЕ ИССЛЕДОВАНИЕ В УКРАИНСКИХ КОМПАНИЯХ

Предметом исследования в статье являются практики выявления требований в IT-проектах. **Цель работы** – определить, как контекст проекта влияет на выбор техник выявления требований, и обнаружить зависимости между техниками выявления требований. В работе решаются следующие **задачи**: изучить отраслевые стандарты и опыт бизнес-аналитиков и инженеров по требованиям в деятельности по сбору требований в IT-проектах; создать и провести опрос по практикам выявления требований в IT-проектах; определить предпочтения практикующих специалистов в отношении техник выявления и определить, как контекст проекта влияет на выбор техники выявления требований; определить зависимости между техниками выявления требований. Используются следующие **методы**: опрос ($N = 324$) бизнес-аналитиков и инженеров по требованиям в Украине относительно использования ими техник выявления требований и контекста их использования. Для определения статистически значимых зависимостей между контекстом проекта и техниками выявления, а также зависимостей между техниками применялись критерий независимости χ^2 -квадрат и критерий размера V -эффекта Крамера. Были получены следующие **результаты**: определены наиболее популярные техники выявления требований, проведено сравнение с другими всесторонними исследованиями; обнаружено 20 статистически значимых ассоциаций для пар "контекст проекта – техника выявления" и "техника выявления – техника выявления" (на основе p -значения и меры размера V -эффекта Крамера). **Вывод**: для ряда техник выявления требований контекст проекта существенно влияет на их использование для извлечения информации в ходе IT-проектов. Также существуют статистически значимые зависимости между техниками выявления требований. Эти зависимости могут быть использованы для формирования начального набора техник выявления или изменения набора используемых техник выявления в ходе планирования и мониторинга эффективности работ по бизнес-анализу.

Ключевые слова: техники выявления требований; IT-проекты; инженерия требований; критерий χ^2 -квадрат; V Крамера.

Бібліографічні описи / Bibliographic descriptions

Гобов Д. А. Залежності між техніками виявлення вимог: оглядове дослідження в українських компаніях. *Сучасний стан наукових досліджень та технологій в промисловості*. 2022. № 3 (21). С. 5–15. DOI: <https://doi.org/10.30837/ITSSI.2022.21.005>

Gobov, D. (2022), "Dependencies between requirements elicitation techniques: a survey study in ukrainian companies", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (21), P. 5–15. DOI: <https://doi.org/10.30837/ITSSI.2022.21.005>

V. YESILEVSKYI, A. KOLIADIN, O. SEREDA

DEVELOPMENT OF A VIDEO PROCESSING MODULE FOR THE TASK OF AIR OBJECT RECOGNITION BASED ON THEIR CONTOURS

The subject of research in the article is the module of automatic segmentation and subtraction of the background, which is created, based on the sequential application of methods of image preprocessing and modified method of interactive segmentation of images and implemented in the system of optical monitoring of the air situation. **The aim** of the work is to develop an image segmentation module to increase the efficiency of recognition of an air object type on a video image in the system of visual monitoring of the air environment by means of qualitative automatic segmentation. To solve this problem, a modified interactive algorithm in the mode of automatic selection of an object in the image, which allows more accurately, without the participation of the operator, to determine the foreground pixels of the image for further recognition of the type of airborne object. The following **tasks** are solved in the article: the analysis of existing methods of binarization of color images for semantic segmentation of images, which are used in image recognition systems; the development of a pipeline of methods for automatic segmentation of images in the system of optical monitoring of the air environment. In the work, the following methods are used: methods of digital image processing, methods of filtering and semantic segmentation of images, methods of graph analysis. The following **results** are obtained: the results of image processing with the proposed module of segmentation and background subtraction confirm the performance of the module procedures. The developed pipeline of methods included in the module demonstrates correct segmentation in 93% of test images in automatic mode without operator participation, which allows us to conclude about the effectiveness of the proposed module. **Conclusions:** The implementation of the developed module of segmentation and background subtraction for the system of optical monitoring of the air environment allowed to solve the problem of segmentation of video images for further recognition of aerial objects in the system of optical monitoring of the air environment in automatic mode with a high degree of reliability, thus increasing the operational efficiency of this system.

Keywords: image segmentation; background subtraction; recognition of air objects; optical monitoring of the air situation.

Introduction

The task of recognition of air objects (AO) as a part of the task of monitoring the air situation is of great importance in the combat situation during military operations. Its solution can also be used in peacetime to monitor the air situation at airports and protected facilities. In addition to (or even instead of) classical radar systems, optical video surveillance systems are increasingly being used for monitoring the air situation. This approach makes it possible to increase the mobility of AO detection systems and to overcome some known problems of radar monitoring associated with masking the characteristics of air objects in the radar detection area.

Information about the type of aerial threat is necessary for a correct assessment of the air situation and for making an operational decision. Currently, the list of types of detected AO has significantly expanded due to the use of artillery shells, various types of missiles, unmanned aerial vehicles (UAVs), quadcopters, helicopters. This has fundamentally changed the range of detectable AO parameters, from shape and size to the dynamic characteristics of motion. In this case, in the

case of optical video surveillance, the task of monitoring the aerial situation becomes a class of tasks that can be solved by methods of digital video image processing and automatic pattern recognition using machine learning and artificial intelligence.

In the field of automatic image recognition, the following terminology is adopted to define tasks (Fig. 1).

Classification – definition of a class of one object on the image with indication of degree of reliability of the accepted automatic decision.

Classification with localization – classification of one object and indication of its place on the image.

Detection – definition of a class of each object on the image.

Video image processing includes a pipeline of tasks, among which detection of a moving flying object, determination of AO type, determination of object characteristics. In the proposed approach, we are in the framework of the task of determining the type of AO, pre-detected by the signs of motion and localized on the digital image. In accordance with the above terminology, the task of determining the type of AO refers to image classification problems.

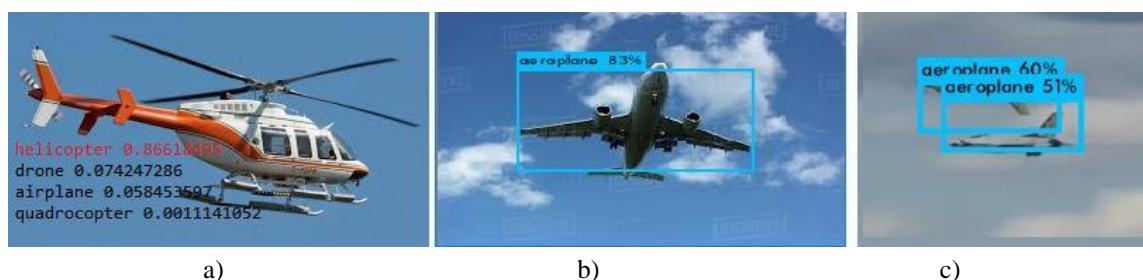


Fig. 1. Classification (a); classification with localization (b); detection (c)

Their solution is usually based on the use of deep convolutional neural networks. Such a solution for the problem we study is somewhat redundant due to its universality. Obviously, in the field of AO-type recognition object contours provide enough information to solve the problem even without taking into account the color and texture. In human visual AO-type recognition, the geometric shape of the object, represented by the outer contour, is usually a minimally sufficient feature for recognition even under the condition of geometric distortion (displacement, orientation, scale).

This problem was solved by the authors in previous studies [1], where it was assumed that the image of the object is in some idealized form: not hidden by clouds and well distinguishable on the background. However, in such conditions and under different illumination conditions, the problem of separating the background of the image from the object under study arises. In the field of image analysis, this is called the background subtraction problem, which is part of the broader problem of semantic segmentation.

This paper discusses the problem of processing video frames from an air monitoring camera mounted on a turntable to construct a binary image with the object under study highlighted. This image is necessary for further solving the problem of AO type recognition by the obtained numerical characteristics of its contour.

The problem of applying different algorithms of segmentation and background subtraction is relevant for real applications, which is confirmed by the works [2, 3], which are devoted to the issues of segmentation and background subtraction for video images, including the tasks of AO tracking.

Analysis of the problem and existing methods

The problem of optical AO recognition and tracking is the subject of many years of research due to the interest in its solution for both civil and military applications [4]. Classical recognition methods were based on comparing

an unknown image or its specific characteristics with a library of known images [5].

Outstanding results in computer vision related to deep neural networks and increasing computer performance have made the task of computer visual tracking one of the most prominent research topics. A study [6] provides a detailed review of advances in the use of deep convolutional neural networks for arbitrary image recognition.

However, as shown in [7, 8], deep learning methods require a large number of training sets and significant computer time to train the network.

The study [9] shows that the use of transfer learning, based on inexpensive additional training of pre-trained network, allows solving part of the problems of deep neural networks. However, at the same time there remains the problem of redundancy of the universal deep neural network-based approach in the recognition of airborne object types. The description of an object image by its contour is enough for the task of air object type recognition. In this case, much less redundant information is used than in the analysis using deep neural networks, which gives a number of advantages. In addition, as noted in [2], in the application of image processing algorithms, there is indeed a gap between the methods used in basic research and in real-world applications. Recently, in practical systems, there is an interest in the use of improved classical image processing methods.

To solve the problem of recognition of types of airborne objects in the general task of air monitoring, we use a camera attached to a turntable. Therefore, the algorithm for the foreground object extraction must take into account the changing illumination of the object and the background. This means that the algorithm must be time-critical in terms of execution time. Classical image processing methods meet these requirements well [10]. In [11] the use of these methods to recognize types of air objects is investigated in detail.

In our study, we consider one of the tasks of image preparation for recognition – the task of image

segmentation and background subtraction. This problem is solved after moving objects are detected and localized in a sequence of video frames by optical flow methods, based on the study of frame difference [12, 13].

Optical flow is one of the most widely used tracking algorithms for computer vision. Solutions are known for background subtraction algorithms in the case of shooting a fixed scene with a static camera [14].

However, in the conditions of our problem, the background is constantly changing, which interferes with the detection of moving targets [15]. Estimates and comparisons for background subtraction methods are studied in detail in [16], among which weighted moving mean, Gaussian mixture model (GMM), adaptive background learning, etc.

Highlighting the previously unsolved parts of the overall problem. Aim of the work

The system of visual monitoring of the air situation is an intelligent hardware-software complex, which includes a camera fixed on a rotating stand, and a computer control and information processing unit. The software for information processing consists of various interconnected functional modules of video image processing. Among these modules we can distinguish a pipeline of software modules, which solve

the problem of determining the type of AO as an image classification problem, as defined in the introduction. The solution of this problem is based on the input information received from the moving air object detection and localization module by optical flow methods, as already described above.

In an ideal situation, image preprocessing with localized AO should be reduced to image binarization followed by contour calculation, as presented in fig. 2.

To obtain the numerical characteristics of a contour, the localized image (fig. 2, a) must first be reduced to the black-and-white form (fig. 2, b). Then it is necessary to select a set of points of a contour (fig. 2, c) and represent it as a sequence of their coordinates. As a result of such processing the image of an air object will look like an ordered sequence of points $z(k) = (x_k, y_k) = f(t_k)$, $k = 0, \dots, N-1$, which is a discretized representation of a continuous two-dimensional curve $f(t) \subset R^2$, describing the contour. The further procedure of air type recognition is described in detail in our paper [1] and is reduced to obtaining a vector of features based on Fourier descriptors for the curve $f(t)$. The feature vector is the input for a pre-trained classifier based on a traditional multilayer neural network.

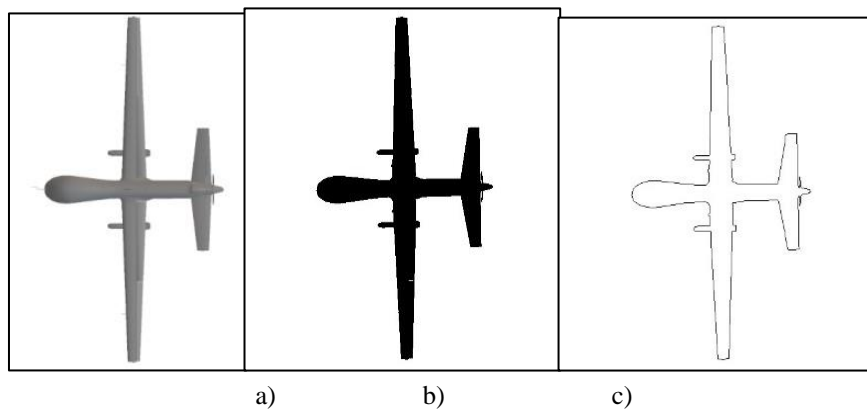


Fig. 2. Stages of image preprocessing:

a – initial image of localized object; **b** – black and white image; **c** – contour image of object

Generally speaking, a block diagram (fig. 3) can represent the algorithm of the air object type recognition procedure.

When implementing this algorithm, a number of problems arise. The AO image can be obtained with varying degrees of distortion, which is associated

with the size of the image, its blurring due to the aberration of light rays (see fig. 4).

In addition, the image of the object may be partially hidden by clouds or poorly distinguishable against their background (fig. 5). The task is to automatically, without human involvement, by means of the software

highlight a clear outline in conditions of visual disturbances (illumination, cloud cover).

Segmentation of objects and background subtraction under conditions of changing background and illumination, occlusion, shape distortion, motion blur, zooming, etc. are important areas of research in computer vision.

The aim of this article is to develop a video image processing module, which allows to select air objects for their further classification based on contours and to increase the efficiency of air object type recognition in a visual air monitoring system. To solve this problem, a modified interactive algorithm in the mode of automatic object silhouette extraction is used.

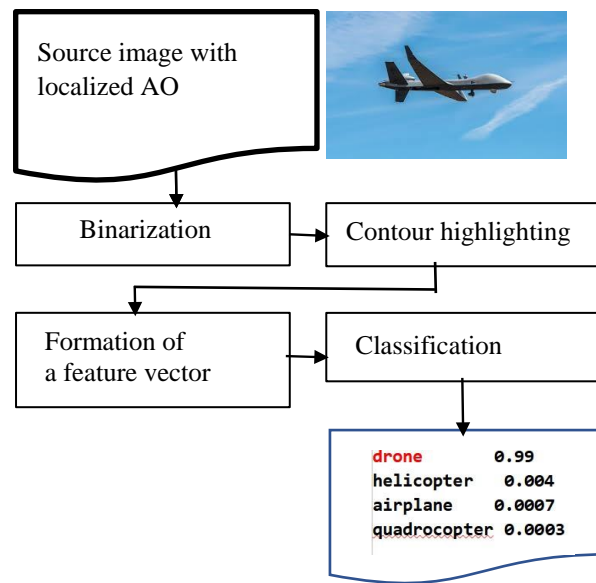


Fig. 3. Generalized block diagram of the air object type recognition procedure algorithm

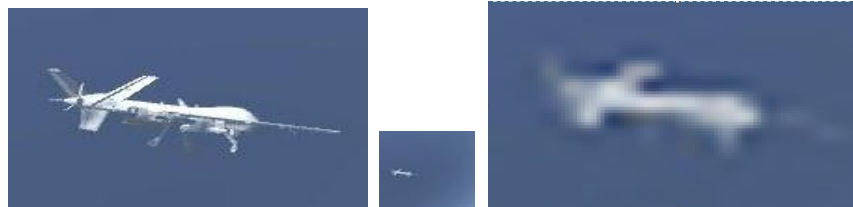


Fig. 4. A good-resolution image of the MQ-9 Predator UAV, the actual image, and its enlarged view



Fig. 5. Image of the MQ-9 Predator UAV in the ideal view for recognition and in conditions of visual interference

Materials and methods

When preparing an image for outline extraction, the main focus is on the brightness characteristics of the pixels in all three color channels of the color image. The optimal way to simplify this problem is to reduce the color three-channel RGB image to

a grayscale view (fig. 6). Such a brightness scale conveys 256 gradations from 0 – black, to 255 – white.

According to [17], the conversion to grayscale is performed by the formula

$$y = 0,2126 r + 0,7152 g + 0,0722 b, \quad (1)$$

where r , g , b are the color components of the original image, and y is the output value for a pixel of the image in grayscale.

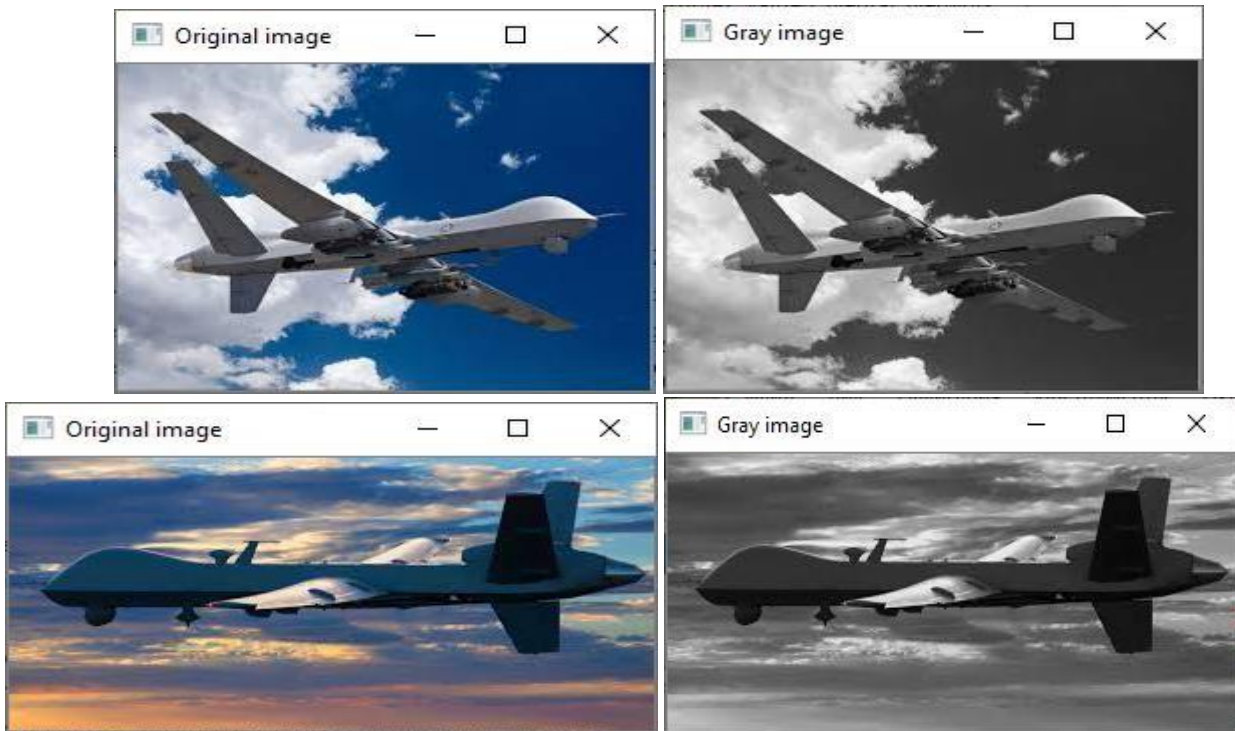


Fig. 6. The result of converting the image to grayscale mode

In an air monitoring system, the camera's shutter and iris control system may not have time to respond to too rapid changes in light intensity when the camera is rotating on a turntable. Sometimes the contrast range is too large for the sensors. In addition, it usually requires different exposure times for dark and light areas for a "perfect" image. In artistic and special photography for better transfer of details there is even a technique of high dynamic range (HDR), which allows to combine several frames of the object with different exposures.

In the project that we are implementing, it is possible to expand the dynamic range of the image and improve the contrast of the image by applying the algorithm of histogram correction. For this purpose a graphical distribution of image intensity in the form of a histogram is created, which determines the number

of pixels for each considered value of intensity in the grayscale range from 0 to 255.

With the correction, the original distribution becomes a broader and more homogeneous distribution of intensity values, so that they are distributed more evenly over the entire range. The algorithm normalizes brightness and increases image contrast (fig. 8).

For contour detection, a good result can be obtained after conversion to a binary (black and white) image, as shown above (fig. 1, b). In this case, the boundary point is unambiguously found by the change in the pixel color.

The binarization transformation can be determined by assigning a threshold value. The original image converted to a grayscale view with a pixel brightness range of 0 to 255 must be reduced to a binary view (0 or 1). If the pixel value is less than the threshold value, it is set to 0, otherwise to 1.

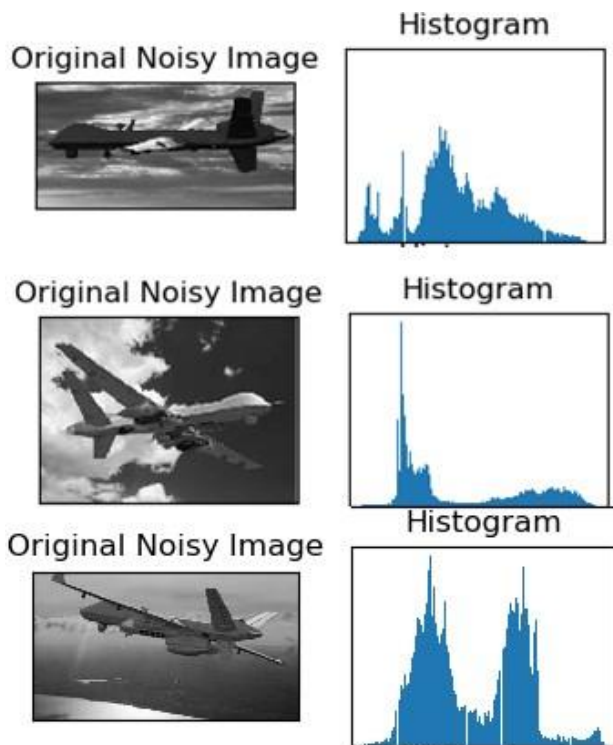


Fig. 7. Determining intensity histograms for grayscale images

Different variations of the threshold binarization algorithm (with lower and upper threshold, with double restriction, incomplete thresholding, etc.) presented in fig. 9 can be used.

Naturally, the result of the threshold conversion depends on the value of the threshold. Fig. 10 shows binarization results for different threshold values.

Using a fixed threshold may not be justified for different lighting conditions in different areas. In this case, adaptive threshold setting is used. The algorithm determines the threshold for a pixel based on a small area around it. In this way, we obtain different thresholds for different areas of the same image, which gives better results for images with different illumination for different averaging modes (fig. 11).

As a binarization method, the adaptive Otsu method oriented to a bimodal image, in which the histogram has two distinctly expressed peaks, has proven itself well. In such images, the optimal threshold is in the middle of these two values. Otsu method determines a global threshold value from image histogram.

Fig. 12 shows examples of the algorithm in the case of a bimodal image.

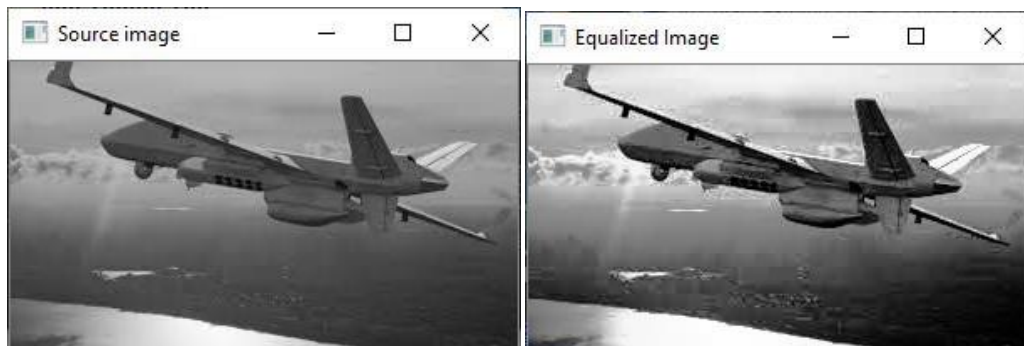


Fig. 8. The result of the alignment of the gray-scale image histogram

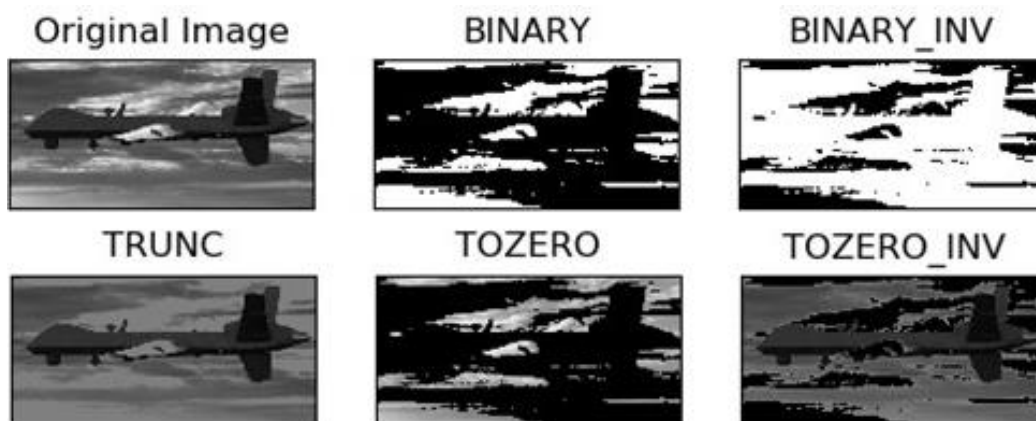


Fig. 9. Results of applying different types of threshold conversion

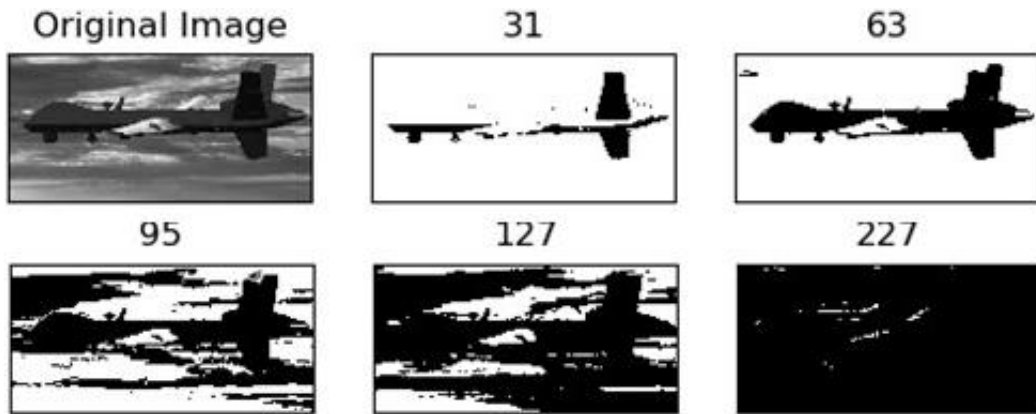


Fig. 10. Binarization results for different threshold values

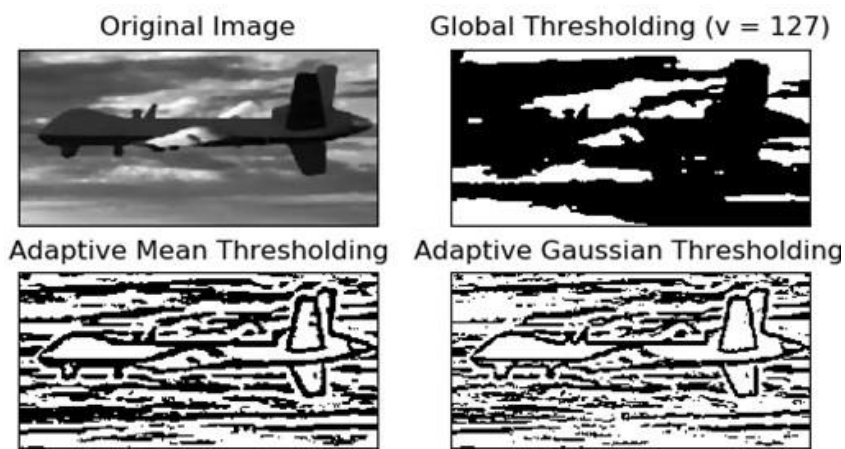


Fig. 11. Adaptive binarization results for different averaging modes

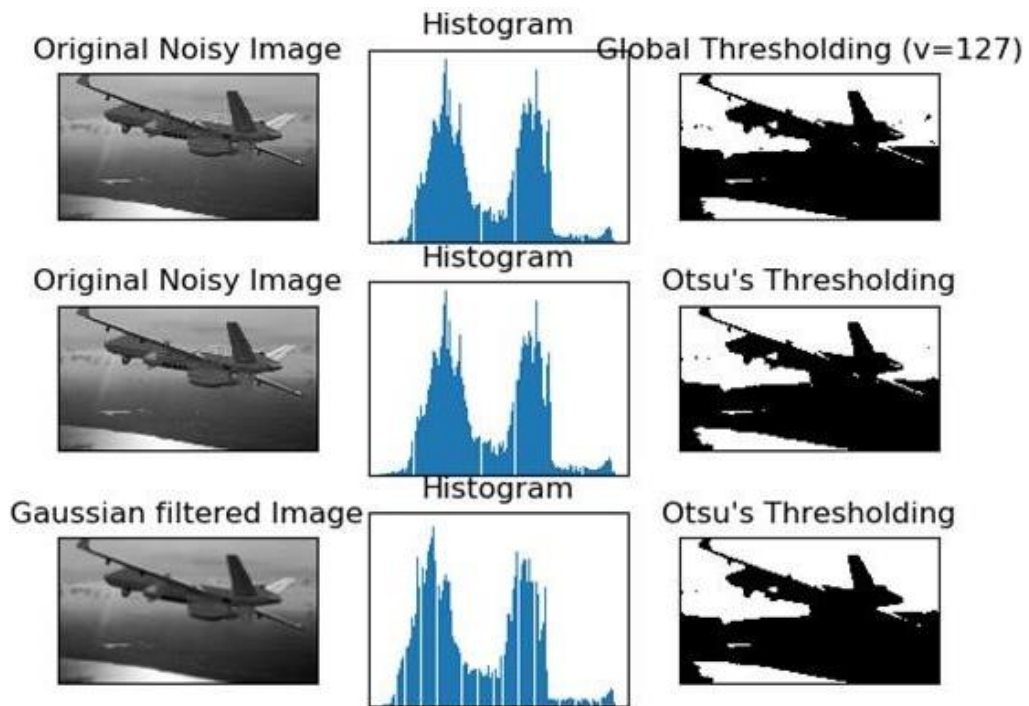


Fig. 12. Otsu binarization results for bimodal image

The disadvantage of finding objects using threshold filtering is the need to individually select the threshold value for each image, and even in this case, the separation from the background is not always possible, as seen in fig. 12.

It is possible to propose a method that goes through all possible thresholds and builds a recognition procedure for each of them. This would be possible when searching for any one particular contour, which is not suitable, for air objects having different contour shapes depending on the position with three degrees of freedom in relation to the camera.

Research results and their discussion

The recognition of airborne object types is based on the procedure of object contour extraction on the binary image obtained as a result of the initial image preprocessing. In this case, the GrabCut algorithm [18] is used. In this algorithm, user labeled foreground and background areas were used to build distribution histograms for these two classes of areas. The algorithm itself assumes that the remaining unlabeled pixels have a distribution similar to one of the two distributions. The energy functional was constructed based on the assumption that the pixels in each of the regions are connected. The assignment of a pixel to one of the regions was performed as a result of minimization of the energy functional.

The algorithm is based on the Gaussian mixture model (GMM), which is used to model the foreground and background. Depending on the data that represents the background and foreground areas, the GMM learns and creates a new pixel distribution. That is, unknown pixels are labeled either probable foreground or probable background, depending on their relationship to other, strictly labeled pixels in terms of color statistics.

References

1. Yesilevskiy, V., Tevyashev, A., Koliadin, A. (2020), "A method of air object recognition based on the normalized contour descriptors and a complex-valued neural network", *Eastern-European Journal of Enterprise Technologies*, No. 6, P. 48–57. DOI: 10.15587/1729-4061.2020.22003
2. Garcia-Garcia, B., Bouwmans, T., Rosales Silva, A. J. (2020), "Background subtraction in real applications: Challenges, current models and future directions", *Computer Science Review*, Vol. 35, P. 1–42. DOI: 10.1016/j.cosrev.2019.100204
3. Molloy, Timothy L., Jason, J. Ford and Luis Mejías Alvarez (2017), "Detection of aircraft below the horizon for vision-based detect and avoid in unmanned aircraft systems", *Journal of Field Robotics*, Vol. 34, Issue7, P. 1378–1391. DOI: 10.1002/rob.21719
4. Huihui, Li, Xing, Jin, Ning Yang, Zhe Yang (2015), "The recognition of landed aircrafts based on PCNN model and affine moment invariants", *Pattern Recognition Letters*, Vol. 51. DOI: 10.1016/j.patrec.2014.07.021

The interactivity of the algorithm can be excluded if we make the following assumptions:

- 1) everything within a given indent from the edge of the image is the background;
- 2) everything within a given area in the center of the image is the foreground.

Fulfillment of these conditions is possible because the system architecture is designed in such a way that the preliminary stage of video processing the system of moving object detection by optical flow method selects pixels that determine the area of interest.

A computational experiment was conducted on a test set of 40 images, which showed the correct result in 37 cases (93%), which confirms the possibility of using the proposed method. The results of the algorithm are shown in fig. 13.

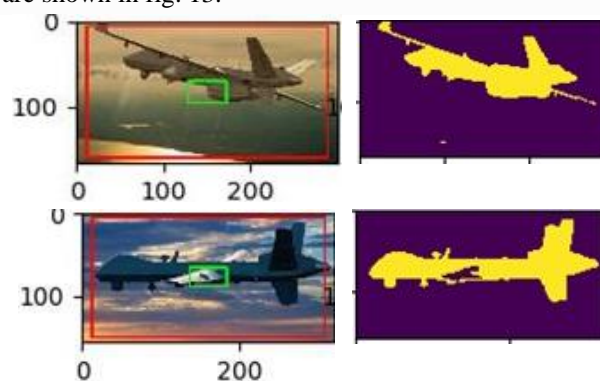


Fig. 13. Results of the background removal algorithm

Conclusion

Implementation of the developed background segmentation and subtraction module for the system of optical monitoring of the air situation allowed to solve the problem of video images segmentation with high reliability for further recognition of air objects in the system of optical monitoring of the air situation in automatic mode, thus increasing the operational efficiency of this system.

5. Suzuki, S. (1985), "Topological structural analysis of digitized binary images by border following", *Computer Vision, Graphics, and Image Processing*, Vol. 30 (1), P. 32–46. DOI: 10.1016/0734-189X(85)90016-7
6. Neha, Sh., Vibhor, J., Anju, M. (2018), "An Analysis Of Convolutional Neural Networks For Image Classification", *Procedia Computer Science*, No. 132, P. 377–384. DOI: 10.1016/j.procs.2018.05.198
7. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017), "ImageNet classification with deep convolutional neural networks", *Commun. ACM* 2017, No. 60, P. 84–90. DOI: 10.1145/3065386
8. Szegedy, C., Liu, W., Jia (2015), "Going deeper with convolutions", 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, P. 1–9. DOI: 10.1109/CVPR.2015.7298594
9. Yesilevskiy, V., Tevyashev, A., Koliadin, A. (2019), "Transfer learning in aircraft classification", *Information systems and technologies IST-2019, The 8-th International Scientific and Technical Conference September 9–14, Kobleve–Kharkiv, Ukraine*, P. 132–135.
10. Ishchenko, A. (2019), "Development of an intelligent processing system module for scanned documents based on the combined image segmentation method", *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (8), P. 44–53. DOI: 10.30837/2522-9818.2019.8.044
11. Strotov, V., Babyan, P., Smirnov, S. (2017), "Aerial object recognition algorithm based on contour descriptor", *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W4, P. 91–95. DOI: 10.5194/isprs-archives-XLII-2-W4-91-2017
12. Thai, V.-P., Zhong, W., Pham, T., Alam, S., Duong, V. (2019), "Detection, Tracking and Classification of Aircraft and Drones in Digital Towers Using Machine Learning on Motion Patterns", 2019 *Integrated Communications, Navigation and Surveillance Conference (ICNS)*, P. 1–8. DOI: 10.1109/ICNSURV.2019.8735240
13. Jie, Z. (2020), "An Aircraft Image Detection and Tracking Method Based on Improved Optical Flow Method", 2020 *Chinese Automation Congress (CAC)*, P. 2512–2516. DOI: 10.1109/CAC51589.2020.9326473
14. Bengtsson, T., Mckelvey, T., Lindström, K. (2016), "COOn Robust Optical Flow Estimation on Image Sequences with Differently Exposed Frames using Primal-Dual Optimization", *Image and Vision Computing*, No. 57. DOI: 10.1016/j.imavis.2016.11.003
15. Rashid, M., Thomas, V. (2016), "A Background Foreground Competitive Model for Background Subtraction in Dynamic Background", *Procedia Technol*, No. 25, P. 536–543. DOI:10.1016/j.protcy.2016.08.142
16. Sobral, A., Vacavant, A. (2014), "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos", *Comput. Vis. Image Underst.*, No. 122, P. 4–21. DOI: 10.1016/j.cviu.2013.12.005
17. Recommendation ITU-R BT.709-6 (2015), Parameter values for the HDTV standards for production and international programme exchange, 19 p.
18. Rother, C., Kolmogorov, V., Blake, A. (2004), "GrabCut": interactive foreground extraction using iterated graph cuts", *ACM SIGGRAPH 2004 Papers*, P. 309–314. DOI:10.1145/1186562.1015720

Received 26.09.2022

Відомості про авторів / Сведения об авторах / About the Authors

Єсілевський Валентин Семенович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри прикладної математики, Харків, Україна; e-mail: valentyn.yesilevskiy@nure.ua; ORCID ID: <https://orcid.org/0000-0002-5935-1505>

Єсілевский Валентин Семенович – кандидат технических наук, доцент, Харьковский национальный университет радиоэлектроники, доцент кафедры прикладной математики, Харьков, Украина.

Yesilevskiy Valentyn – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Applied Mathematics, Kharkiv, Ukraine.

Колядін Антон Володимирович – Харківський національний університет радіоелектроніки, молодший науковий співробітник кафедри прикладної математики, Харків, Україна; e-mail: anton.koliadin@nure.ua; ORCID ID: <https://orcid.org/0000-0001-5552-5080>

Колядин Антон Владимирович – Харьковский национальный университет радиоэлектроники, младший научный сотрудник кафедры прикладной математики, Харьков, Украина.

Koliadin Anton – Kharkiv National University of Radio Electronics, Junior Researcher at the Department of Applied Mathematics, Kharkiv, Ukraine.

Середа Олена Григорівна – Харківський національний університет радіоелектроніки, старший викладач кафедри прикладної математики, Харків, Україна; e-mail: olena.sereda@nure.ua; ORCID ID: <https://orcid.org/0000-0003-3579-0092>

Середа Елена Григорьевна – Харьковский национальный университет радиоэлектроники, старший преподаватель кафедры прикладной математики, Харьков, Украина.

Sereda Olena – Kharkiv National University of Radio Electronics, Senior Lecturer at the Department of Applied Mathematics, Kharkiv, Ukraine.

РОЗРОБЛЕННЯ МОДУЛЯ ОБРОБКИ ВІДЕО ДЛЯ ЗАВДАННЯ РОЗПІЗНАВАННЯ ПОВІТРЯНИХ ОБ'ЄКТІВ ЗА ЇХНІМИ КОНТУРАМИ

Предметом дослідження в статті є модуль автоматичної сегментації та віднімання фону, створений на основі конвеєра методів попереднього оброблення зображення та модифікованого методу інтерактивної сегментації зображень, і впроваджений у систему оптичного моніторингу повітряної обстановки. **Метою** роботи є розроблення модуля сегментації зображень для підвищення ефективності розпізнавання типу повітряного об'єкта на відеозображенні в системі візуального моніторингу повітряної обстановки шляхом якісної автоматичної сегментації. Для вирішення цього завдання використовується модифікований інтерактивний алгоритм у режимі автоматичного виділення об'єкта на зображенні, що дає змогу точніше без участі оператора визначати пікселі переднього плану зображення для подальшого розпізнавання типу повітряного об'єкта. У статті вирішуються такі **завдання**: аналіз наявних методів бінаризації кольорового зображення для семантичної сегментації зображень, що використовуються в системах розпізнавання образів; розроблення конвеєра методів автоматичної сегментації зображень у системі оптичного моніторингу повітряної обстановки. У роботі використовуються **методи** цифрового оброблення зображень, фільтрації та семантичної сегментації зображень, аналізу графів. Отримано такі **результати**: унаслідок оброблення зображень за допомогою запропонованого модуля сегментації та віднімання фону підтверджено працездатність процедур модуля. Розроблений конвеєр методів, вміщений у модуль, демонструє правильну сегментацію 93% тестових зображень в автоматичному режимі без участі оператора, що дозволяє зробити висновок про ефективність застосування запропонованого модуля. **Висновки**: упровадження розробленого модуля сегментації та віднімання фону для системи оптичного моніторингу повітряної обстановки дозволило з високим ступенем достовірності вирішити завдання сегментації відеозображень для подальшого розпізнавання повітряних об'єктів у системі оптичного моніторингу повітряної обстановки в автоматичному режимі, завдяки чому збільшилася ефективність експлуатації цієї системи.

Ключові слова: сегментація зображень; віднімання фону; оброблення зображень; розпізнавання повітряних об'єктів; оптичний моніторинг повітряної обстановки.

РАЗРАБОТКА МОДУЛЯ ОБРАБОТКИ ВИДЕО ДЛЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ВОЗДУШНЫХ ОБЪЕКТОВ ПО ИХ КОНТУРАМ

Предметом исследования в статье является модуль автоматической сегментации и вычитания фона, созданный на основе конвейера методов предварительной обработки изображения и модифицированного метода интерактивной сегментации изображений, и внедренный в систему оптического мониторинга воздушной обстановки. **Целью** работы является разработка модуля сегментации изображений для повышения эффективности распознавания типа воздушного объекта на видеозображении в системе визуального мониторинга воздушной обстановки путем качественной автоматической сегментации. Для решения этой задачи используется модифицированный интерактивный алгоритм в режиме автоматического выделения объекта на изображении, который позволяет точнее без участия оператора определять пиксели переднего плана изображения для дальнейшего распознавания типа воздушного объекта. В статье решаются следующие **задачи**: анализ существующих методов бинаризации цветного изображения для семантической сегментации изображений, которые используются в системах распознавания образов; разработка конвейера методов для автоматической сегментации изображений в системе оптического мониторинга воздушной обстановки. В работе используются **методы** цифровой обработки изображений, фильтрации и семантической сегментации изображений, анализа графов. Получены следующие **результаты**: результаты обработки изображений с помощью предложенного модуля сегментации и вычитания фона подтверждают работоспособность процедур модуля. Разработанный конвейер методов, включенный в модуль, демонстрирует правильную сегментацию в 93% тестовых изображений в автоматическом режиме без участия оператора, что позволяет сделать вывод об эффективности применения предложенного модуля. **Выводы**: внедрение разработанного модуля сегментации и вычитания фона для системы оптического мониторинга воздушной обстановки позволило с высокой степенью достоверности решить задачу сегментации видеозображений для дальнейшего распознавания воздушных объектов в системе оптического мониторинга воздушной обстановки в автоматическом режиме, благодаря чему увеличилась эффективность эксплуатации данной системы.

Ключевые слова: сегментация изображений; вычитание фона; обработка изображений; распознавание воздушных объектов; оптический мониторинг воздушной обстановки.

Бібліографічні описи / Bibliographic descriptions

Есілевський В. С., Колядін А. В., Серєда О. Г. Розроблення модуля обробки відео для завдання розпізнавання повітряних об'єктів за їхніми контурами. *Сучасний стан наукових досліджень та технологій в промисловості*. 2022. № 3 (21). С. 16–25. DOI: <https://doi.org/10.30837/ITSSI.2022.21.016>

Yesilevskiy, V., Koliadin, A., Sereda, O. (2022), "Development of a video processing module for the task of air object recognition based on their contours", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (21), P. 16–25. DOI: <https://doi.org/10.30837/ITSSI.2022.21.016>

M. KORABLYOV, S. LUTSKYY

SYSTEM-INFORMATION MODELS FOR INTELLIGENT INFORMATION PROCESSING

The subject of the study is system-information models of processes and systems and their use for intelligent processing of information in production tasks. The use of intelligent information processing in production management systems is currently one of the key areas of development of informatics. **The aim** of the work is to develop system-information models of processes and systems for intelligent information processing allowing to analyze and solve production problems, in conditions of uncertainty. In the article **the following tasks** are solved: to analyze approaches to the definition of information characteristics of processes and systems; to develop the basis for modeling of system-information processes and systems for intelligent information processing; to develop system-information models and ways of their application for intelligent information processing in the tasks of production. **The following methods** are used: system-information approach to processes and systems; system-information modeling of processes and systems. **The following results** were obtained: the analysis of approaches to the definition of information characteristics of processes and systems; developed principles of modeling system-information processes and systems for intelligent processing of information; introduced the concepts of system information and information measure; developed system-information models and methods of their application for the intelligent processing of information in the tasks of production. **Conclusions.** The development of methods for solving various classes of practical problems using intelligent information processing is one of the key areas of research in computer science. The developed system-information models of processes and systems for intelligent information processing allow analyzing and solving problems. Thereby increase the efficiency of solving problems of analysis, synthesis and forecasting of production systems and technologies, as well as problems of production management. The system-information approach to processes and systems operates with new concepts – system information and information measure, it allowed developing system-information models for intelligent processing of information, as well as ways of their application at stages of product life cycle, which allowed solving problems of production. System-information models of processes and systems describe interaction between source and receiver on information level on the basis of sensitivity threshold. The communication channel between the source and the receiver of information operates, as a rule, under conditions of uncertainty, which can lead to the loss of information during transmission due to possible changes in the characteristics of the system. To describe their interaction, some models of intelligent information processing can be used, in particular, neural network models or fuzzy inference models. Their use will improve the efficiency of receiver state prediction, taking into account the state of the transmitter and the conditions of communication channel operation. The presented article has shown the relevance of developing system-information models for intelligent information processing at the levels of data reception, interpretation and communication, which allows expanding the class of solved production tasks.

Keywords: system-information models; system information; information measure; sensitivity threshold; intelligent information processing.

Introduction

The development of methods for intelligent information processing of processes and systems is currently one of the key areas of development of computer science.

The definition of the term information plays a key role in the development of intelligent information processing methods. Despite its widespread occurrence, the concept of information remains one of the most debatable in science, and the term can have different meanings in different branches of human activity [1, 2]. The development of approaches to the definition of the meaning of the term information allows us to expand the boundaries of the application

of intelligent information processing in specific branches of human activity.

The systems-information approach to processes and systems is one of the approaches to defining the meaning of the term information, and is used in developing models for intelligent information processing. As a scientific direction in information theory, it was summarized in [3, 4]. From the position of the system-information approach the amount of a priori and a posteriori information, which the object possesses, as well as the information process of interaction between objects are subject to measurement. The model of a priori information characterizes an object before interaction, the model of information process – in the process of interaction, and the model of posterior information – after interaction.

System-information models consist of attributes [3, 4]:

- 1) extent information (space) in three coordinates;
- 2) duration information (time);
- 3) intensity information (value) of objects properties.

The above interrelated information attributes represent a system-information space as a whole. Each manifested property of a system (internal and external) in the information space is characterized by the amount of information. This quantity is determined by a set of information of intensity, extent and duration of the system's properties and characterizes the process of its functioning.

Interaction of system and environment is reflected by a fragment of reality of manifestation of system in the information space. The information space structures attributes by quanta of sensitivity threshold interval of lower and upper values of intensity of properties, duration and extent. The change of system state characteristics depends on the change of attributes and their values that make up the information space. In this case, the change of the state of the system occurs in the presence of information connection between the system and the environment.

The methodology of system-information modeling of processes and systems is based on the principle of determining the information measure (number), which is the ratio of the total (limits of possible) value of an attribute to its partial (variable) value of sensitivity threshold. This dimensionless measure indicates the information of the place of the particular in the general.

The development of system-information models for intelligent information processing in tasks of various classes, including production, and is an important direction in the development of computer science.

Analysis of approaches to defining information characteristics

The object of research in the science of computer science is "information". With all the different interpretations of the concept of information, it is indisputable that information is manifested in material and energy form in the form of a "signal". In its most general form, a "signal" is understood as a value that reflects in some way the state of the system. In this sense, it is natural to consider a signal as the result of some measurements carried out over the system in the process of its observation. The signal reflects both the information of the state of the system and is the cause of the change in the state of the interdependent system.

Information in general can be structured into the following main groups:

- 1) information of non-living matter (physical information, including technical information),
- 2) information of living matter (biological, including social),
- 3) artificial information (coded information), and
- 4) a combination of information from these three components.

Each of the presented information groups is characterized by its own type of information carrier – «signal», which can be represented in a material-energetic or immaterial-energetic form.

Material-energy signal is a time-varying physical quantity described by a function of time and can be classified:

- 1) by the physical nature of the information carrier – electrical, electromagnetic, optical, acoustic and others;
- 2) according to the way the signal is defined – regular (deterministic), defined by an analytical function and irregular (random), taking arbitrary values at any time.

The apparatus of probability theory is used for irregular signals. Depending on the function that describes the parameters of the signal, there are: continuous (analog), continuous-quantized, discrete-continuous and discrete-quantized signals.

Intangible-energy signal is a carrier of information, which so far has not been sufficiently investigated. An example is a signal, which is a representation of such phenomenon as entropy of a system, or a signal in a quantum system.

A special kind of signals are "signs", which, unlike signals of natural origin, are created by self-organizing systems and are intended for the transmission and storage of information. A sign is a material sensually perceived signal (phenomenon, action), which acts as a representative of another object of property or attitude. The science that studies the properties of signs and sign systems is called semiotics. The subject of semiotics is the questions of connection of signs with each other, signs with the phenomena of the external world and signs with the subjects using them as a tool for the purposes of communication.

With all the variety, complexity and multi-level manifestation of information processes, they all contain the basic components:

- 1) perception by the system of some phenomenon as a signal;

2) change of the system's state under the influence of a received signal;

3) spatial and temporal transmission of signals.

These three components of the information process, which at different levels of forms of reflection acquire specific characteristics and can be expressed in such concepts as reception, interpretation and communication.

One of the important issues in the development of information technologies that are implemented in intelligent information processing systems is the use of the mathematical form of representation of the processed information. This affects the limitations of the possible choice of information technology for the list of the types of tasks to be solved in intelligent systems, as well as the indicators of the effectiveness of their application.

Information technologies of intellectual information processing systems in conditions of uncertainty are widely used for solving complex tasks that are difficult to formalize [5]. The lack of a universal approach to the form of information representation when solving such problems leads to the need to develop and apply new methods, information technologies, and their complex combination with traditional methods of artificial intelligence.

The main limitations of the known methods and information technologies currently used in intelligent systems for solving problems that are difficult to formalize are due to the methods of information representation and insufficient efficiency in solving the problems of learning, adjustment and adaptation to the problem domain, processing the measured values of incomplete and inaccurate source information, data interpretation and knowledge accumulation, uniform presentation of information coming from different sources, etc. These limitations can be eliminated on the basis of using the system-information approach to the formalization of the measured values of the system information and promising hybrid methods of soft computing – Soft computing [6]. With the help of methods that use a hybrid approach to the representation of information values, it is possible to generate new knowledge under conditions of knowledge deficit and uncertainty, which allows expanding the logical possibilities of intelligent systems that use these technologies.

Significant scientific interest and important practical importance is the development of information models of the subject area in conditions of uncertainty,

characterized by incompleteness, unreliability, uncertainty of the initial information, diversity and complexity of the influence of various factors on the decision process, etc. Depending on the degree of uncertainty, information processing tasks can be divided into [7]:

1) tasks, for which an objective assessment of solution results or comparative assessment of at least two solutions is possible;

2) tasks for which an objective assessment is impossible, and it is replaced by expert assessments of specialists in the subject area under study.

The first type of problems refers to poorly structured and difficult to formalize problems characterized by the presence of both quantitative and qualitative dependencies between the elements of the systems under study, with qualitative dependencies tending to dominate. Tasks of this type can be structured as system information on the basis of the theory of systems-information approach.

The second type of problems refers to unstructured (unformalizable) problems whose peculiarity consists only in a qualitative description of the systems, while quantitative dependences between the elements of the system are absent, which determines the necessity of using expert evaluations [7].

Currently, the most effective models of intelligent systems, implemented within the hybrid approach for intelligent information processing, are neuro-fuzzy models [8], which combine the universal approximating capabilities of neural networks, as well as the transparency and interpretability of fuzzy inference systems. The structures of neuro-fuzzy models of intelligent systems in most cases are fixed and, as a rule, redundant, which leads to a decrease in the speed of learning and adaptation of models when external conditions change. A way out in this situation is associated with the possibility of building neuro-phase models of intelligent systems, in which not only parameters, but also the structure of the model as a whole is trained and adapted, which allows to obtain the most effective hybrid models for specific problems of intelligent information processing.

The problem arising in the problems of intelligent information processing in conditions of uncertainty is the problem of constructing an adequate information model of the investigated subject area, as well as the formation of effective models and procedures for finding optimal or close to them solutions. Methods for building information models, allowing a reliable description of design objects and processes occurring

in them, are determined by the nature of the tasks being solved. That is why the formation and choice of this or that information model can be carried out only after the definition of a particular problem, and even for the same problem, it is possible to use different types of models.

The process of finding solutions for a problem that is difficult to formalize can be represented by two methods.

The first one consists of the following stages [9]:

- 1) extraction (acquisition), representation and structuring of knowledge about the problem domain;
- 2) construction of adequate mathematical models of the subject area;
- 3) prediction of possible variants of solutions (search for solutions);
- 4) analysis and evaluation of effectiveness of proposed solutions;
- 5) choice of solution.

Since system information combines quantitative and qualitative characteristics of objects in the information description of the elements of the systems under study, the second method consists in formalizing the problem on the basis of system-information models [10] and consists of the following stages:

- 1) construction of adequate system-information models of the subject area;
- 2) analysis and evaluation of the information process;
- 3) determination of the main characteristics of information objects.

When building system-information models, the key issue is to determine the sensitivity threshold of the subject domain attributes. For example, fuzzy sensitivity threshold is determined using various expert methods, including statistical methods.

An information system is characterized by the presence of internal and external connections, which may have an energy or material character. A distinctive property of an information system is that the structure of energy and material connections is a carrier of information about the properties of the objects of the external world and the internal environment of the system. These links with essential structural features are information links.

The basis of the classical theory of information is based on the results of solving a number of problems on the syntactic level. It relies on the concept of «quantity of information», which is a measure of the frequency of use of signs, which in no way reflects the meaning or importance of the messages transmitted.

The exchange of information between two objects occurs due to deterministic or stochastic information process in accordance with the expressions [3, 4]:

$$I = \log_2 \frac{X}{\Delta x}, \quad I = \log_2 \frac{\mu}{\sigma}, \quad (1)$$

$$\log_2 \frac{X_1}{\Delta x_1} = \log_2 \frac{X_2}{\Delta x_2}, \quad \log_2 \frac{\mu_1}{\sigma_1} = \log_2 \frac{\mu_2}{\sigma_2}.$$

where: Δx – sensitivity threshold X , σ – standard deviation, μ – expected value.

If the system information is transmitted in error, in the equations of the information process the sign of equality (=) is replaced by a sign of approximation (\approx).

System information is characterized by quantity, quality and value. The quantity of system information is calculated as the logarithm of the ratio of the sign value to the sensitivity threshold in accordance with (1). The quality of system information refers to information processes and characterizes the value of possible losses of received information during the transfer of optimal values of the quantity of system information. The less possible losses in the received information, the more qualitative the received system information, which is calculated on the basis of a posteriori information [3, 4]:

$$\sum_{i=1}^m I(x_i) = \sum_{j=1}^n I(y_j) \quad I(y)_{\text{кач}} = \frac{\sum_{i=1}^n I_{x_i}}{\sum_{j=1}^n (1 + |\Delta y_j - U_{y_j}|)}. \quad (2)$$

where: Δy – sensitivity threshold, U – expanded uncertainty.

The value of system information refers to information processes and is characterized by the efficiency of using the optimal value of the quality of system information:

$$\sum_{i=1}^m I(x_i) = \sum_{j=1}^n I(y_j)$$

$$I(y)_{\text{вал}} = \frac{\sum_{j=1}^n I_{y_j}}{\sum_{i=1}^m I_a(x_i) \times \sum_{j=1}^n (1 + |\Delta y_j - U_{y_j}|)}. \quad (3)$$

where: Δy – sensitivity threshold, U – expanded uncertainty.

The analysis of approaches to determining the information characteristics of processes and systems showed the relevance of developing system-information models for intelligent information processing at the levels of data reception, interpretation and communication, which will expand the class of solved production tasks.

System-information models and their application to intelligent information processing

The principles of developing system-information models are based on the scientific provisions in the field of information theory outlined in [10]. The significance of these provisions for science is that they eliminate the drawback of the modern scientific paradigm – the desire to describe everything using the language of dimensionless models, and also allow to reflect such a property of systems and processes as discreteness and specify their absolute optimum dimensions.

The information object of research of the system-information approach is a system – a set of elements X , which are in relations and information links with each other, which forms a certain integrity and unity. System information possessed by elements of the set X is characterized by the interval between the upper X_{up} and lower X_{low} limits of its manifestation, as well as the sensitivity threshold $\Delta x = (x - X_{low})/n$, where x is a discrete variable, n is a multiple of Δx – sensitivity threshold on the interval $X_{up} - X_{low}$. In this case the sensitivity threshold takes on the value on the interval $X_{low} - X_0$.

The system information of an object is characterized by the information measure, which is equal to the proportion of the ratio of the total value of the attribute to its private value. The system information indicates the place of the particular in the general. The information measure $|I(X)|$ is a function of the absolute value of the qualitative and/or quantitative proportion of the ratio

$$|I(X)| = f\left(\frac{X_{up} - X_{low}}{\Delta x}\right) \quad (4)$$

where: $\Delta x = x - X_{low}/n$ – sensitivity threshold.

The information measure is a dimensionless quantity in any system of physical quantities, and is a number. Information $I(X)$ is a dimensional value of a logarithmic function, and is measured in bits.

From the presented function of the information measure the following tasks are solved.

1. Definition of the discrete variable of a discrete value x :

$$x = \frac{(X_{up} - X_{low}) \times n}{|I(X)|} + X_{low}. \quad (5)$$

2. Determination of the upper limit value of X_{up} :

$$X_{up} = |I(X)| \Delta x + X_{low}. \quad (6)$$

3. Determination of the lower limit value X_{low} :

$$X_{low} = X_{up} - |I(X)| \Delta x. \quad (7)$$

4. Determination the sensitivity threshold value $\Delta x = \frac{x - X_{low}}{n}$:

$$\Delta x = \frac{X_{up} - X_{low}}{|I(X)|}. \quad (8)$$

5. Determination of the multiplicity n of the sensitivity threshold Δx in a discrete variable value x :

$$n = \frac{x - X_{low}}{\Delta x}. \quad (9)$$

The above formulas of transformation of information measure are used in intelligent information processing for tasks of analysis, synthesis of objects and forecasting based on the methodology of system-information modeling.

The presented approach to the definition of the concept of information differs from the generally accepted one, where numerical values of data X_{up} , X_{low} , Δx , x , but not their relations, are taken as information.

The next step is to formalize the information connection between elements of the set X , i.e. the information process of transmission of system information.

All qualitative and quantitative characteristics of information assume the presence of a transmitter and a receiver (sender and receiver) of information, i.e. in some kind of information interaction of objects. Formalization of information process of information transfer from sender to receiver describes the equilibrium of information measures of interacting objects:

System-information models of the information process allow us to solve prediction problems on the basis of knowledge of the information parameters of causes, which are informationally related to the information parameters of the receiver. Knowledge of receiver information parameters is obtained directly by the measurement method.

The system-information model of the information process looks like

$$\begin{aligned}
|I(X_i)| &= |I(X_j)|, \quad |I(X)| = f((X_{up} - X_{low})/\Delta x), \\
f_{trans} \left(\frac{X_{iup} - X_{ilow}}{\Delta x_i} \right) &= f_{rec} \left(\frac{X_{jup} - X_{jlow}}{\Delta x_j} \right), \\
\frac{X_{(trans)up} - X_{(trans)low}}{\Delta x_{(trans)}} &= \frac{X_{(rec)up} - X_{(rec)low}}{\Delta x_{(rec)}}, \\
n = \frac{x - X_{up}}{\Delta x}, \quad \Delta x &= \frac{x - X_{low}}{n}, \\
\frac{(X_{(trans)up} - X_{(trans)low})n_{(trans)}}{x_{(trans)} - X_{(trans)low}} &= \frac{(X_{(rec)up} - X_{(rec)low})n_{(trans)}}{x_{(rec)} - X_{(rec)low}}, \\
\log_2 \frac{(X_{(trans)up} - X_{(trans)low})n_{(trans)}}{x_{(trans)} - X_{(trans)low}} &= \log_2 \frac{(X_{(rec)up} - X_{(rec)low})n_{(trans)}}{x_{(rec)} - X_{(rec)low}}.
\end{aligned} \tag{10}$$

From the equation of system-information model of information process on the basis of proportional ratio of information parameters of attribute (number) of transmitter and receiver, we get knowledge about parameters of information transmitted by transmitter in conditions of uncertainty of sensitivity threshold.

That is, from the equation we determine all previously unknown knowledge about parameters of information of the transmitter beyond the limits of previous experience at this particular moment.

System-information models of processes and systems describe interaction between source and receiver on information level based on sensitivity threshold. The communication channel between source and receiver of information operates, as a rule, under conditions of uncertainty, which can lead to loss of information during transmission due to possible changes in system characteristics. To describe their interaction, some models of intelligent information processing can be used, in particular, neural network models or fuzzy inference models. Their use will improve the efficiency of receiver state prediction, taking into account the state of the transmitter and the conditions of communication channel operation.

The physical interpretation of the information transmission process can be represented as follows. Variable value of the transmitter Δx_2 increases under the influence of various factors till the moment of achievement of information connection to the threshold of sensitivity of its receiver X_{low} . At the moment of achievement of the threshold of sensitivity of the receiver to the transmitter the information resonance occurs between the transmitter and the

receiver, at that Δx_1 jumps up to the proportional value Δx_2 of the transmitter:

$$\Delta x_i = \frac{(X_{iup} - X_{ilow}) \times \Delta x_j}{X_{jup} - X_{jlow}}. \tag{11}$$

Thus, at the moment of information resonance between objects, the information of the proportional ratio of the common to the particular is transmitted from the transmitter to the receiver. With the known information of the transmitter $I(X_j)$ the information of the receiver $I(X_i)$ is calculated. With N transmitters independent of each other the receiver's information is added and the measures are multiplied:

$$\begin{aligned}
I(X_i) &= \sum_{j=1}^N I(X_j), \quad |I(X_i)| = \prod_{j=1}^N |I(X_j)| \\
|I(X_i)| &= \frac{X_{iup} - X_{ilow}}{\Delta x_i} = \frac{\prod_{j=1}^N (X_{jup} - X_{jlow})}{\prod_{i=1}^N \Delta x_j}.
\end{aligned} \tag{12}$$

If N transmitters from the set X are interdependent, then the formula for the information measure of the receiver when transmitting information from N interdependent transmitters has the following form

$$\begin{aligned}
I(X_i) &= \sum_{i=1}^N I(X_j), \\
|I(X_i)| &= \frac{X_{iup} - X_{ilow}}{\Delta x_i} = \frac{\prod_{j=1}^N (X_{jup} - X_{jlow})}{\prod_{j=1}^N (\Delta x_j + \Delta x'_j)}, \\
\Delta x'_j &= \frac{\prod_{j=1}^N (X_{N-jup} - X_{N-jlow})}{\prod_{j=1}^N \Delta x_{N-j}}.
\end{aligned} \tag{13}$$

In a closed system with a set of X elements with $N \rightarrow X$ interdependent elements, $\Delta x \rightarrow X$ up, and $I(X) \rightarrow 0$, i.e. the system tends to information zero equilibrium.

The general approach to system-information modeling has particular cases depending on the kind of information that is formalized. This occurs when the value of the set X is an argument function of a particular particular kind of information $B(I_i) = fI(X_i)$.

From the general approach, the following private options can be distinguished.

1. For identification of properties of physical world mankind uses a principle of information measure, at that fixes a value of threshold of sensitivity of physical property in the form of a standard of physical quantity $\Delta x = 1St(\text{standard})$, having presented value $X_{low} = 0$, at that $X_{up} = x$ is discrete variable X , and value of information measure of physical quantity has a form:

$$|I(X)| = f\left(\frac{x - X_{low}}{\Delta x}\right) \quad (14)$$

All natural sciences are built on this principle, which uses the proportion of the general to the particular as a generally accepted notion of information. When solving analysis and synthesis problems in the traditional way on the basis of a physical quantity to determine and optimize its place in the system, various rather complex scientific approaches have been developed.

The system-information approach greatly simplifies the solution of analysis and synthesis problems due to the possession of physical quantity of system information, which determines the place of the particular in the general. The equation of the amount of system information of a physical quantity has the form:

$$I(X) = \log_2 \frac{x}{\Delta x}. \quad (15)$$

where: x – discrete variable of the value X , Δx – sensitivity threshold.

2. When $\Delta x = f(IT)$ is a function of the tolerance on the accuracy of the parameter, then the system-information model acquires the characteristics of the information of the accuracy of the parameter. The higher the IT tolerance on the accuracy of the parameter, the more complex and costly the technology of its production. In this case, the system information characterizes the technological costs of the products.

3. When $\Delta x = f(PL)$ is a function of the Planck unit [12], and X_{up} is the sought variable, then the system-information model acquires the information characteristics of the Planck units of physical quantities. Since the values of the Planck units are derived from the fundamental physical constants, in this case the information characterizes the equilibrium of the system at the optimal numerical value of the objects' properties.

4. When $\Delta x = f(U)$ is a function of the interval of expanded uncertainty of the value of a physical quantity, then the system-information model acquires the characteristics of uncertainty when solving problems in intelligent systems. In this case the information characterizes the probability of quality of the system under conditions of uncertainty.

5. When $\Delta x = f(\mu_A(x))$, $x \in X$, where $\mu_A(x)$ – the membership function, then the system-information model acquires the characteristics of fuzzy information in an intelligent system.

6. When $\Delta x = f(p)$ is a function of probability. Such an information measure in the form of a ratio was used by K. Shannon in his formula for calculating the amount of information: $I(p) = \log_2 \frac{1}{p}$, $p = \frac{n}{m}$, $|I(p)| = \frac{1}{p} = \frac{m-0}{n-0}$; where: $X_{up} = m$, $X_{low} = 0$, $x = n$, $\Delta x = n-0$, $|I(p)|$ – informational measure, m – (total) number of attempts, n – (private) value of realized.

Thus, the characteristic of a particular kind of information to be formalized depends on the properties that are attributed to the elements of the information measure function.

In the transition from the abstract mathematical formulation to the physical interpretation of system information the set X is represented by the set of physical properties of objects that are identified as physical quantities and their values are estimated due measurement. The material physical world can be modeled on the basis of system information of space, time and physical properties of objects. Each of the attributes presented can be formalized as a function of the information measure. The union of these attributes represents the information space 5D.

The objects of systems-information research are various systems: physical, chemical, biological, social,

technical, informational, etc. For each of these systems, there is both a general and a private information measure of transformation and development. For each system-information model of the hierarchical level of a material object, methods for calculating quantitative values of information may be different. Numerical methods for determining the quantity, quality and value of information are based on the information measure.

The presented analysis of information characteristics is based on information interaction. The interaction of objects, leading to a change in knowledge of at least one of them, is called information interaction. Despite the nature of the definition of information interaction, it is useful for analyzing different definitions of information. Thus, the entropic approach describes information at the signal level, the algorithmic and algebraic approach – at the linguistic level, and the logical approach – at the semantic level.

For information interaction it is not enough only to transmit a message, it is necessary that the receiver (addressee) has a possibility to perceive it adequately. From this follows the thesaurus principle: the importance of having a priori information, sufficient to decode and assimilate the message received. This means, in particular, that participants in information interaction must have consistent information about the codes, languages and their semantics being used. This principle emphasizes the paramount importance for informatics of linguistic and semantic research in the broad sense of the word. First of all, we are talking about fragments of human communication languages and their semantics, which have become the basis for the development of human-machine dialogue tools today.

System-information models in production tasks

The product in production is characterized by the product life cycle (PLC), which includes the following stages: marketing, research and development, design, technological preparation of production, production, quality control, operation, maintenance and disposal. Each stage of the PLC is an open information system, in which the information comes from external sources, is processed and issued to the next stage. At each stage of PLC both deterministic and stochastic information processes between objects which possess the certain information necessary for performance of a concrete industrial task proceed.

On the basis of the system-information approach considered, four basic models of information technology can be distinguished for use in intelligent information processing systems for product manufacturing [13].

Absolute system-information model (ASIM)

The design of information technology uses absolute numerical values of system information of the parameters of design documentation (CD), technological system (TS) and technological process (TP) on the basis of deterministic and stochastic system-information models. ASIM defines: the numerical value of the quantity of system information of production objects parameters, which characterizes the complexity of production; quality of system information parameters, which characterizes the level of used technology; value of system information parameters, which characterizes the efficiency of production. These system information models are used in the analysis of the state of production in solving economic and managerial problems.

ASIM are used in the early stages of PLC in solving the problems of optimizing the redundancy of both product design and technological process based on the ratio of the amount of system information incorporated in the product parameters to the indicators of technical characteristics of the technological system and process, as well as to the parameters of the product service purpose.

Relative system-information model (RSIM)

RSIM characterizes the quality of production processes in the manufacture of the product and is equal to the coefficient of information connection ASIM and KPI (technical and economic indicators) of production:

$$RSIM_{Old} = (K_{inf}) = \frac{KPI_{Old}}{ASIM_{Old}}, \quad (16)$$

$$KPI_N = K_{inf} \times ASIM_N.$$

RSIM is used in the stages of PLC when solving problems of optimization and forecasting of production resources for the manufacture of the product: energy, time, material and others at the level of technological equipment, production area, shop and production as a whole.

Example 1. Conduct a preliminary prediction of the cost of resources: electrical energy C_{El}^N , the time

of manufacture C_T^N and the estimated cost of C_{PROD}^N , when launching a new (n) product in production

on the basis of existing enterprise technologies in the manufacture of an old (*old*) product:

$$K_{el} = \frac{C_{el}^{old}}{\sum_{i=1}^n I_{i(prod)}^{old}}, \quad K_T = \frac{T_{prod}^{old}}{\sum_{i=1}^n I_{i(prod)}^{old}}, \quad K_c = \frac{C_{prod}^{old}}{\sum_{i=1}^n I_{i(prod)}^{old}},$$

$$C_{el}^n \approx K_{el} \times \sum_{i=1}^n I_{i(prod)}^n, \quad T_{prod}^n \approx K_T \times \sum_{i=1}^n I_{i(prod)}^n, \quad C_{prod}^n \approx K_c \times \sum_{i=1}^n I_{i(prod)}^n. \quad (17)$$

where: K_{el} , K_T , K_c – coefficients of information communication for existing production technologies; *OLD* – old product, N – new product, I_i – system information of the product parameter according to the design documentation.

Forecasting tasks according to the given method to calculate resource costs for launching a new product have not been solved in production before. The method is simple and effective using computer technology. The traditional method of solving such problems requires a large expenditure of resources, both labor, time and financial, it consists in the development of preliminary (approximate) technological processes for the manufacture of product elements and their assembly with the calculation of economic indicators.

Equivalent system-information model (ESIM)

One of the conceptual features of the system-information approach is the possibility to represent system information of physical quantities on the basis of Planck units. The values of Planck units (l_{pl} – length, t_{pl} – time, m_{pl} – mass, E_{pl} – energy and other derived physical quantities are calculated on the basis of the fundamental physical constants C (speed of light, m/s);

\hbar (Dirac constant, $J \cdot s$); G (gravitational constant, $m^3/kg \cdot s^2$) [12]:

$$l_{pl} = \sqrt{\frac{G\hbar}{C^3}}m, \quad t_{pl} = \sqrt{\frac{G\hbar}{C^5}}s, \quad m_{pl} = \sqrt{\frac{Ch}{G}}kg, \quad e_{pl} = \sqrt{\frac{\hbar C^5}{G}}J.$$

Basic Planck units, as well as fundamental physical constants, have invariable values of relative uncertainties and have approximately the same order – 10^{10} [14]. The value of the information measure is the inverse of the relative uncertainties of the Planck units and their numerical values are equal to each other. Therefore in the system-information approach the value of Planck units are taken as a threshold of sensitivity of reference physical quantities. Thus, the value of the amount of system information of the reference physical quantity, calculated on the basis of the Planck unit, can be equivalently expressed through another physical quantity by means of the coefficient of agreement between them.

The universal ESIM parameter is the physical quantity energy, i.e. the quantity of system information of any physical quantity on the basis of the Planck unit can be equivalently expressed through the system information energy.

Example 2: Determine the E_{PL}^{eq} equivalent of the P power value through the energy equivalent using a matching factor $K_{match}(E_{PL}/P_{PL})$:

$$E_{pl} = \sqrt{\frac{\hbar C^5}{G}}, \quad LSI(E_{PL}) = \log_2\left(\frac{1}{E_{PL}}\right), \quad LSI(P_{PL}) = \log_2\left(\frac{1}{P_{PL}}\right),$$

$$P \left[kg \cdot \frac{m^2}{s^3} \right], \quad \log_2(P) = \log_2(m) + 2\log_2(l) - 3\log_2(t),$$

$$\log_2\left(\frac{1}{P_{PL}}\right) = \log_2\left(\frac{1}{m_{PL}}\right) + 2\log_2\left(\frac{1}{l_{PL}}\right) - 3\log_2\left(\frac{1}{t_{PL}}\right), \quad (18)$$

$$P_{pl} = \frac{1}{2^{(\log_2(1/m_{pl}) + 2\log_2(1/l_{pl}) - 3\log_2(1/t_{pl}))}}, \quad K_{match}\left(\frac{E_{PL}}{P_{PL}}\right) = \frac{LSI(E_{PL})}{LSI(P_{PL})},$$

$$\log_2 \frac{E_P^{eq}}{E_{PL}} = K_{match}\left(\frac{E_{PL}}{P_{PL}}\right) \times \log_2 \frac{P}{P_{PL}}; \quad E_P^{eq} = E_{PL} \times 2^{K_{match}(E_{PL}/P_{PL}) \times \log_2(P/P_{PL})}.$$

where: LSI – the logarithmic single index of the Plank unit of physical value; E_{pl} , P_{pl} – Plank units of energy and power; E_p^{eq} – equivalent of the power value in energy units E_{pl} ; P – power, p – power symbol in the equivalent of energy E_p^{eq} ; $K_{match}(E_{pl}/P_{pl})$ – matching factor.

Problems solved on the basis of the ESIM method refer to engineering problems, both in science and in production.

Mixed system-information model (MSIM)

System-information models MSIM represent a combination of the above models ASIM, RSIM, ESIM used for intelligent information processing in solving a wide range of problems. Economic and social tasks are some of the complex tasks that are currently being solved using intelligent information processing systems. Research of information processes of these systems and development of information models for solving problems are labor- and resource-intensive.

When analyzing complex information systems (e.g., economic or social), MSIMs with limited information resources are used based on the use of the three system-information models presented above. The information elements of the system are modeled on the basis of ASIM, the information process in the system is modeled on the basis of RSIM, the system information for analysis is presented by ESIM. The use of MSIM allows you to get the most objective result of the economic task of determining the effectiveness of processes and systems, which is calculated on the basis of the ratio of system information result to the costs of the work done.

In the theory of economics, the task of determining the efficiency of systems in a strict plan is currently not solved [15], because the result and cost of work indicators use different qualitative and quantitative attributes, measured by different units on different scales. ESIM eliminates this contradiction, it allows to evaluate the effectiveness of the system, representing the values of the result and cost of work in the form of system information equivalent to the energy measured in unified units of bits (bits).

Application of system-information models for intelligent information processing in production tasks

System-information models of production processes and systems are used to solve production problems. System-information models for intelligent information

processing, provide the definition of optimal ways in solving the problems of resource planning and building a strategy for achieving the goal of production.

To use system-information models for solving various practical problems, different models based on the use of certain methods of intelligent information processing can be applied, the main ones are [16].

- 1) artificial neural networks (NN), the advantage of which is the ability to represent some limited continuous function with any small approximation error;
- 2) fuzzy logic, which allows inputs or variables in a decision problem to be represented as people reason about them;
- 3) expert systems, used to describe a problem and using the intelligence of one or more identified experts;
- 4) evolutionary computing, characterized by the ability to adapt to the environment by modeling the emergence, survival, and improvement of a population of individuals;
- 5) multi-agent systems, consisting of groups of agents with different goals and tasks, with certain characteristics and which are an active area of research in complex applications.

For the analysis of system-information models as objects of information interaction we will use both neural network approach and fuzzy logic, which are effective tools for solving a wide range of problems of intelligent data analysis, which can include and production tasks.

Let us consider some characteristics of functional dependences of technical and economic production indicators to the numerical values of system information (I) of product parameters DD (design documentation), TS (technological system) and TP (technological process). Timely and justified assessment of production indicators on the basis of information technology provides effective forecasting and management of the state of production.

1. Any mode of production is displayed through the system information (I) of the monitored parameters of the production objects.
2. The parameters of the product, technological process and system have the final numerical value of the system information (I) and are the most objective characteristics of production.
3. Quantitative indicators of resource costs of material, labor and energy to produce the product is a function of the numerical value of the system information (I) of the parameters of DD, TS and TP:

$$\begin{aligned} \text{Resource costs}(\text{material, labor, energy}) &= \\ &= f(I(\text{DesDoc}); I(\text{TechSist}); I(\text{TechProc})). \end{aligned} \quad (19)$$

4. Technical and economic indicators of production (KPI) is a function of redundancy or lack of system information of DD parameters of the product, TS,

$$KPI = f(I_{excess}(DesDoc, TechSist, TechProc), I_{deficiency}(DesDoc, TechSist, TechProc)). \quad (20)$$

5. Production efficiency is characterized as a function of the ratio of transferred system information by the technological system and process (costs) to the received system information by the product (result) (characteristic of the value of system information in the production of the product):

$$Production\ efficiency = f(I_{TechProc} / I_{DesDoc}). \quad (21)$$

The presented system-information models and ways of their use can be applied in systems of intelligent information processing to solve a wide class of practical problems.

Conclusion

The development of methods for solving various classes of practical problems using intelligent information processing is one of the key research areas in computer science. The developed system-information models of processes and systems for intelligent information processing allow analyzing and solving production problems. Thereby increase the efficiency of solving problems of analysis, synthesis and forecasting of production systems and technologies, as well as problems of production management.

The system-information approach to processes and systems introduced new concepts of system information and information measure, which allowed to develop system-information models for intelligent information processing.

The concept of systemic information corresponds to the generally accepted classical definition of "information". The information process is based on the information interaction of system elements, which leads to a change in their state and is caused by a positive or negative increase in the amount of system information.

References

1. Cattell, R. B. (1971), *Abilities: Their structure, growth, and action*, New York, Houghton Mifflin, 312 p.
2. Kazakhstan (2005), *National Encyclopedia*, Almaty, Vol. 2, No. 2, 420 p.
3. Lutsky, S. (2008), *Theoretical foundations of the system-information approach to technological processes and systems (monograph)*, 238 p.
4. Lutskyy, S. V. (2021), "System-information approach to uncertainty of process and system parameters", *Innovative technologies and scientific solutions for industries*, No. 3 (17), P. 91–106. DOI: <https://doi.org/10.30837/ITSSI.2021.17.091>
5. Chi Leung, Patrick Hui (2011), *Application of Artificial Neural Networks and Hybrid Methods in the Solution of Inverse Problems*, 586 p.

TP (characteristic of the quality of system information in the production process):

The paper proposes:

1. The concept of a term of information as the value of a number equal to the proportional ratio of the total value of the attribute to its private value.

2. In the information process, the proportion (number) of the ratio of the general to the particular is to be transmitted.

3. The notion of the type of information is given, which is determined by the type of sensitivity threshold of the attribute that possesses the information (examples are given).

4. System-information models are used to obtain knowledge of any processes or systems beyond previous experience. That is, by getting information (number) of different data of design documentation parameters, technological documentation, technological process we determine the new knowledge, characterizing the production.

System-information models of processes and systems describe the interaction between source and receiver at the information level based on the sensitivity threshold. The communication channel between the source and the receiver of information operates, as a rule, under conditions of uncertainty, which can lead to loss of information during transmission due to possible changes in system characteristics. To describe their interaction, some models of intelligent information processing can be used, in particular, neural network models or fuzzy inference models. Their use will improve the efficiency of receiver state prediction, taking into account the state of the transmitter and the conditions of communication channel operation.

The presented article has shown the relevance of developing system-information models for intelligent information processing at the levels of data reception, interpretation and communication, which allows to expand the class of solved production tasks.

6. Komartsova, L. G., Maksimov, A.V. (2012), *Neurocomputers*, 157 p.
7. Larichev, O. I. (2008), *Theory and methods of decision making*, 290 p.
8. Merkert, Mueller, Hubl (2015), *A Survey of the Application of Machine Learning in Decision Support Systems*, University of Hoffenheim, 186 p.
9. Sanchez i Marre, Gibert (2012), *Evolution of Decision Support Systems*, University of Catalunya, 136 p.
10. Ruzhentsev, I. V., Lutsky, S. V. (2017), "Discrete probabilistic information laws factor of efficiency", *Ukrainian Metrological Journal*, No. 1, P. 67–71.
11. Egorov, A. A. (2001), "Application for discovery", *International Association for Scientific Discovery*, No. A-242, P. 36–52.
12. Planck, M. (1899), *Proceedings of the Royal Prussian Academy of Sciences in Berlin*, No. 5, P. 440–480.
13. Tomilin, K. A. (2002), "Planck values – 100 years of quantum theory. History. Physics. Philosophy", *Proceedings of the international conference*, Moscow, NIA-Nature, P. 105–113.
14. Nogin, V. D. (2020), *The set and the Pareto principle*. St. Petersburg: Publishing and Printing Association of Higher Educational Institutions, 100 p.
15. Stuart, J. Russell, Peter Norvig (2020), *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition, 1136 p.

Received 19.09.2022

Відомості про авторів / Сведения об авторах / About the Authors

Корабльов Микола Михайлович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри комп'ютерних інтелектуальних технологій і систем, Харків, Україна; e-mail: mykola.korablyov@nure.ua; ORCID ID: <https://orcid.org/0000-0002-8931-4350>

Кораблев Николай Михайлович – доктор технических наук, профессор, Харьковский национальный университет радиоэлектроники, профессор кафедры компьютерных интеллектуальных технологий и систем, Харьков, Украина.

Korablyov Mykola Mykhailovych – Doctor of Technical Science, Professor of the Department of Computer Intelligent Technologies and Systems, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Луцький Сергій Володимирович – кандидат технічних наук, Харківський національний університет радіоелектроніки, старший викладач кафедри метрології та технічної експертизи, Харків, Україна; e-mail: lutsk.sv6@gmail.com; ORCID: <https://orcid.org/0000-0002-5327-6591>

Луцкий Сергей Владимирович – кандидат технических наук, Харьковский национальный университет радиоэлектроники, старший преподаватель кафедры метрологии и технической экспертизы, Харьков, Украина.

Lutskyu Sergey – PhD (Engineering Sciences), Kharkiv National University of Radio Electronics, Senior Lecturer of the Department of Metrology and Technical Expertise, Kharkiv, Ukraine.

СИСТЕМНО-ІНФОРМАЦІЙНІ МОДЕЛІ ДЛЯ ІНТЕЛЕКТУАЛЬНОГО ОБРОБЛЕННЯ ІНФОРМАЦІЇ

Предметом дослідження є системно-інформаційні моделі процесів і систем та їхнє використання для інтелектуального оброблення інформації у виробничих завданнях. Застосування інтелектуального оброблення інформації в системах управління виробництвом є нині одним із ключових напрямів розвитку інформатики. **Мета роботи** – розроблення системно-інформаційних моделей процесів і систем для інтелектуального оброблення інформації, що дають змогу аналізувати й виконувати виробничі завдання, які виходять за межі попереднього досвіду в умовах невизначеності. У статті вирішуються такі **завдання**: аналіз підходів до визначення характеристик інформації процесів та систем; розроблення основ моделювання системно-інформаційних процесів і систем для інтелектуального оброблення інформації; створення системно-інформаційних моделей і способів їхнього застосування для інтелектуального оброблення інформації в завданнях виробництва. Використовуються такі **методи**: системно-інформаційний підхід до процесів та систем; системно-інформаційне моделювання процесів і систем. Отримано такі **результати**: проаналізовано підходи до визначення характеристик інформації процесів і систем; розроблено основи моделювання системно-інформаційних процесів і систем для інтелектуального оброблення інформації; запроваджено поняття системної інформації та інформаційної міри; розроблено системно-інформаційні моделі та способи їхнього застосування для інтелектуального оброблення інформації в завданнях виробництва. **Висновки**. Розроблення методів для вирішення різного класу практичних завдань із використанням інтелектуального оброблення інформації є одним із ключових напрямів досліджень в інформатиці. Створені системно-інформаційні моделі процесів і систем для інтелектуального оброблення інформації дають змогу аналізувати й вирішувати завдання, що виходять за межі попереднього досвіду. Тим самим вони підвищують ефективність вирішення завдань аналізу, синтезу та прогнозування виробничих систем і технологій, а також завдань управління виробництвом. Системно-інформаційний підхід до процесів і систем оперує новими поняттями – "системна інформація" та "інформаційна міра".

Завдяки цьому розроблено системно-інформаційні моделі для інтелектуального оброблення інформації, а також способи їхнього застосування на етапах життєвого циклу виробу, що дало змогу вирішувати завдання виробництва. Системно-інформаційні моделі процесів та систем описують взаємодію джерела та приймача на інформаційному рівні на основі порога чутливості. Канал зв'язку між джерелом та приймачем інформації працює, як правило, в умовах невизначеності, що може спричинити втрату інформації під час передачі внаслідок можливої зміни характеристик системи. Для опису їхньої взаємодії можуть бути використані ті чи інші моделі інтелектуального оброблення інформації, зокрема нейромережні моделі або моделі нечіткого виведення. Їхнє використання дасть змогу підвищити ефективність прогнозування стану приймача з урахуванням стану передавача та умов функціонування каналу зв'язку. Ця стаття показала актуальність розроблення системно-інформаційних моделей для інтелектуального оброблення інформації на рівнях рецепції, інтерпретації та комунікації даних, що дозволяє розширити клас виробничих завдань, які вирішуються.

Ключові слова: системно-інформаційні моделі; системна інформація; інформаційна міра; поріг чутливості; інтелектуальне оброблення інформації.

СИСТЕМНО-ИНФОРМАЦИОННЫЕ МОДЕЛИ ДЛЯ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ ИНФОРМАЦИИ

Предметом исследования являются системно-информационные модели процессов и систем и их использование для интеллектуальной обработки информации в производственных задачах. Использование интеллектуальной обработки информации в системах управления производством является в настоящее время одним из ключевых направлений развития информатики. **Цель** работы – разработка системно-информационных моделей процессов и систем для интеллектуальной обработки информации, позволяющих анализировать и решать производственные задачи в условиях неопределенности. В статье решаются следующие **задачи**: проведение анализа подходов к определению характеристик информации процессов и систем; разработка основ моделирования системно-информационных процессов и систем для интеллектуальной обработки информации; разработка системно-информационных моделей и способов их применения для интеллектуальной обработки информации в задачах производства. Используются следующие **методы**: системно-информационный подход к процессам и системам; системно-информационное моделирование процессов и систем. Получены следующие **результаты**: проведен анализ подходов к определению характеристик информации процессов и систем; разработаны принципы моделирования системно-информационных процессов и систем для интеллектуальной обработки информации; введены понятия системной информации и информационной меры; разработаны системно-информационные модели и способы их применения для интеллектуальной обработки информации в задачах производства. **Выводы.** Разработка методов для решения различного класса практических задач с использованием интеллектуальной обработки информации является одним из ключевых направлений исследований в информатике. Разработанные системно-информационные модели процессов и систем для интеллектуальной обработки информации позволяют анализировать и решать задачи, которые выходят за рамки предыдущего опыта. Тем самым они повышают эффективность решения задач анализа, синтеза и прогнозирования производственных систем и технологий, а также задач управления производством. Системно-информационный подход к процессам и системам оперирует новыми понятиями – "системная информация" и "информационная мера". Это позволило разработать системно-информационные модели для интеллектуальной обработки информации, а также способы их применения на этапах жизненного цикла изделия, что помогло решать задачи производства. Системно-информационные модели процессов и систем описывают взаимодействие источника и приемника на информационном уровне на основе порога чувствительности. Канал связи между источником и приемником информации работает, как правило, в условиях неопределенности, что может приводить к потере информации при передаче вследствие возможного изменения характеристик системы. Для описания их взаимодействия могут быть использованы те или иные модели интеллектуальной обработки информации, в частности нейросетевые модели или модели нечеткого вывода. Их использование позволит повысить эффективность прогнозирования состояния приемника с учетом состояния передатчика и условий функционирования канала связи. Представленная статья показала актуальность разработки системно-информационных моделей для интеллектуальной обработки информации на уровнях рецепции, интерпретации и коммуникации данных, что позволяет расширить класс решаемых производственных задач.

Ключевые слова: системно-информационные модели; системная информация; информационная мера; порог чувствительности; интеллектуальная обработка информации.

Бібліографічні описи / Bibliographic descriptions

Кораблев М. М., Луцкий С. В. Системно-інформаційні моделі для інтелектуального оброблення інформації. *Сучасний стан наукових досліджень та технологій в промисловості*. 2022. № 3 (21). С. 26–38. DOI: <https://doi.org/10.30837/ITSSI.2022.21.026>

Korablyov, M., Lutskyu, S. (2022), "System-information models for intelligent information processing", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (21), P. 26–38. DOI: <https://doi.org/10.30837/ITSSI.2022.21.026>

O. SOLOVEI

NEW ORGANIZATION PROCESS OF FEATURE SELECTION BY FILTER WITH CORRELATION-BASED FEATURES SELECTION METHOD

The **subject** of the article is feature selection techniques that are used on data preprocessing step before building machine learning models. In this paper the focus is put on a Filter technique when it uses Correlation-based Feature Selection (further CFS) with symmetrical uncertainty method (further CFS-SU) or CFS with Pearson Correlation (further CFS-PearCorr). The **goal** of the work is to increase the efficiency of feature selection by Filter with CFS by proposing a new organization process of feature selection. The **tasks** which are solved in the article: review and analysis of the existing organization process of feature selections by Filter with CFS; identify the routs cause the performance degradation; propose a new approach; evaluate the proposed approach. To implement the specified tasks, the following **methods** were used: information theory, process theory, algorithm theory, statistics theory, sampling techniques, data modeling theory, science experiments. **Results.** Based on the received results are proved: 1) the chosen features subset's evaluation function couldn't be based only on CFS merit as it causes a learning algorithm's results degradation; 2) the accuracies of the classification learning algorithms had improved and the values of determination coefficient of the regression learning algorithms had increased when features are selected according to the proposed new organization process. **Conclusions.** A new organization process for feature selection which is proposed in current work combines filter and learning algorithm properties in evaluation strategy which helps to choose the optimal feature subset for predefined learning algorithm. The computation complexity of the proposed approach to feature selection doesn't depend on dataset's dimensions which makes it robust to different data varieties; it eliminates the time needed for feature subsets' search as subsets are selected randomly. The conducted experiments proved that the performance of the classification and regression learning algorithms with features selected according to the new flow had outperformed the performance of the same learning algorithms built with without applied new process on data preprocessing step.

Keywords: Correlation-based Feature Selection (CFS); symmetrical uncertainty (SU); Pearson Correlation (PearCorr); merit; accuracy; determination coefficient.

Introduction

When data which is gathered for pattern recognitions or machine learning models includes a lot of observations and features then it became difficult to perform effective data visualization; data mining or to build a machine learning model with high accuracy and performance. Therefore, sampling and feature selections methods are developed to cope with high-dimensional datasets [1].

Feature selection is a widely used instrument to remove irrelevant and redundant information from the dataset to avoid overfitting and reduce memory usage and computational costs. The goal of feature selection is to choose an optimal feature subset according to predefined evaluation criterion [2]. The recent trend to have a small number of samples in dataset versus a lot of features may cause problems to machine learning algorithm regarding learning performance therefore feature selection process plays increasingly import role while building machine learning model.

Analysis of the current state of the problem and methods of its solution

The techniques to select features for machine learning model are specified as: wrappers, embedded, filters, dimensionality reduction and hybrid [3]. A different Fast Correlation-Based Feature Selection (FCBFS) algorithm for filter had been considered in the study [4]. It proposes to use a threshold value δ , which is identified by user e.g. for dataset with N features and class C when $merit_{i,c}$ measures the correlation between a feature F_i and the class C then F_i is added in subset if $\forall F_i \in S', 1 < i \leq N, merit_{i,c} > \delta$. Formed in such way subset is processed the 2^{nd} time in order to retain only predominant feature. After one round of filtering features, algorithm takes the remaining features as a new subset and repeat starting to add a new feature. The algorithm stops when there is no feature to be removed. The worst case could be none features are removed. FCBF's performance for ten datasets had been compared with wrapper for two learning algorithms C4.5 and naive Bayes and

because FCBF improved the accuracy of both learning algorithms FCBF was concluded as practical for feature selection for classification of high dimensional data.

The idea to use Hybrid approach according to which evaluation strategy uses a filter method and learning algorithm had been considered in several studies [5–7]. In work [5], it was proposed to form feature subsets using FBS with CFS and then to use dominance-based rough set approach (DBRSA) to select the final feature set. As DBRSA is an extension of the classical rough set approach (CRSA) which utilizes a decision tree, so the final feature subset is selected by the means of the learning algorithm.

Adaptive Hybrid Feature Selection methodology (AHFS) is proposed in work [6]. It utilizes the fact that there is no «best of» metric/method to select a feature subset and the choice of the metric/method can be realized by using the applied learning algorithm. AHFS uses SFS to form a subset in each iteration and iterates through possible evaluation methods to assign feature subset a set of ranks corresponding to the method. The final feature subset is selected by artificial neural networks model.

Hybrid feature selection that significantly reduces dimensionality of features was proposed in work [7]. The approach uses the combination of ReliefF and Principle Component Analysis (PCA) algorithms which are applied in the following sequence: 1) features are weighted by ReliefF and a candidate feature subsets are formed from features which weights exceed threshold; 2) PCA is applied on the candidate feature subset to reduce the dimension.

Highlight of the earlier unresolved parts of the general problem.

Aim of the study

FCBF algorithm proposes to solve the first suggested in this paper problem by adding to evaluation strategy of feature selection a hyper parameter threshold δ , which value to be decided by user. It makes an algorithm dependent on the dataset and considering possible data variety the chosen value of δ may not be optimal.

Hybrid approaches [5–7] have a common idea that evaluation strategy couldn't be based on CFS method only but requires additional evaluation criteria which is a learning algorithm. However, in studies [5–7] are left not considered the dependencies between

a predefined learning algorithm and selected by evaluation strategy a final feature subset.

This paper proposes to change the organization process of feature selection by Filter with CFS that will take into consideration the dependencies between a predefined learning algorithm and chosen feature subset. The experiment tests of the new organization process will be performed for Filter with CFS-SU method – when dataset has features with discrete values and for Filter with CFS and Filter with CFS-PearCorr method – when dataset has features with discrete values.

Materials and methods

All Feature selection techniques have four steps in common:

1. Starting point – selects the feature from which to begin the feature selection.

2. Search organization – specify the algorithm for feature subset identification. Covers a type of search: exhaustive; complete; random and heuristic. A search can start by adding a new feature to an initially empty set and then a feature subset is expanded with one additional variable in each iteration step is called Sequential Feature Selection (SFS) or add all features and start removing irrelevant or redundant features (backward elimination) or Best First Selection (BFS) – the search is started with the most predictive feature according to chosen metric and then in each iteration step – the most predictive subset is expanded with a feature.

3. Evaluation strategy – specifies how a goodness of feature subset to be evaluated. It can be independent of the machine learning algorithm (common for filter technique) or by performance metrics of the learning algorithm (common for wrappers).

4. Stopping criterion – a rule to decide when to stop searching the feature subsets.

The current study is focused on filter and wrapper techniques, therefore the details for other techniques are omitted on purpose.

Wrapper method is aimed to select the feature subset that will ultimately provide a better estimate of accuracy. Wrapper uses a predefined machine learning algorithm to evaluate the quality of a selected feature subset. In forward selection, it calculates the accuracy of adding a new unselected feature to the subset and according to received accuracy decides to keep or remove the feature. Wrapper method produces good feature subsets because estimated performance of the learning

algorithm is the best heuristic for measuring the goodness of feature subsets, however it computationally inefficient and it does not scale well to large datasets.

Filter – independent of any learning algorithm because its evaluation strategy is based on different statistical measures. When filter uses one of the methods: Low Variance [8], Fisher Score [9], T-score [10], distance measure [11], Chi-square [12], information gain [13], Gini index [14] and others [15] then the features are ranked and a selection strategy, e.g. "select best n_{features} " is applied to extract the final set of features to be used. Filter methods had been proved as fast and effective while capturing the relevance of features to the target, therefore filter methods are chosen when dataset has a big number of samples. However, filters with the mentioned methods cannot discover redundancy among features whereas redundant features along with irrelevant feature negatively affect the speed and accuracy of learning algorithm [16]. Another point, which makes filters technique weaker is that the features are ranked with no consideration for learning algorithm but different algorithms may perform better or worse for the same feature subsets [17]. In wok [18] was introduced a new Filter with CFS method it returns the set of features from which simultaneously are removed both irrelevant and redundant features. CFS finds feature subset that is useful to predict the target variable and do not strongly interact with other features. The rule is formalized as specified in equation 1

$$CFS_score(S) = \frac{k \overline{r_{cf}}}{\sqrt{k + k(k-1) \overline{r_{ff}}}}, \quad (1)$$

where CFS score is an evaluation score (further «merit») of subset S with k features; $\overline{r_{cf}}$ – the average of the correlations between feature subset and target variable; $\overline{r_{ff}}$ – the average feature-feature correlation. In eq. 1 the numerator indicates the predictive power of the feature set while the denominator shows how much redundancy the feature set has. Correlations $\overline{r_{cf}}$ and $\overline{r_{ff}}$ can be calculated using symmetrical uncertainty formula when features in dataset have discrete values; or Pearson's correlation, for features with continuous values. When Filter uses CFS_PearCorr then $\overline{r_{cf}}$ from eq. 1 is the average of absolute values of Pearson correlations between feature subset and target variable; $\overline{r_{ff}}$ – the average of absolute values of feature-feature

Pearson correlation. As a result of conducted experiments of features selection based on CFS score (eq. 1) in study [18] had been concluded that while solving a classification problem with discrete features dataset – CFS-SU can be used as a standard. The effectiveness of CFS-SU was evaluated by comparing it with wrapper which uses predefined learning algorithms: naive Bays, C4.5. The Filter with CFS-SU was organized as:

1. Starting point – from the 1st feature add/remove features one by one.

2. Search organization – features are added to subset with SFS technique until merit shows increasing value five time in sequence. The process is repeated 50 times. The resulted subsets are ordered by merit in descending order.

3. Evaluation strategy – consists of steps: 1) to merge 1st and 2nd best subsets; 2) to calculate the merit of the new composed subset; 3) if the new merit is within 10% of the merit of the best subset then accept the new best subset; 4) form a new subset by merging the best subset with next not used subset from the list with 50 subsets.

4. Stopping criterion – repeats steps 2–4 until the condition in step 3 isn't met.

The following results from the comparison had been shared: accuracy of naive Bayes with feature subset selected by CFS-SU had shown a degradation for one dataset: audiology (au) from 80.24% to 75.55%. Accuracy of C4.5 had shown a degradation for five datasets: mushroom (mu) from 99.59% to 99.37%, audiology (au) from 78.48% to 77.14%, soybean (sb) from 89.16 to 86.80, horse-colic(hc) from 84.02 to 78.79, king-rook vs. king-pawn (kr-vs-kp) from 99.16 to 94.13.

Potentially, there are several problems caused the accuracy's degradation:

1. SFS with CFS-SU merit as evaluation technique doesn't return the feature subset which is good for predefined learning algorithm.

2. Merging approach used for final feature subset works well for naive Bayes but caused the performance degradation for C4.5 because added redundancies to the final subset and algorithm C4.5 isn't efficient when correlated feature are included. This fact means that an evaluation of "goodness" of feature subset is impacted by the chosen learning algorithm but that impact wasn't taken into consideration.

Further, in this paper, those assumptions will be verified.

Study results and their discussion

The characteristics of datasets for which a performance degradation had been captured in study [18] are specified in table 1. The number of observations ($n_{\text{observations}}$) and number of features (n_{features}) are different compared to raw datasets

Table 1. The characteristics of datasets with discrete and categorical features

Datasets	n_observations	n_features	Machine learning task	Feature Type			Missing values, Y/N?
				Continuous	Discrete	Categorical	
audiology	194	39	Multi classification			+	Y
horse-colic	299	12	Binary classification		+	+	Y
mushroom	8124	96	Binary classification			+	Y
king-rook vs. king-pawn	3196	35	Binary classification			+	N
soybean	306	35	Multi classification		+	+	Y

The learning algorithms: 1) naive Bayes was chosen in study [18] because, its classification accuracy is negatively affected by present of the redundant features as break the assumption that feature values are independent given the class; 2) C4.5 – is a popular learning algorithm for solving a classification task therefore an accuracy of C4.5 obtained as a result of experiments with feature selection approaches can be considered as a benchmark. In the current study we will continue to use those learning algorithms to compare the results and evaluate the correctness of the new proposals.

To study the result of feature evaluation strategy based on CFS-SU merit, we will visualize and analyze relationships between 1) accuracy of learning algorithm and CSF-SU merit; 2) number of features included in subset vs CSF-SU merit. To visualize the relationship

from UCI Machine Learning Repository due to performed features engineering: 1) "one hot encoding" and "label encoding" – in order to adapt categorical values to CFS-SU method; 2) data clean up – observations with missing values were removed in case the majority of feature's values are missing.

line for each dataset from table 1 we do: 1) randomly select N feature subsets S'_i ; 2) calculate CSF-SU merit and accuracy of learning algorithms for each S'_i , $1 \leq i < N$.

The results are presented on fig. 1 – fig. 5.

Fig. 1 – on picture (a) the highest accuracy 83.67% corresponds to merit 0.145, calculated with 27 features (picture (c)) and the highest merit 0.181 corresponds to lower accuracy 79.59%. On picture (b) the highest accuracy 81.63% corresponds to merit 0.119, calculated for 27 features and the highest merit 0.18 corresponds to lower accuracy 73.47%. Therefore, for audiology dataset, the height merit of S'_i doesn't correspond to the best accuracy of both learning algorithms; the number of features can be reduced from 39 to 27 for naive Bayes and C4.5 algorithms.

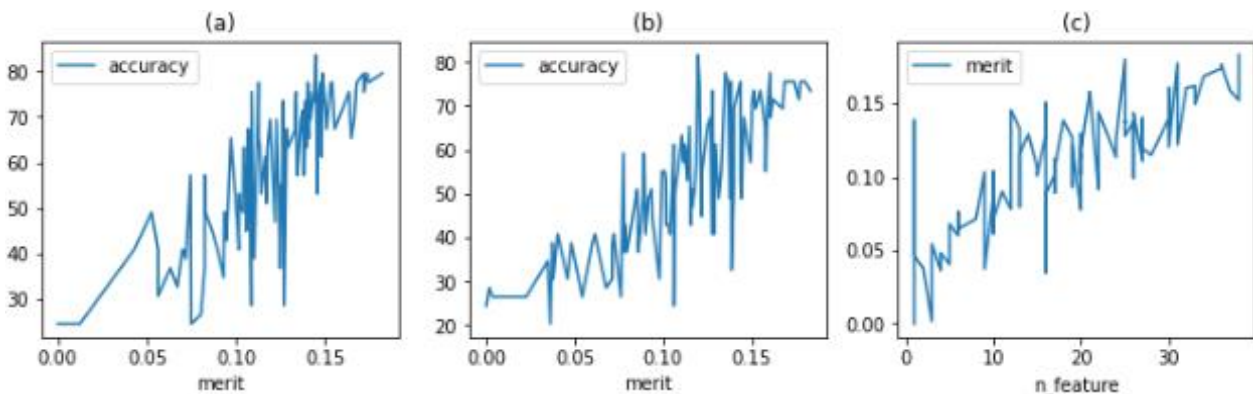


Fig. 1. (a) – accuracy of naive Bayes vs merit of au dataset with S'_i and target variable; (b) – accuracy of C4.5 vs merit of au dataset with S'_i and target variable; (c) – merit vs number of features in S'_i for au dataset

Fig. 2 – on picture (a) the highest accuracy 82% corresponds to merits 0.016 calculated with 6 features (picture (c)) and the highest merit 0.03 corresponds to lower accuracy 60%. On picture (b) the highest accuracy 85.28% corresponds to merit 0.023, calculated for 2 features (picture (c)) and the highest merit 0.027 corresponds to lower accuracy 76.58%. Therefore,

for horse-colic dataset, the highest merit of S'_i doesn't correspond to the best accuracy of both learning algorithms and the number of features can be reduced from 12 to 6 for naïve Bayes and from 12 to 2 for C4.5.

Fig. 3 – on picture (a) the highest accuracy 99.7% corresponds to merit 0.027, calculated with 62 selected features (picture (c)) and the highest

merit 0.03 corresponds to lower accuracy 98.13%. On picture (b) the highest accuracy 100% the 1st time corresponds to merit 0.021, calculated for 52 features and that level of accuracy is kept with further growing merits. Therefore, for mushroom dataset, the height merit of S_i doesn't correspond to the best accuracy of both learning the number of features can be reduced from 96 to 62 for naïve Bayes and from 96 to 52 for C4.5.

Fig. 4 – on picture (a) the highest accuracy 93.24% corresponds to merit 0.008, calculated with 25 selected

features (picture (c)) and the highest merit 0.014 corresponds to lower accuracy 67.83%. On picture (b) the highest accuracy 99.62% corresponds to merit 0.009, calculated with 35 features and the highest merit 0.01 corresponds to lower accuracy 90.99%. Therefore, for kr-vs-kp dataset, the height merit of S_i doesn't correspond to the best accuracy of both learning algorithms, the number of features can be reduced from 35 to 25 for naïve Bayes algorithm and no feature reduction is expected for C4.5.

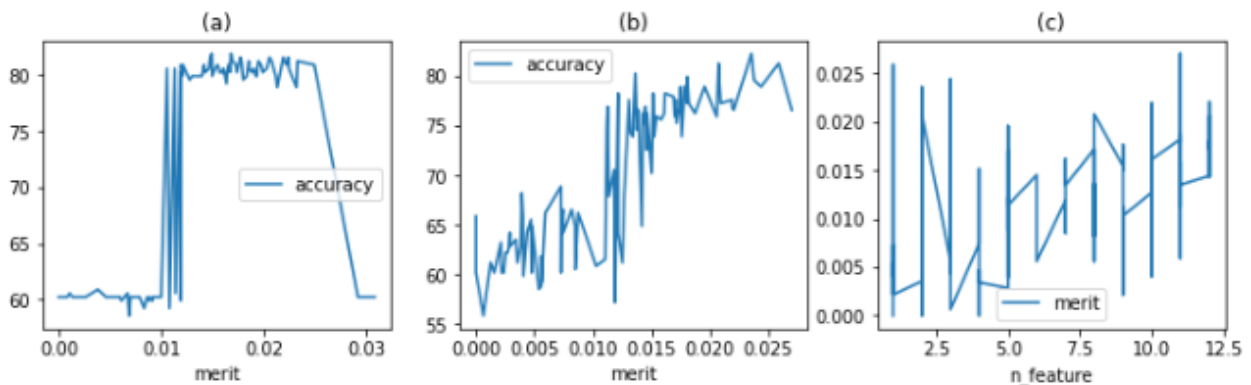


Fig. 2. (a) – accuracy of naïve Bayes vs merit of hc dataset with S_i and target variable; (b) – accuracy of C4.5 vs merit of hc dataset with S_i and target variable; (c) – merit vs number of features in S_i for hc dataset

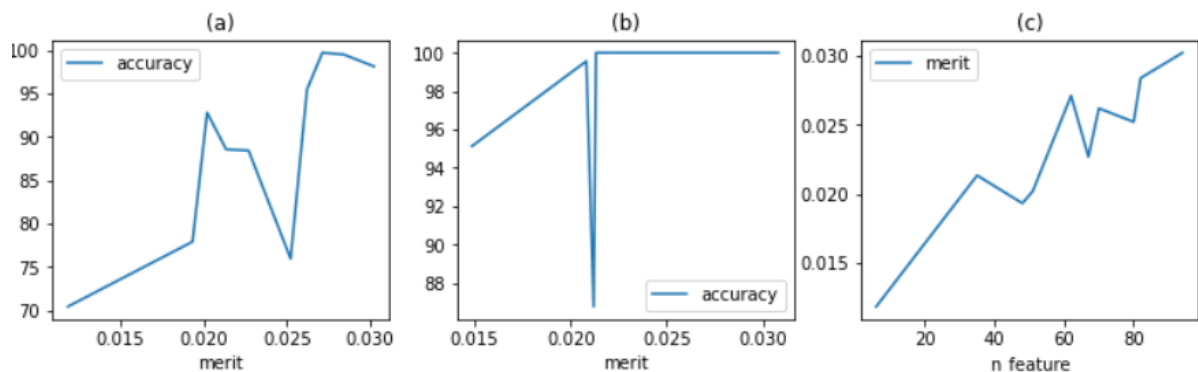


Fig. 3. (a) – accuracy of naïve Bayes vs merit of mu dataset with S_i and target variable; (b) – accuracy of C4.5 vs merit of mu dataset with S_i and target variable; (c) – merit vs number of features in S_i for mu dataset

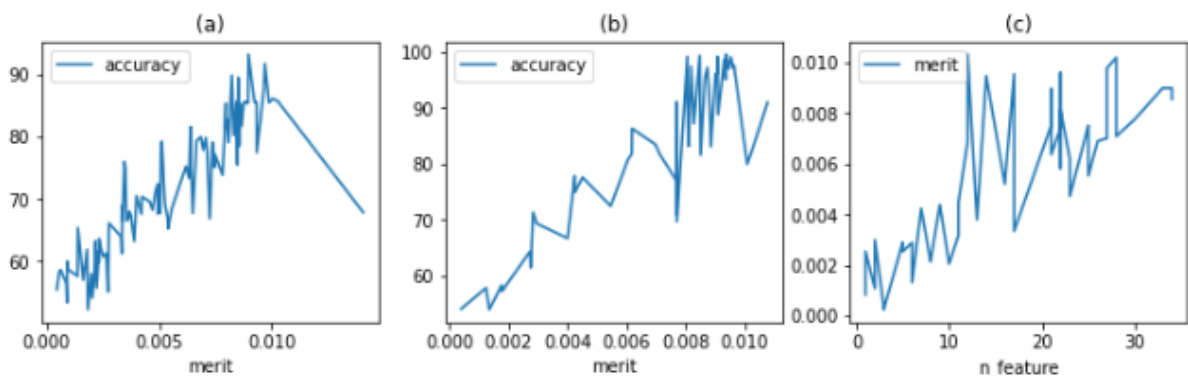


Fig. 4. (a) – accuracy of naïve Bayes vs merit of kr-vs-kp dataset with S_i and target variable; (b) – accuracy of C4.5 vs merit of kr-vs-kp dataset with S_i and target variable; (c) – merit vs number of features in S_i for kr-vs-kp dataset

Fig. 5 – on picture (a) the highest accuracy 93.51% corresponds to merit 0.255, calculated with 23 selected features (picture (c)) and the highest merit 0.3 corresponds to lower accuracy 88.31%. On picture (b) the highest accuracy 89.61% corresponds to merit 0.281, calculated with 27 features. Therefore, for soybean dataset, the height merit of S'_i doesn't correspond to the best accuracy of both learning algorithms, the number of features can be reduced from 35 to 23 for naïve Bays and from 35 to 27 for C4.5.

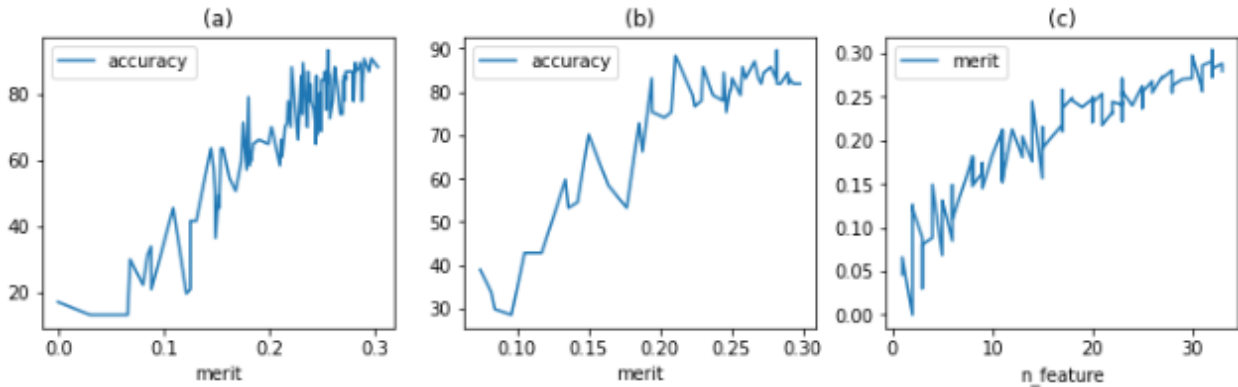


Fig. 5. (a) – accuracy of naive Bayes vs merit of soybean dataset with S'_i and target variable; (b) – accuracy of C4.5 vs merit of soybean dataset with S'_i and target variable; (c) – merit vs number of features in S'_i for soybean dataset

To tackle the identified problems, the following improvements in organization of feature selection process by Filter with CFS method are proposed:

1. Starting point: randomly select N feature subsets S'_i ; calculate CFS merit (M_i) for each S'_i , $1 \leq i < N$; sort a vector with S'_i by M_i in ascending order.

Note: As a result of the sorting the higher accuracy will likely correspond to subset from the 2nd part of a vector $\overline{S'}$.

2. Search strategy: is not required.

3. Evaluation strategy: calculate accuracy $A_{S'_i}$ of learning algorithm with subset S'_i , where $i = N/2$ and calculate accuracy $A_{S'_N}$ of learning algorithm with subset S'_N .

If $A_{S'_i}$ is less than $A_{S'_N}$ then save N ; increment i by $(N-i)divstep$; ENDIF

IF $A_{S'_i}$ is higher than $A_{S'_N}$ then save i ; decrement N by 1; ENDIF

IF $A_{S'_i}$ is equal to $A_{S'_N}$ then save i ; increment i by 1; ENDIF

4. Stopping criterion: repeat Evaluation until (i is less than N) or $((N-i)divstep > 0)$

The summary from the conducted analysis (fig. 1 – fig. 5):

– a direct correlation between CFS-SU merit and accuracy of learning algorithm exists, however which value of merit corresponds to the best accuracy couldn't be formalized. As a result: evaluation strategy can't be based only on CFS-SU merit value.

– both naïve Bayes and C4.5 algorithms works better with different number of features, therefore feature subset selected with CFS-SU only without knowing a learning algorithm may also negatively impact the accuracy of the built model.

The above steps 1–4 are formalized on fig. 6 and a calculation example to illustrate as the "best" feature subset can be chosen is illustrated in table 2 for horse-colic dataset with 10 feature subsets.

In table 2. CFS-SU merits are marked by bold to show how start and end indices had been moved: end index N is always decremented by constant 1 in order to not miss the feature subset which fits the best and likely correspond to the higher merit; start index i is incremented by $(N-i)divstep$ to make a search more computationally efficient. Step value δ is a hyper parameter.

Time complexity of algorithm (fig. 2) depends on values of two parameters: N and δ and doesn't depend on dataset dimension. In worth scenario its time complexity is $O\left(\frac{N}{2\delta}\right)$. At the same time, the

algorithm (fig. 2) eliminates "search strategy" from commonly used feature selection process which reduces over all time and simplifies the process.

To understand whether the new process can be used when feature is selected by Filter with CFS and other method from the specified in [15], we include in the experiments the datasets (table 3) which features have continuous values.

Input: \vec{S}' ; dataset; δ
Output: Index of feature subset which predicts the best accuracy
1. calculate \vec{M}
2. sort \vec{S}' by \vec{M} in ascending order
3. $i := \text{int}(\text{len}(\vec{S}') / 2) - 1$
4. $N := \text{len}(\vec{S}') - 1$
5. Divide dataset on train/test.
6. WHILE True
7. Fit learning algorithm with train data with features \vec{S}_i ;
8. Calculate A_{S_i} ;
9. Fit learning algorithm with train data with features \vec{S}_N ;
10. Calculate A_{N_i} ;
11. IF $A_{S_i} < A_{N_i}$ THEN $\text{index} := N$; $i := i + \text{int}((N - i) / \delta)$ ENDIF
12. IF $A_{S_i} > A_{N_i}$ THEN $\text{index} := i$; $N := N - 1$ ENDIF
13. IF $A_{S_i} = A_{N_i}$ THEN $\text{index} := i$; $i := i + 1$ ENDIF
14. IF $i \geq N$ or $(\text{int}(\frac{N - i}{\delta}) == 0)$ THEN Break ENDIF
15. ENDWHILE

Fig. 6. Algorithm of the new organization process for feature selection by Filter with CFS

Table 2. Calculation example to illustrate algorithm fig. 6 for horse-colic dataset

Merit	0.002	0.003	0.005	0.007	0.0076	0.013	0.0158	0.0162	0.0175	0.0215
Accuracy					54.67					72.00
Merit	0.002	0.003	0.005	0.007	0.0076	0.013	0.0158	0.0162	0.0175	0.0215
Accuracy							85.33			72.00
Merit	0.002	0.003	0.005	0.007	0.0076	0.013	0.0158	0.0162	0.0175	0.0215
Accuracy							85.33		70.67	
Merit	0.002	0.003	0.005	0.007	0.0076	0.013	0.0158	0.0162	0.0175	0.0215
Accuracy							85.33	72.00		

Table 3. The characteristics of datasets with continuous and categorical features

Datasets	n_observations	n_features	Machine learning task	Feature Type			Missing values, Y/N?
				Continuous	Discrete	Categorical	
automobile	205	26	Regression	+		+	Y
forecast order	60	12	Regression	+	+		N
rental building	372	108	Regression	+	+		N
boston house prices	506	13	Regression	+	+	+	N
Computer hardware	209	9	Regression	+		+	N

Experiments on continuous data follow a similar methodology as was applied for dataset from table 1. The only difference is the learning algorithms – three algorithms representing diverse approaches to learning are chosen for experiments with continuous values:

a decision tree learner for regression (C4.5 for regression); Linear Regression (LR) – a linear predictor function is used to fit a prediction model; Locally Weighted Linear Regression (LWR) – non-linear learning algorithm for fitting a regression surface

to data through multivariate smoothing. The quality of the built models is evaluated by determination coefficient (further R^2 score). The visualization of the relationships between R^2 score of C4.5, LR, LWR algorithms and CFS_PearCorr merit (fig. 7 – fig. 9) shows similar tendency as on fig. 1 – fig. 5, i.e. the highest

value of merit doesn't correspond to the highest value of R^2 score which means that an evaluation strategy can't be based only on CFS_PearCorr merit and new proposed process is applicable to be applied to get better quality of machine learning model.

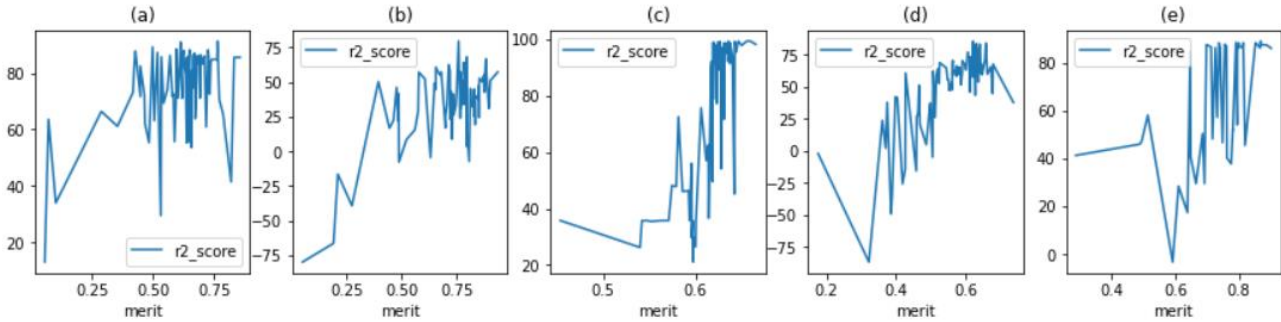


Fig. 7. R^2 score of C4.5 for regression vs CFS_PearCorr merit with (a) – automobile dataset; (b) – forecast order dataset; (c) – rental building dataset; (d) – boston house prices dataset; (e) – computer hardware datasets

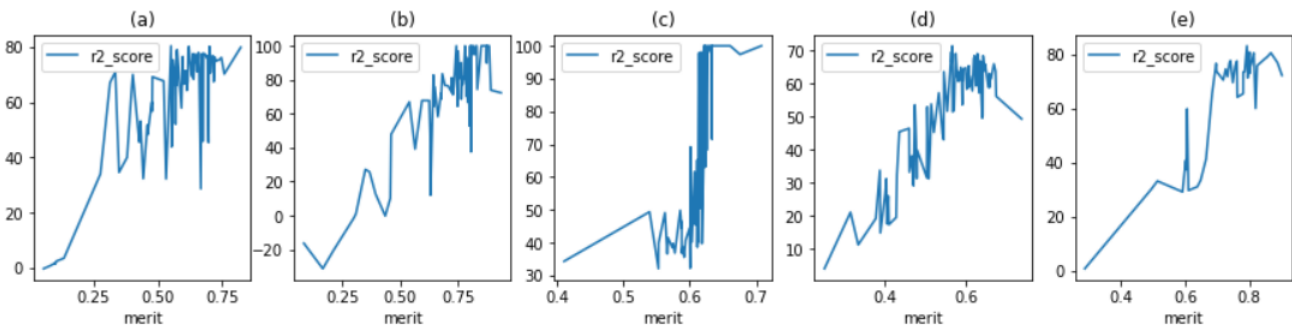


Fig. 8. R^2 score of Linear Regression vs CFS_PearCorr merit with (a) – automobile dataset; (b) – forecast order dataset; (c) – rental building dataset; (d) – boston house prices dataset; (e) – computer hardware datasets

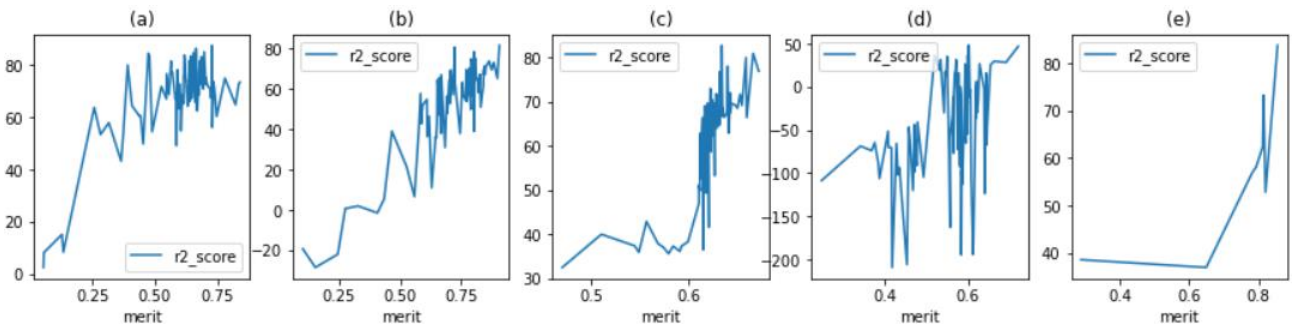


Fig. 9. R^2 score of Locally weighted linear regression vs CFS_PearCorr merit with (a) – automobile dataset; (b) – forecast order dataset; (c) – rental building dataset; (d) – boston house prices dataset; (e) – computer hardware datasets

In our experiments for datasets from table 1 the algorithm (fig. 2) had been ran with $\delta=4$ and $N=100$ and the received results are included in table 4. Columns: "all features, %" record the accuracy of naive Bays and C4.5 learning algorithms which were trained on dataset with all features. Columns: "Filter CFS_SU, %" record the accuracy of naive Bays and C4.5 learning algorithms which were trained on dataset with feature selected by approach from

work [18]. Columns: "novel Filter CFS_SU, %" record the accuracy of naive Bays and C4.5 learning algorithms which were trained on dataset with feature selected by proposed process (fig. 2). Columns "Number of selected features" record the number of features selected by process (fig. 2) with regard to learning algorithm.

Accuracies in table 4 show that there is an improvement in the performance of naive Bays

and C4.5 learning algorithms when features are selected according to the proposed process (fig. 2).

Different numbers of selected features for naive Bays and C4.5 learning algorithms proved that the same feature subset fits differently to learning algorithms and evaluation strategy of the feature selection process must take this into consideration.

The results of tests of the new flow (fig. 2) for datasets from table 3 is recorded in table 5. Columns: "all features, %" record R^2 score of C4.5, LR and LWR learning algorithms which were trained on dataset with all features. Columns: "CFS_PearCorr, %" record R^2 score of C4.5, LR and LWR algorithms which were trained on dataset with feature selected by proposed process (fig. 2). Columns "Selected features" record

the number of features selected by process (fig. 2) with regard to the learning algorithm.

R^2 scores in table 5 show that there is an improvement in the performance of C4.5 and LR algorithms when features are selected according to the proposed process (fig. 2) for all dataset. R^2 scores of LWR algorithm had improved for three datasets included in the test, however, showed the degradation for "boston house pricing" and "computer hardware" datasets. The degradation happened because the values of hyper parameters of LWR algorithm are changing when feature subset is changed therefore if new flow to be used with LWR algorithm, it should have an additional step "selection of hyper parameters" before steps 7 and 9 on fig. 2.

Table 4. Accuracy of learning algorithms for classification

Datasets	naive Bays				C4.5			
	all features, %	Filter CFS_SU, %	novel Filter CFS_SU, %	Number of selected features	all features, %	Filter CFS_SU, %	novel Filter CFS_SU, %	Number of selected features
au	77.55	75.55	83.67	27	71.43	77.14	81.63	27
hc	80	88.76	82	6	76.6	78.79	85.28	4
mu	94.49	97.53	99.7	62	99.7	99.37	100	52
kr-vs-kp	85.48	90.20	93.24	25	99.62	94.13	99.62	35
sb	88.31	91.18	93.51	23	83.12	86.80	89.61	27

Table 5. R^2 score of learning algorithms for regression

Datasets	C4.5			LR			LWR		
	all features, %	CFS_PearCorr, %	Selected features	all features, %	CFS_PearCorr, %	Selected features	all features, %	CFS_PearCorr, %	Selected features
automobile	84.35	90.55	6	77.41	81.55	8	78.52	87.55	9
Forecast order	38	71.66	7	100	100	6	55.5	81.51	2
Residential building	90.91	99.48	25	98.23	100	28	66.29	82.83	23
Boston house prices	76.17	86.12	10	68.4	70.95	11	91.9	50.5	9
computer hardware	87.89	88.5	5	80.45	83.06	5	92.22	85.53	2

P-value of paired t-test for accuracies from table 4 are recorded in tables 6 and 7 for naive Bayes and C4.5 learning algorithms correspondingly. P=0.0049 indicates statistically significant difference in accuracies of naive Bayes algorithm is obtained when features are selected by the proposed process vs accuracies of naive Bayes algorithm with all features.

P=0.018 (table 7) indicates statistically significant difference in accuracy of C4.5 algorithm is obtained

when features are selected by the proposed process vs accuracies of C4.5 algorithm with features selected by Filter with CFS-SU [18].

P-value of paired t-test for R^2 scores from table 5 are recorded in tables 8. P=0.03 indicates statistically significant difference in R^2 score of Linear Regression algorithm.

Table 6. Accuracies and p-value of paired t-test of naive Bayes

Accuracy of Naive Bayes						P-value of paired t-test
all features, %	77.55	80	94.49	85.48	88.31	0.0049
novel Filter CFS_SU, %	83.67	82	99.7	93.24	93.51	
Filter CFS_SU, %	75.55	88.76	97.53	90.2	91.18	0.49
novel Filter CFS_SU, %	83.67	82	99.7	93.24	93.51	

Table 7. Accuracies and p-value of paired t-test of naive Bayes of C4.5

Accuracy of C4.5						P-value of paired t-test
all features, %	71.43	76.6	99.7	99.62	83.12	0.07
novel Filter CFS_SU, %	81.63	85.28	100	99.62	89.61	
Filter CFS_SU, %	77.14	78.79	99.37	94.13	86.8	0.018
novel Filter CFS_SU, %	81.63	85.28	100	99.62	89.61	

Table 8. R^2 score and p-value of paired t-test of regression algorithms

R^2 score of C4.5 for regression, %						P-value of paired t-test
all features	84.35	38	90.91	76.17	87.89	0.1
CFS_PearCorr	90.55	71.44	99.48	86.12	88.5	
R^2 score of Linear Regression, %						
all features	77.41	100	98.23	68.4	80.45	0.03
CFS_PearCorr	81.55	100	100	70.95	83.06	
R^2 score of Locally weighted Linear Regression, %						
all features	78.52	55.5	66.29	91.9	92.22	0.95
CFS_PearCorr	87.55	81.51	82.83	50.5	85.53	

Conclusion and perspectives of further development

This paper has presented new organization process for feature selection by Filter with CFS. The proposed process eliminates a time consuming "search strategy" step which is commonly included in feature selection procedure but is a time consuming and not always efficient. Time complexity of the new process (fig. 2) doesn't depend on dataset's dimension which makes it robust to different varieties of datasets which is often visible.

The conducted experiments with five datasets which features have discrete values and two predefined classification algorithms: naive Bayes and C4.5 have shown that by using a new process the performance results of learning algorithms are improved. P-value of paired t-test records statistically significant difference

in the accuracies: for naive Bayes when features are selected by the proposed process compared with the accuracies of naive Bayes when all features are included in the model; for C4.5 when features are selected by the proposed process compared with the accuracies of C4.5 when features are selected by Filter with CFS-SU.

The conducted experiments with five datasets which features have continuous values and three predefined regression algorithms: C4.5 for regression, Linear Regression and Locally Weighted Linear Regression have shown that by using a new process the performance results of learning algorithms are improved. It is also noted that the new process should include additional step which is aimed to select values of hyper parameter when it is used with Locally Weighted Linear Regression algorithm.

Future work will be to extend a new approach by applying to different filter methods.

References

- Guyon, I., Elisseeff, A. (2003), "An introduction to variable and feature selection", *J. Machine Learning Research*, No. 3, P. 1157–1182.
- Dernoncourt, D., Hanczar, B., & Zucker, J.-D. (2014), "Analysis of feature selection stability on high dimension and small sample data", *Computational Statistics & Data Analysis*, No. 71, P. 681–693. DOI: <https://doi.org/10.1016/j.csda.2013.07.012>

3. Luan, C., Dong, G. (2018), "Experimental identification of hard data sets for classification and feature selection methods with insights on method selection", *Data Knowl. Eng.* No. 118, P. 41–51.
4. Senliol, B., Gulgezen, G., Yu, L., Cataltepe, Z. (2008), "Fast Correlation Based Filter (FCBF) with a different search strategy", *23rd international symposium on computer and information sciences*, P. 1–4.
5. Yu, L., Liu, H. (2021), "Enhancing Big Data Feature Selection Using a Hybrid Correlation-Based Feature Selection", No. 10, 2984 p. DOI: <https://doi.org/10.3390/electronics10232984>
6. Alzami, F., Tang, J., Yu, Z., Wu, S., Chen, P., You, J., Zhang, J. (2018), "Adaptive Hybrid Feature Selection-Based Classifier Ensemble for Epileptic Seizure Classification", *IEEE Access.*, No. 6, P. 29132–29145. DOI: <https://doi.org/10.1109/ACCESS.2018.2838559>
7. Jaina, D., Singhb, V. (2018), "An Efficient Hybrid Feature Selection model for Dimensionality Reduction", *Procedia Computer Science*, No. 132, P. 333–341.
8. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. (2011), "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research*, No. 12, P. 2825–2830.
9. Duda, R., Hart, P., Stork D. (2012), *Pattern classification*, John Wiley & Sons.
10. Mundra, P., Rajapakseab, J. (2016), "Gene and sample selection using T-score with sample selection", *Journal of Biomedical Informatics*, No. 59, P. 31–41. DOI: <https://doi.org/10.1016/j.jbi.2015.11.003>
11. Tan, H., Wang, G., Wang, W., Zhanga, Z. (2022), "Feature selection based on distance correlation: a filter algorithm", *Journal of Applied Statistics*, No. 49 (2), P. 411–426.
12. Zhai, Y., Song, W., Liu, X., Liu, L. (2018), "A Chi-Square Statistics Based Feature Selection Method in Text Classification", *IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*. DOI: <https://doi.org/10.1109/ICSESS.2018.8663882>
13. Ircioa, J., Lojo, A., Morib, U., Lozanobc, J. (2020), "Mutual information based feature subset selection in multivariate time series classification", *Pattern Recognition*, No. 108. DOI: <https://doi.org/10.1016/j.patcog.2020.107525>
14. Sarkar, D., Goswami, S. (2013), "Empirical Study on Filter based Feature Selection Methods for Text Classification", *International Journal of Computer Applications*, No. 6, P. 38–43.
15. Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R., Tang, J., Liu, H. (2018), "Feature Selection: A Data Perspective", *ACM Computing Surveys*, No. 50, P. 1–45. DOI: <https://doi.org/10.1145/3136625>
16. Koller, D., Sahami, M. (1996), "Toward optimal feature selection", *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*, P. 284–292.
17. Ahn, E., Mullen, T., Yen, J. (2011), "A two-population evolutionary algorithm for feature extraction: Combining filter and wrapper", *IEEE Congress of Evolutionary Computation (CEC)*. DOI: <https://ieeexplore.ieee.org/document/5949692>
18. Hall, M. A. (1998), *Correlation-based Feature Selection for Machine Learning*, Ph.D diss. Dept. of Computer Science, Waikato Univ.

Received 14.09.2022

Відомості про авторів / Сведения об авторах / About the Authors

Соловей Ольга Леонідівна – кандидат технічних наук, Київський національний університет будівництва і архітектури, доцент кафедри інформаційних технологій проектування та прикладної математики, Київ, Україна; e-mail: solovey.ol@knuba.edu.ua; ORCID: <https://orcid.org/0000-0001-8774-7243>

Соловей Ольга Леонидовна – кандидат технических наук, Киевский национальный университет строительства и архитектуры, доцент кафедры информационных технологий и прикладной математики, Киев, Украина.

Solovei Olga – PhD (Technical Sciences), Associate Professor, Kyiv National University of Construction and Architecture, Kyiv, Ukraine.

НОВА ОРГАНІЗАЦІЯ ПРОЦЕСУ ВИБОРУ ОЗНАК ЗА ДОПОМОГОЮ ФІЛЬТРА НА ОСНОВІ КОРЕЛЯЦІЇ

Предметом статті є методи вибору ознак, що використовуються на етапі попереднього оброблення даних перед побудовою моделей машинного навчання. Увага надається методу фільтра в разі застосування вибору ознак на основі кореляції (далі CFS) та методу симетричної невизначеності (далі CFS-SU) або кореляції Пірсона (далі PearCorr). **Метою** статті є підвищення ефективності вибору ознак за допомогою фільтра

з CFS шляхом нової організації процесу вибору ознак. **Завдання**, які вирішуються в роботі: огляд та аналіз наявної організації процесу виділення ознак фільтром із CFS; визначення причин, що спричиняють погіршення якості моделі; розроблення нового підходу; оцінювання запропонованого підходу. Для реалізації поставлених завдань використовувалися такі **методи**: теорія інформації, теорія процесів, теорія алгоритмів, теорія статистики, методи вибірки, теорія моделювання даних, наукові експерименти. **Результати**. На основі отриманих результатів доведено: 1) функція оцінки обраної підмножини ознак не може ґрунтуватися лише на CFS-оцінці, оскільки це спричиняє погіршення результатів алгоритму навчання; 2) точність алгоритмів навчання класифікації покращилася, а значення коефіцієнта детермінації алгоритмів регресії зросли, коли ознаки обиралися відповідно до запропонованого процесу. **Висновки**. Новий процес організації для вибору ознак, що пропонується в цій роботі, поєднує властивості фільтра та алгоритму навчання в стратегії оцінювання, що дає змогу обрати оптимальну підмножину ознак для попередньо визначеного алгоритму навчання. Обчислювальна складність запропонованого підходу не залежить від розмірів набору даних, що робить його стійким до будь-яких різновидів даних; запропонований процес також дає змогу заощадити час, необхідний для пошуку підмножин функцій, оскільки підмножини обираються у випадковий спосіб. Проведені експерименти довели, що продуктивність алгоритмів класифікації та регресії покращилася, порівняно з продуктивністю тих самих алгоритмів навчання, але без застосування запропонованого процесу на етапі попереднього оброблення даних.

Ключові слова: вибір ознак на основі кореляції (CFS); симетрична невизначеність (SU); кореляція Пірсона (PearCorr); критерій якості; точність; коефіцієнт детермінації.

НОВАЯ ОРГАНИЗАЦИЯ ПРОЦЕССА ВЫБОРА ПРИЗНАКОВ С ПОМОЩЬЮ ФИЛЬТРА НА ОСНОВЕ КОРРЕЛЯЦИИ

Предметом статьи являются методы выбора признаков, которые используются на этапе предварительной обработки данных перед построением моделей машинного обучения. Внимание уделяется методу фильтра при использовании выбора признаков на основе корреляции (далее CFS) и методу симметричной неопределенности (далее CFS-SU) или корреляции Пирсона. **Целью** статьи является повышение эффективности выбора признаков с помощью фильтра CFS-SU путем новой организации процесса выбора признаков. **Задачи**, решаемые в статье: обзор и анализ существующей организации процесса выделения признаков фильтром с CFS; определение причин, вызывающих ухудшение качества модели; разработка нового подхода; оценка предложенного подхода. Для реализации поставленных задач использовались следующие **методы**: теория информации, теория процессов, теория алгоритмов, теория статистики, методы выборки, теория моделирования данных, научные эксперименты. **Результаты**. На основе полученных результатов доказано: 1) функция оценки выбранного подмножества признаков не может базироваться только на CFS-оценке, поскольку это приводит к ухудшению результатов алгоритма обучения; 2) точность алгоритмов обучения классификации улучшилась, а значение коэффициента детерминации алгоритмов регрессии выросли, когда признаки выбирались в соответствии с предложенным процессом. **Выводы**. Новый процесс организации для выбора признаков, который предлагается в данной работе, сочетает свойства фильтра и алгоритма обучения в стратегии оценки, что помогает выбрать оптимальное подмножество признаков для предварительно определенного алгоритма обучения. Вычислительная сложность предлагаемого подхода не зависит от размеров набора данных, что делает его устойчивым к разным разновидностям данных; также предложенный процесс позволяет экономить время, необходимое для поиска подмножества функций, поскольку подмножества выбираются случайным образом. Проведенные эксперименты доказали, что производительность алгоритмов классификации и регрессии улучшилась по сравнению с производительностью тех же алгоритмов обучения, но без применения предложенного процесса на этапе предварительной обработки данных.

Ключевые слова: выбор признаков на основе корреляции (CFS); симметричная неопределенность (SU); корреляция Пирсона (PearCorr); критерий качества; точность; коэффициент детерминации.

Бібліографічні описи / Bibliographic descriptions

Соловей О. Л. Новая организация процессу вибору ознак за допомогою фільтра на основі кореляції. *Сучасний стан наукових досліджень та технологій в промисловості*. 2022. № 3 (21). С. 39–50. DOI: <https://doi.org/10.30837/ITSSI.2022.21.039>

Solovei, O. (2022), "New organization process of feature selection by filter with correlation-based features selection method", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (21), P. 39–50. DOI: <https://doi.org/10.30837/ITSSI.2022.21.039>

O. FEDOROVICH, V. KOSENKO, L. LUTAI, I. ZAMIRETS

METHODS AND MODELS OF RESEARCH OF INVESTMENT ATTRACTIVENESS AND COMPETITIVENESS OF PROJECT-ORIENTED ENTERPRISE IN THE PROCESS OF CREATING INNOVATIVE HIGH-TECH

The subject of the research is the investment attractiveness of enterprises producing competitive products with modern component architecture. **The aim** of the article is to develop complex methods for research of investment attractiveness of projects on creation of new high-tech products with component architecture. **Tasks** to be solved are: to use component method for the analysis of innovativeness of created high-tech products; to study investment attractiveness of an innovation project with the use of expert assessments and the method of planning experiments; to substantiate the composition of a diversification project portfolio for investment taking into account the limited possibilities of the enterprise based on the method of integer optimization; to study the performance of the project portfolio using agent-based imitational modeling. **Methods** used are: system analysis, component design, multi-criteria optimization, expert assessment and agent-based simulation modeling in the form of applied information technology. **The results** have been obtained: the analysis of investment attractiveness of projects on creation of new high-tech products with component architecture is carried out. The architecture of a complex product built with components from previous positive development experience is investigated. A classification of possible components in the architecture of a complex product is created. A set of indicators for assessing the investment attractiveness of an innovative component-based product is formed. The qualitative assessments of experts for the main indicators of the component composition of the product in the form of linguistic variables are proposed. The use of multifactor experiment to assess the investment attractiveness of a project to create a complex product with a component-based architecture is proposed. An optimization model for selecting the rational composition of components in an innovative product is developed. An agent-based simulation model is built to study the performance of work on the preparation of production of innovative products on a component basis. **Conclusions:** The use of a set of developed methods and applied information technology for the creation of high-tech products with component-based architecture allows you to plan a diversification portfolio of projects with investment attractiveness and competitiveness of products.

Keywords: component architecture; innovation components; investment attractiveness; diversification project portfolio; optimization of key project indicators; agent-based simulation model of design work.

Introduction

The interest of investors often depends on the innovativeness and competitiveness of the products that an enterprise can produce. It is relevant for an enterprise to form a new, diversified portfolio of projects in order to attract finances from possible investors. Innovativeness of a product is determined by the novelty of its components. That's why the research of innovativeness of high-tech products, based on the component architecture of the products, is relevant, which allows investors in the process of forming a promising diversification project portfolio by the enterprise to assess the possibility of its financing.

Increased competition between manufacturers of high-tech products, as well as the unstable behavior of sales markets leads to the need for innovative development of enterprises, diversification of their activities, and the formation of a portfolio of future projects that are of interest to investors. The transition

to new innovative production technologies that fully meet the requirements of Industry.4.0 depends on the influence of risks of political and economic nature. Consequently, there is a contradiction between the desire of developing enterprises to implement innovations and the lack of scientific justification for the diversification of their production activities to attract investment. There is a difficult scientific and technical problem of a compromise nature associated with the formation of such a diversification project portfolio of the developing enterprise, which, on the one hand, is interesting to investors due to its innovative attractiveness and possible competitiveness in the market of high-tech products consumption, and on the other hand, the increase in terms due to possible risks, as well as unjustified economic costs in the process of creating innovative products.

Therefore, the development of methodological support for scientific justification of investments in innovative development of the enterprise is an urgent task.

Analysis of recent research and publications

To date, the component approach in the creation of new complex mechanical engineering products exists in practical developments [1–3]. But the works do not pay enough attention to the systematization and formalization of the component approach with the identification of its qualitative characteristics [4, 5]. Explicitly there is no classification of components with components from development experience and new innovative components [6, 7]. Not enough attention has been paid to the evaluation of the component-based performance of a complex product project [8–10]. There is no formal synthesis of the component architecture of complex products based on the use of new and old components [11]. The search for a rational ratio of "new" and "old" components in the composition of a complex product has not been given proper attention [12, 13]. The innovativeness of component-based complex products has not been analyzed [14, 15]. Optimization of the component product structure has not been explicitly performed [16–19]. The investment attractiveness of component-based innovative product projects was practically not considered [20–22]. The process of creating an innovative component-based product has not been modeled [23]. There are practically no publications for expert evaluation of the component architecture of a complex product using linguistic variables [24, 25].

The aim of the article

The aim of the article is to develop a set of methods to study the investment attractiveness of projects to create new high-tech products with component architecture.

Problems being solved:

- to use the component method to analyze the innovativeness of created high-tech products;
- to research investment attractiveness of an innovation project using expert assessments and the method of planning experiments;
- justify the composition of the diversification project portfolio for investment, taking into account the limited opportunities of the enterprise based on the method of integer optimization;
- to simulate the implementation of the project portfolio using agent-based information technology.

Materials and methods of the study

The component method is based on the breakdown of the architecture of a high-tech product into components that have a modular structure and the possibility to combine them with each other to create new products. The basic platform simultaneously consists of many basic components and is developed by a separate group of highly qualified designers with the prospect of forming a new family of competitive products. The selection and structuring of basic components with their possible adaptation to the requirements to the technical characteristics of a new product is carried out for a specific project.

In the component representation of a complex high-tech product, as mentioned earlier, the following components can be distinguished:

- simple components used at the lower levels of the product architecture;
- complex components, which are structurally related simple components;
- components of previous experience with positive experience in a number of product generations;
- new components, previously used and which need to be developed, i.e. a full life cycle of creation.

It is important to note that in the creation of new components it is possible to increase the design time, retraining of personnel and new technological equipment. The more new components there will be in the architecture of the new product, the design time as well as the cost of creating the new product will increase significantly. In addition, the risks of creating the product are likely to increase, which affects the possible risks of attracting investors.

To assess the capabilities of the developing enterprise and conducting diversification to perform new innovative projects, we will use component architecture in the representation of a complex product, which can contain both "old", known, components that need to adapt (change) to the technical requirements of the new product, as well as new, innovative components that give investment attractiveness due to the possible competitiveness in the market of high-tech products.

To evaluate the creation of a new product we will apply the following key indicators:

- percentages of new and old components – P(%);
- innovation attractiveness (as a level or qualitative estimates) – Q;
- costs of creating a new product – W;
- terms of creation of a new product – T;
- risks of creating a new product – R.

To investigate the innovativeness of an individual project we will use the opinion of experts, formalized in the form of a set of estimates and presented by means of a multifactor experiment plan.

A complete factor experiment allows us to evaluate the innovativeness of the product by changing the values of the factors. Individual components of the product in the form of new or "old" components can be used as factors.

Let the singular component as a factor x_j have the value:

$x_j=1$ if the j -th component is innovative, i.e. it is new;

$x_j=0$ if the j -th component is taken from the previous experience (in the form of "old") with possible further adaptation to the technical requirements of the new product.

Full factor experiment (FFE) contains a complete enumeration of the values of all factors and three components in the composition of the new product and has the form (fig. 1).

№	x_1	x_2	x_3	P(%)	Q	W	T	R
1	0	0	0					
2	0	0	1					
3	0	1	0					
4	0	1	1					
5	1	0	0					
6	1	0	1					
7	1	1	0					
8	1	1	1					

Fig. 1. Full factor experiment

The combination 010, for example, means that as part of a new product, the first and third components were taken from previous experience, and the second is a new component. As an example, we can consider different models of the AN-178 aircraft family being created, where x_1 – fuselage; x_2 – engine; x_3 – avionics can be considered as the main components. Then 111 means complete innovation of the product containing new components. On the right side of the FFE plan there are "feedbacks" of the research, which are expert evaluations of innovativeness as values – P (%), and also evaluations on investment attractiveness – Q, costs of creation – W, terms of production – T, risks of product creation – R.

Such a plan of experiment is multivariant and allows estimating simultaneously the indicators of interest: P (%), Q, W, T, R. As estimates of experts, it is possible

to use simultaneously quantitative and qualitative values of indicators. For the quantitative values of indicators, it is possible, based on the results of the study conducted with the participation of experts, to construct a regression dependence and identify significant factors (components) influencing the investment attractiveness. Let us give an illustrated example of using FFE for expert evaluation of possible modifications of AN-178 (fig. 2).

№	x_1	x_2	x_3	P(%)	Q	W	T	R
1	0	0	0	0	1	A	A	A
2	0	0	1	10	2	B	B	A
3	0	1	0	10	3	B	C	C
4	0	1	1	20	5	C	C	C
5	1	0	0	10	4	B	D	C
6	1	0	1	20	6	C	D	C
7	1	1	0	20	7	D	D	D
8	1	1	1	30	10	E	E	E

Fig. 2. Illustrated example of a multiple recall complete factor experiment

In this example, P (%) corresponds to the percentage ratio of "new" / "old" components, Q – quantitative assessment on a ten-point scale of investment attractiveness; W, T, R – costs, manufacturing time, risks, presented in a qualitative scale:

A – the best value of the indicator;

B – excellent value of the indicator;

C – good value of the indicator;

D – good value of the indicator;

E – satisfactory value of the index.

FFE allows, using the standard calculation technique, to obtain the following regressive dependence with the help of quantitative values of investment attractiveness indicators:

$$Q = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3.$$

In the process of calculations for FFE it is necessary to represent zero as the lower value of the factor "-1", and one as the upper value of the factor "+1".

Since we are interested only in the linear part of the regression dependence to estimate the importance of the factors, we obtain:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 = 4,75 + 2x_1 + 1,5x_2 + x_3.$$

Hence, we see that for the proposed illustrated example of assessing the investment attractiveness of the AN-178 aircraft, the factor in the form of the fuselage modification is important, the engine modification is less important, and the avionics is the least important.

To choose the model of the aircraft of interest to investors it is necessary to take into account the values of all indicators P (%), Q, W, T, R both in quantitative and qualitative forms of representation.

Let us use the lexicographic ordering of qualitative values of indicators. To do this, we must place the values of the indicators as a series, with the most important indicator on the first place, and the least important indicator on the last.

Let the row of importance of the indicators look like this:

Q, W, T, R.

Then estimates of all possible options for modernizing the AN-178 can be presented as a set:

1. 1, A, A, A
2. 2, B, B, A
3. 3, B, C, C
4. 5, C, C, C
5. 4, B, D, C
6. 6, C, D, C
7. 7, D, D, D
8. 10, E, E, E.

Let us order the variants of the aircraft modification lexicographically. We get:

8. 10, E, E, E
7. 7, D, D, D
6. 6, C, D, C
4. 5, C, C, C
5. 4, B, D, C
3. 3, B, C, C
2. 2, B, B, A
1. 1, A, A, A.

The investment attractiveness will be the highest in option 8. But this option has the worst values of all the last characteristics (W, T, R). In order to select a compromise option, we will use the "threshold" value of the indicators, set by the experts in the form of a "word" **5, B, B, B**. Let's place this "word", taking into account lexicographic ordering, in the obtained ordered list of aircraft modification variants. We get:

8. 10, E, E, E
7. 7, D, D, D
6. 6, C, D, C
- 5. 5, B, B, B**
4. 5, C, C, C
5. 4, B, D, C
3. 3, B, C, C
2. 2, B, B, A
1. 1, A, A, A.

From here we see that the compromise option with innovative appeal is placed in the set:

8. 10, E, E, E
7. 7, D, D, D
6. 6, C, D, C.

Given the values of indicators W, T, R, it is reasonable to invest the project of modernization 7.7, D, D, D with good values of indicators W, T, R and good enough investment attractiveness.

To form a diversification portfolio of projects of a developing enterprise, it is necessary to take into account the possible large dimensionality of the problem, which is associated with both the number of components in the architecture of innovative products, and the number of projects in the portfolio itself. Therefore, to solve this problem we will use the methods of integer (Boolean) programming. Let us introduce the variables x_{kj} , where:

$$x_{kj} = \begin{cases} 1, & \text{if the } k\text{-th product will use} \\ & \text{the } j\text{-th innovative component;} \\ 0, & \text{if the } k\text{-th product will use} \\ & \text{the } j\text{-th component from a previous experiment.} \end{cases}$$

Then w_{kj} – the cost of creating the j -th innovative component in the k -th product;

w'_{kj} – costs of adapting the j -th component, taken from previous experience, to the technical requirements of the k -th product;

t_{kj} – time of creation of the j -th innovative component for the k -th product;

t'_{kj} – time of adaptation of the j -th component, taken from previous experience, to the technical requirements of the k -th product;

r_{kj} – risks associated with the creation of the j -th innovation component of the k -th product;

r'_{kj} – risks of adaptation of the j -th component, taken from previous experience, to the technical requirements of the k -th product.

In this case, the cost of implementing a diversified project portfolio containing N projects (products) has two components:

$$W = \sum_{k=1}^N \sum_{j=1}^{n_k} (1 - x_{kj}) w'_{kj} + \sum_{k=1}^N \sum_{j=1}^{n_k} x_{kj} \cdot w_{kj},$$

where $(1 - x_{kj}) w'_{kj}$ – costs associated with the use (adaptation) of the j -th component, which was taken from previous experience, to the technical requirements of the k -th product;

$x_{kj} \cdot w_{kj}$ – costs associated with the creation of an innovative component in the k -th product.

For the timing of the project portfolio, we obtain:

$$T = \sum_{k=1}^N \sum_{j=1}^{n_k} (1-x_{kj}) t'_{kj} + \sum_{k=1}^N \sum_{j=1}^{n_k} x_{kj} \cdot t_{kj}.$$

For the risks of implementing a diversified portfolio of projects, we will have:

$$R = \sum_{k=1}^N \sum_{j=1}^{n_k} (1-x_{kj}) r'_{kj} + \sum_{k=1}^N \sum_{j=1}^{n_k} x_{kj} \cdot r_{kj}.$$

The main criterion related to the investment attractiveness of a diversification portfolio of N projects (products) will be as follows:

$$Q = \sum_{k=1}^N \sum_{j=1}^{n_k} (1-x_{kj}) q'_{kj} + \sum_{k=1}^N \sum_{j=1}^{n_k} x_{kj} \cdot q_{kj},$$

where q'_{kj} – investment attractiveness of the k -th project, which uses the j -th component taken from the previous experiment;

q_{kj} – investment attractiveness of the k -th project, which uses the j -th innovative component.

It is necessary to maximize the investment attractiveness of a diversified portfolio of projects:

$$\max Q, Q = \sum_{k=1}^N \sum_{j=1}^{n_k} (1-x_{kj}) q'_{kj} + \sum_{k=1}^N \sum_{j=1}^{n_k} x_{kj} \cdot q_{kj}.$$

At the same time, it is necessary to meet the constraints on costs, timing of the project portfolio and risks:

$$W \leq W, W = \sum_{k=1}^N \sum_{j=1}^{n_k} (1-x_{kj}) w'_{kj} + \sum_{k=1}^N \sum_{j=1}^{n_k} x_{kj} \cdot w_{kj},$$

$$T \leq T, T = \sum_{k=1}^N \sum_{j=1}^{n_k} (1-x_{kj}) t'_{kj} + \sum_{k=1}^N \sum_{j=1}^{n_k} x_{kj} \cdot t_{kj},$$

$$R \leq R, R = \sum_{k=1}^N \sum_{j=1}^{n_k} (1-x_{kj}) r'_{kj} + \sum_{k=1}^N \sum_{j=1}^{n_k} x_{kj} \cdot r_{kj},$$

where W – acceptable costs associated with the implementation of a diversified portfolio of projects;

T – acceptable terms of implementation of the diversification portfolio of projects;

R – acceptable risks associated with the implementation of a diversified portfolio of projects.

To model the diversification portfolio of projects, an applied information technology of agent-based simulation modeling was developed. The modeling will allow to take into account the time of the project portfolio implementation, as well as the impact of risks on the timing and implementation of projects. Fig. 3 shows a block diagram of the agent model for the implementation of a diversification portfolio of projects of a developing enterprise.

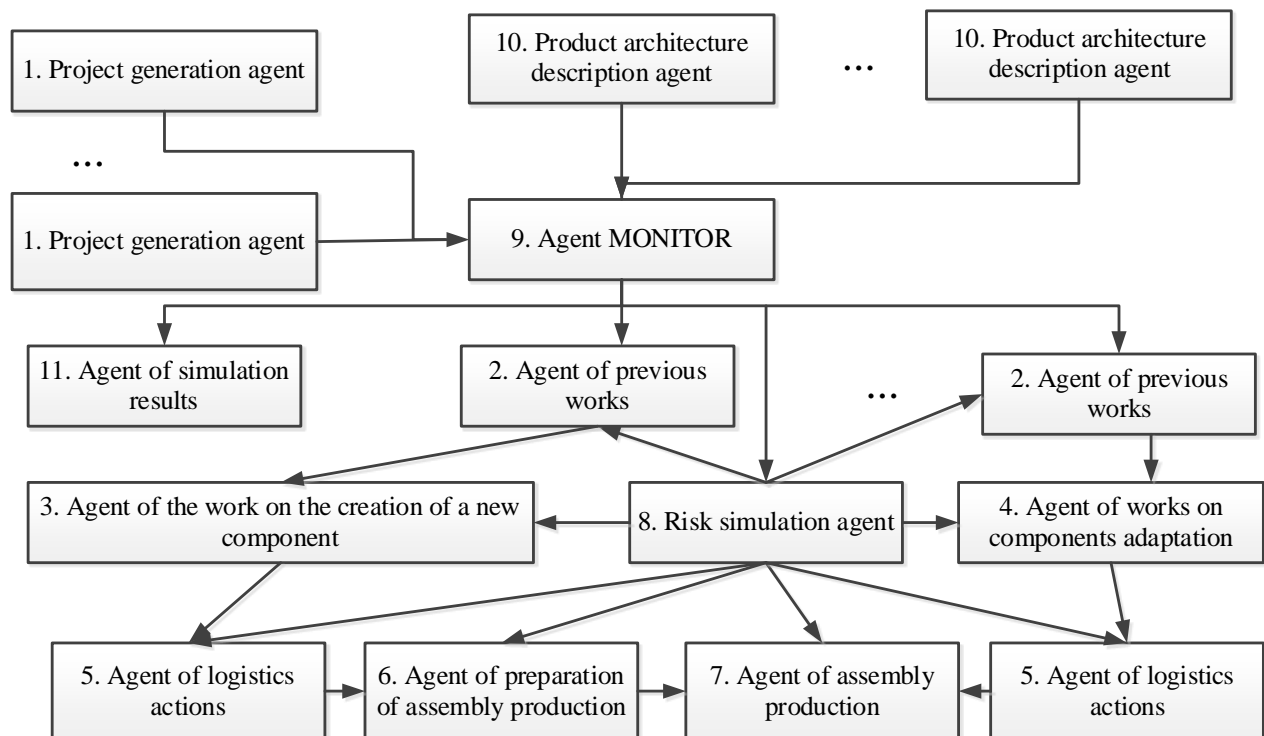


Fig. 3. Block diagram of agent-based simulation of the diversification portfolio of enterprise projects

In order to simulate the implementation of a diversified portfolio of enterprise projects, it is necessary to use the following types of agents:

1. Agent for generating the start of project execution.
2. Agent of preliminary work related to the preparation of the production of a product component ("new" or "old").
3. Agent of the execution of work on the creation of a new component.
4. Agent of works on adaptation of the component, which was taken from the previous experience, to the technical requirements of the new product.
5. Agent related to the logistics of moving the component that has been produced to the assembly production.
6. Agent for the preparation of the assembly production of a new high-tech product.
7. Agent of the assembly work on the manufactured new product.
8. Agent of simulating the risks of performing work to create a new product.
9. MONITOR agent to manage the process of simulation event modeling.
10. Agent for describing the component architecture of a new product.
11. Agent of simulation results.

Conclusion

The analysis of publications and results on this topic was carried out and the lack of a systematic and formal presentation of component synthesis for the creation of innovative engineering products was revealed. A classification of the component composition of a technical product based on the following components is proposed: components from previous positive development experience, new innovative components and combined components. The investment attractiveness of a project to create a complex product with a component architecture and using the method of the theory of experiment planning and expert assessments of designers is investigated. Both quantitative and qualitative assessments are used to determine the investment attractiveness of a diversification project to create a new complex product on a component basis. To solve the optimization problem of large dimensionality, with the choice of a rational composition of components of innovative products, integer (Boolean) programming is used. In solving a multi-criteria problem, cost, time and project risks are used to create a complex product on a component basis. An agent-based simulation model has been developed to study the implementation of a project to create a complex engineering product with a component architecture. The complex of methods is used with the help of applied information technology, which allows planning the innovative activity of the enterprise to ensure investment attractiveness.

References

1. Karimov, T., Khamadeev, Sh. (2019), "The production quality management based on the precedent approach and the clustering of use case", *Género & Direito*, No. 8 (5), P. 173–188. DOI: <https://doi.org/10.22478/ufpb.2179-7137.2019v8n5.48626>
2. Hänel, A., Schnellhardt, T., Wenkler, E., Nestler, A., Brosius, A., Corinth C., Fay, A., Ihlenfeldt, S. (2020), "The development of a digital twin for machining processes for the application in aerospace industry", *Procedia CIRP*, No. 93, P. 1399–1404. DOI: <https://doi.org/10.1016/j.procir.2020.04.017>
3. Dobrynin, A., Gudkov, M., Koynov, R. (2020), "A precedent approach to incident management in automated process control systems", *Software systems and computational methods*, No. 2, P. 45–52. DOI: <https://doi.org/10.7256/2454-0714.2020.2.31040>
4. Fedorovich, O., Lutai, L. (2010), "Formation the architecture of complex aerospace products based on a precedent approach", *Radioelectronic and computer systems*, No. 4, P. 138–142.
5. De Oliveira Moura Santos, L. F., Osiro, L., Palma Lima, R. H. (2017), "A model based on 2-tuple fuzzy linguistic representation and Analytic Hierarchy Process for supplier segmentation using qualitative and quantitative criteria", *Expert Systems with Applications*, No. 79, P. 53–64. DOI: <https://doi.org/10.1016/j.eswa.2017.02.032>
6. Fedorovich, O., Lutai, L. (2010), "Formation the architecture of complex aerospace products based on a precedent approach", *Radioelectronic and computer systems*, No. 4, P. 138–142.

7. Fedorovich, O., Lutai, L., Pisklova, T. (2011), "Component design based on positive experience from past developments", In: *IX International scientific and practical conference "Modern information technologies in economics and management of enterprises, programs and projects"*, Kharkiv, Ukraine, P. 65–66.
8. Matuszek, J., Seneta T., Plinta D., Więcek D. (2020), "Manufacturability assessment in assembly processes", *IFAC-PapersOnLine*, No. 53 (2), P. 10536–10541. DOI: <https://doi.org/10.1016/j.ifacol.2020.12.2801>
9. Kafiev, I., Romanov, P., Romanova, I. (2019), "The Method of Selecting the Best Technical System by the Set of Characteristics", In: *International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, IEEE. DOI: <https://doi.org/10.1109/FarEastCon.2019.8934392>
10. Guo, X, Liu, Y., Zhao, W., Wang, J., Chen, L. (2021), "Supporting resilient conceptual design using functional decomposition and conflict resolution", *Advanced Engineering Informatics*, No. 48. DOI: <https://doi.org/10.1016/j.aei.2021.101262>
11. Wang, Z., Rodríguez, R. M., Wang, Y., Martínez, L. (2021), "A two-stage minimum adjustment consensus model for large scale decision making based on reliability modeled by two-dimension 2-tuple linguistic information", *Computers & Industrial Engineering*, No. 151. DOI: <https://doi.org/10.1016/j.cie.2020.106973>
12. Jiao, Z., Yu, B., Wu, Sh., Shang, Y., Huang, H., Tang, Zh., Wei, R., Li, Ch. (2019), "An intelligent design method for actuation system architecture optimization for more electrical aircraft", *Aerospace Science and Technology*, No. 93. DOI: <https://doi.org/10.1016/j.ast.2019.03.048>
13. Li, T., Lockett, H., Lawson, C. (2020), "Using requirement-functional-logical-physical models to support early assembly process planning for complex aircraft systems integration", *Journal of Manufacturing Systems*, No. 54, P. 242–257. DOI: <https://doi.org/10.1016/j.jmsy.2020.01.001>
14. Meissner, D., Burton, N., Galvin, P., Sarpong, D., Bach, N. (2021), "Understanding cross border innovation activities: The linkages between innovation modes, product architecture and firm boundaries", *Journal of Business Research*, No. 128, P. 762–769. DOI: <https://doi.org/10.1016/j.jbusres.2019.05.025>
15. Tang, H. (2018), "An Integrated Product-Process Hierarchical Modeling Method for Development of Complex Assembly Manufacturing Systems", *Procedia CIRP*, No. 76, P. 2–6. DOI: <https://doi.org/10.1016/j.procir.2018.01.023>
16. Yuan, Zh., Chen, H., Xie, P., Zhang, P., Liu, J., Li, T. (2021), "Attribute reduction methods in fuzzy rough set theory: An overview, comparative experiments, and new directions", *Applied Soft Computing*, No. 107. DOI: <https://doi.org/10.1016/j.asoc.2021.107353>
17. Wang, H., Zhang, Y.-M., Yang Zh. (2019), "A risk evaluation method to prioritize failure modes based on failure data and a combination of fuzzy sets theory and grey theory", *Engineering Applications of Artificial Intelligence*, No. 82, P. 216–225. DOI: <https://doi.org/10.1016/j.engappai.2019.03.023>
18. Dong, J., Hou, J. (2017), "Output feedback fault-tolerant control by a set-theoretic description of T–S fuzzy systems", *Applied Mathematics and Computation*, No. 301, P. 117–134. DOI: <https://doi.org/10.1016/j.amc.2016.12.017>
19. Zambetti, M., Khan, M. A., Pinto, R., Wuest, Th. (2020), "Enabling servitization by retrofitting legacy equipment for Industry 4.0 applications: benefits and barriers for OEMs", *Procedia Manufacturing*, No. 48, P. 1047–1053. DOI: <https://doi.org/10.1016/j.promfg.2020.05.144>
20. Liu, H., Chu, X., Xue, D. (2019), "An optimal concurrent product design and service planning approach through simulation-based evaluation considering the whole product life-cycle span", *Computers in Industry*, No. 111, P. 187–197. DOI: <https://doi.org/10.1016/j.compind.2019.07.008>
21. Kobayashi, H., Matsumoto, T., Fukushige, Sh. (2018), "A simulation methodology for a system of product life cycle systems", *Advanced Engineering Informatics*, No. 36, P. 101–111. DOI: <https://doi.org/10.1016/j.aei.2018.03.001>
22. Bertoni, M., Pezzotta, G., Scandella, B., Wall, J., Jonsson, P. (2019), "Life cycle simulation to support cross-disciplinary decision making in early PSS design", *Procedia CIRP*, No. 83, P. 260–265. DOI: <https://doi.org/10.1016/j.procir.2019.03.138>
23. Fedorovich, O., Uruskiy, O., Pronchakov, Y., Lukhanin, M. (2021), "Method and information technology to research the component architecture of products to justify investments of high-tech enterprise", *Radioelectronic and computer systems*, No. 1, P. 150–157. DOI: <https://doi.org/10.32620/reks.2021.1.13>
24. Bustio-Martínez, L., Letras-Luna, M., Cumplido, R., Hernández-León, R., Feregrino-Uribe, C., Bande-Serrano, J. M. (2019), "Using hashing and lexicographic order for Frequent Itemsets Mining on data streams", *Journal of Parallel and Distributed Computing*, No. 125, P. 58–71. DOI: <https://doi.org/10.1016/j.jpdc.2018.11.002>
25. Pai, K.-J., Chang, J.-M., Wu, R.-Y., Chang, Sh.-Ch. (2019), "Amortized efficiency of generation, ranking and unranking left-child sequences in lexicographic order", *Discrete Applied Mathematics*, No. 268, P. 223–236. DOI: <https://doi.org/10.1016/j.dam.2018.09.035>

Відомості про авторів / Сведения об авторах / About the Authors

Федорович Олег Євгенович – доктор технічних наук, професор, завідувач кафедри комп'ютерних наук та інформаційних технологій, Національний аерокосмічний університет ім. М. С. Жуковського "Харківський авіаційний інститут", Харків, Україна; e-mail: o.fedorovych@khai.edu; ORCID ID: <https://orcid.org/0000-0001-7883-1144>

Федорович Олег Євгеньевич – доктор технических наук, профессор, заведующий кафедрой компьютерных наук и информационных технологий, Национальный аэрокосмический университет им. Н. Е. Жуковского "Харьковский авиационный институт", Харьков, Украина.

Fedorovich Oleg – Doctor of Sciences (Engineering), Professor, Head of the Department of Computer Sciences and Information Technologies, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine.

Косенко Віктор Васильович – доктор технічних наук, професор, Національний університет "Полтавська політехніка імені Юрія Кондратюка", професор кафедри автоматизації, електроніки та телекомунікацій, м. Полтава, Україна; kosvict@gmail.com; ORCID ID: <http://orcid.org/0000-0002-4905-8508>

Косенко Виктор Васильевич – доктор технических наук, профессор, Национальный университет "Полтавская политехника имени Юрия Кондратюка", профессор кафедры автоматизации, электроники и телекоммуникаций, г. Полтава, Украина.

Kosenko Viktor – Doctor of Sciences (Engineering), Professor of Automation, Electronic and Telecommunication Department of National University "Yuri Kondratyuk Poltava Polytechnic", Poltava, Ukraine.

Лутай Людмила Миколаївна – кандидат технічних наук, доцент кафедри мехатроніки та електротехніки, Національний аерокосмічний університет ім. М. С. Жуковського "Харківський авіаційний інститут", Харків, Україна; e-mail: l.lutay@khai.edu; ORCID ID: <https://orcid.org/0000-0003-1822-8938>

Лутай Людмила Николаевна – кандидат технических наук, доцент кафедры мехатроники и электротехники, Национальный аэрокосмический университет им. Н. Е. Жуковского "Харьковский авиационный институт", Харьков, Украина.

Lutai Liudmyla – PhD (Engineering Sciences), Associate Professor of the department Mechatronics and Electrical Engineering, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine.

Замірець Ярослав Олегович – молодший науковий співробітник, Державне підприємство "Науково-дослідний технологічний інститут приладобудування", Харків, Україна; e-mail: y.zamirets@gmail.com; ORCID ID: <https://orcid.org/0000-0002-3522-6937>

Замирец Ярослав Олегович – младший научный сотрудник, Государственное предприятие "Научно-исследовательский технологический институт приборостроения", Харьков, Украина.

Zamirets Yaroslav – junior researcher, State Enterprise Scientific Research Technological Institute of Instrument Engineering, Kharkov, Ukraine.

МЕТОДИ ТА МОДЕЛІ ДОСЛІДЖЕННЯ ІНВЕСТИЦІЙНОЇ ПРИВАБЛИВОСТІ ТА КОНКУРЕНТОСПРОМОЖНОСТІ ПРОЕКТНО-ОРІЄНТОВАНОГО ПІДПРИЄМСТВА ПРИ СТВОРЕННІ ІННОВАЦІЙНИХ ВИСОКОТЕХНОЛОГІЧНИХ ВИРОБІВ

Предметом дослідження є інвестиційна привабливість підприємств, що створюють конкурентоспроможні вироби із сучасною компонентною архітектурою. **Мета** статті – розроблення методу дослідження інвестиційної привабливості проєктів зі створення нових високотехнологічних виробів із компонентною архітектурою. **Вирішені завдання:** використання компонентного методу для аналізу інноваційності створюваних високотехнологічних виробів; дослідження інвестиційної привабливості інноваційного проєкту з використанням експертних оцінок і методу планування експериментів; обґрунтування складу диверсифікаційного портфелю проєктів для інвестування з урахуванням обмежених можливостей підприємства на основі методу цілочисельної оптимізації; дослідження виконання портфелю проєктів з використанням агентного імітаційного моделювання. **Методи,** що застосовуються: системний аналіз, компонентне проєктування, багатокритеріальна оптимізація, експертне оцінювання та агентне імітаційне моделювання у вигляді прикладної інформаційної технології. **Отримані результати:** проведено аналіз інвестиційної привабливості проєктів зі створення нових високотехнологічних виробів із компонентною архітектурою. Досліджено архітектуру складного виробу, яка побудована за допомогою компонент із минулого позитивного досвіду розробок. Створено класифікацію можливих компонент в архітектурі складного виробу. Сформовано множину показників для оцінювання інвестиційної привабливості інноваційного виробу на компонентній основі. Запропоновано якісні оцінки

експертів для основних показників компонентного складу виробу у формі лінгвістичних змінних. Запропоновано використання багатофакторного експерименту для оцінювання інвестиційної привабливості проєкту зі створення складного виробу з компонентною архітектурою. Розроблено оптимізаційну модель щодо вибору раціонального складу компонент в інноваційному виробі. Побудовано агентну імітаційну модель для дослідження виконання робіт із підготовки виробництва інноваційних виробів на компонентній основі. **Висновки:** використання запропонованого методу, що містить комплекс розроблених моделей та прикладну інформаційну технологію, для створення високотехнологічних виробів із компонентною архітектурою дозволяє планувати диверсифікаційний портфель проєктів з інвестиційною привабливістю та конкурентоспроможністю виробів.

Ключові слова: компонентна архітектура виробу; інноваційні компоненти; інвестиційна привабливість; диверсифікаційний портфель проєктів; оптимізація основних показників проєкту; агентна імітаційна модель проєктних робіт.

МЕТОДЫ И МОДЕЛИ ИССЛЕДОВАНИЯ ИНВЕСТИЦИОННОЙ ПРИВЛЕКАТЕЛЬНОСТИ И КОНКУРЕНТОСПОСОБНОСТИ ПРОЕКТНО-ОРИЕНТИРОВАННОГО ПРЕДПРИЯТИЯ ПРИ СОЗДАНИИ ИННОВАЦИОННЫХ ВЫСОКОТЕХНОЛОГИЧЕСКИХ ИЗДЕЛИЙ

Предметом исследования является инвестиционная привлекательность предприятий, создающих конкурентоспособные изделия с современной компонентной архитектурой. **Цель** статьи – разработка метода исследования инвестиционной привлекательности проектов по созданию новых высокотехнологичных изделий с компонентной архитектурой. **Решаемые задачи:** использование компонентного метода для анализа инновационности создаваемых высокотехнологичных изделий; проведение исследования инвестиционной привлекательности инновационного проекта с использованием экспертных оценок и метода планирования экспериментов; обоснование состава диверсификационного портфеля проектов для инвестирования с учетом ограниченных возможностей предприятия на основе метода целочисленной оптимизации; исследование выполнения портфеля проектов с использованием агентного имитационного моделирования. **Использованные методы:** системный анализ, компонентное проектирование, многокритериальная оптимизация, экспертные оценки и агентное имитационное моделирование в виде прикладной информационной технологии. **Полученные результаты:** проведен анализ инвестиционной привлекательности проектов по созданию новых высокотехнологичных изделий с компонентной архитектурой. Исследована архитектура сложного изделия, которая построена с помощью компонентов из прошлого положительного опыта разработок. Создана классификация возможных компонентов в архитектуре сложного изделия. Сформировано множество показателей для оценки инвестиционной привлекательности инновационного изделия на компонентной основе. Предложены качественные оценки экспертов для основных показателей компонентного состава изделия в виде значений лингвистических переменных. Предложено использование многофакторного эксперимента по оценке инвестиционной привлекательности проекта по созданию сложного изделия с компонентной архитектурой. Разработана оптимизационная модель по выбору оптимального состава компонент в инновационном изделии. Построена имитационная агентная модель для исследования выполнения работ по подготовке производства инновационных изделий на компонентной основе. **Выводы:** использование предложенного метода, включающего в себя комплекс разработанных моделей и прикладную информационную технологию, для создания высокотехнологичных изделий с компонентной архитектурой позволяет планировать диверсификационный портфель проектов с инвестиционной привлекательностью и конкурентоспособностью изделий.

Ключевые слова: компонентная архитектура изделия; инновационные компоненты; инвестиционная привлекательность; диверсификационный портфель проектов; оптимизация основных показателей проекта; агентная имитационная модель проектных работ.

Бібліографічні опису / Bibliographic descriptions

Федорович О. Є., Косенко В. В., Лутай Л. М., Замірець Я. О. Методи та моделі дослідження інвестиційної привабливості та конкурентоспроможності проєктно-орієнтованого підприємства при створенні інноваційних високотехнологічних виробів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2022. № 3 (21). С. 51–59. DOI: <https://doi.org/10.30837/ITSSI.2022.21.051>

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>

K. YUVCHENKO, V. YESILEVSKYI, O. SEREDA

HUMAN EMOTION RECOGNITION SYSTEM USING DEEP LEARNING ALGORITHMS

The **subject** of research in the article is the software implementation of a neural image classifier. The work examines emotions as a special type of mental processes that express a person's experience of his attitude to the surrounding world and himself. They can be expressed in different ways: facial expressions, posture, motor reactions, voice. However, the human face has the greatest expressiveness. Technologies for recognizing companies to improve customer service use human emotions make decisions about interviewing candidates and optimize the emotional impact of advertising. Therefore, the purpose of the work is to find and optimize the most satisfactory in terms of accuracy algorithm for classifying human emotions based on facial images. The following tasks are **solved**: review and analysis of the current state of the problem of "recognition of emotions"; consideration of classification methods; choosing the best method for the given task; development of a software implementation for the classification of emotions; conducting an analysis of the work of the classifier, formulating conclusions about the work performed, based on the received data. An image classification **method** based on a densely connected convolutional neural network is also used. **Results**: the results of this work showed that the method of image classification, based on a densely connected convolutional neural network, is well suited for solving the problems of emotion recognition, because it has a fairly high accuracy. The quality of the classifier was evaluated according to the following metrics: accuracy; confusion matrix; precision, recall, f1-score; ROC curve and AUC values. The accuracy value is relatively high – 63%, provided that the data set has unbalanced classes. AUC is also high at 89%. **Conclusions**. It can be concluded that the obtained model with weights has high indicators of recognition of human emotions, and can be successfully used for its purpose in the future.

Keywords: object detection; classification of objects; supervised learning; recognition of emotions.

Introduction

Emotion is a special type of mental processes that expresses a person's feelings and attitude to the world around him or her. Emotions can be transmitted in different ways: facial expressions, posture, motor reactions, voice. However, the most expressive is the human face. Technologies for recognizing companies to improve customer service use human emotions make decisions about interviewing candidates and optimize the emotional impact of advertising.

Using facial emotion recognition, businesses can process images and videos in real time to monitor video streams or automate video analytics, thereby reducing costs and making the lives of their users better.

Recognizing human emotions in an image with complex background structure is the first and main challenge in solving the emotion recognition task.

The task of emotion recognition is very relevant today in various fields, including medicine (helps people with autism to recognize the emotions of others and establish normal communication), sociology, robotics, marketing, psychology, etc.

Face recognition is still a challenging problem due to the variety of head poses and background settings.

People always try to hide or even falsify their true facial expressions. The huge challenges in facial expression recognition are finding and understanding facial expressions. Over the past few years, science has made sufficient progress, and tasks such as landmark detection, object modeling and much more can be performed using computer vision.

However, researchers usually pretended that their systems should work in a controlled environment. In other words, the face and background should not be difficult to identify. Meanwhile, other scientists are recognizing facial expressions in the wild using deep neural networks so that the machine can react like a human to various factors without limitations, such as dynamic lighting, head position, etc.

In the nineteenth century, one of the most important works on the analysis of facial emotions, which is directly related to the modern field of automatic facial expression recognition, was the work of Charles Darwin [1]. In 1872, the scientist wrote a dissertation on this problem. This laid the foundation for general provisions for determining the emotions of humans and animals. Darwin grouped the names of emotions. The classification looks like this:

- bad mood, anxiety, grief, despair and despair;
- joy, elation, love, tender feelings and devotion;

- reflection, meditation, irritability and gloom;
- hatred and anger;
- contempt, scorn, disgust, guilt and pride;
- surprise, amazement, fear and terror;
- attention to oneself, shame, bashfulness and modesty.

Darwin also classified the facial changes that occur for each of the above classes of expressions. For example, "clenching of the muscles around the eyes in case of sadness", "tight closing of the mouth during contemplation", etc.

Another important milestone in the study of facial expressions and human emotions was the work of Paul Ekman and his colleagues from the 1970s. Their work had a huge impact on the development of modern automatic facial expression recognizers.

Suwa et al conducted the earliest study of automatic facial expression recognition in 1978. As a result, a model was created to learn facial expressions from a sequence of images using 20 tracking arguments. Research has been conducted since the late 1980s and early 1990s, when mobile computing power of the economy became available. This helped to develop face detection and tracking algorithms in the early 1990s. At the same time, research on human-computer interaction (HCI) and affective computing (AC) began [2].

The aim of this work is to find and optimize the most satisfactory in terms of accuracy algorithm for classifying human emotions from face images.

Analysis of recent research and publications

FaceReader

Many researchers have turned to using automated facial expression analysis software to more objectively assess emotions. FaceReader is used worldwide in over 1000 universities, research institutes and companies in most industries. The consumer and psychological benefits of FaceReader and its usability are being researched.

FaceReader software is fast, flexible, objective, accurate and easy to use. It immediately analyzes your data (live, video or still images), saving valuable time. The ability to record audio and video allows you to hear what people say, for example, during human-computer interaction or while viewing stimuli. FaceReader uses the following emotion categories: joy, surprise, sadness, anger, disgust, contempt, fear. Ekman [3] described these emotional categories as basic or universal emotions. Obviously, facial expressions vary in intensity and are

often a combination of several emotions. In addition, there are quite a few interpersonal varieties.

FaceReader works in three stages:

1) Face detection. FaceReader uses the popular Viola-Jones algorithm [4] to detect the presence of a face in an image.

2) Accurate face modeling using an algorithmic approach based on the Active Appearance method [5] described by Cootes and Taylor [6]. At this stage, the model is trained using a database of annotated images. This method describes over 500 key points on the face and the facial texture, e.g.:

a) points that cover the face (the part of the face analyzed by FaceReader);

b) points on the face that are easy to recognize (lips, eyebrows, nose and eyes).

3) The actual classification of facial expressions is performed by training an artificial neural network [7]. More than 10000 manually annotated images were used as training material.

Texture is also important as it gives additional information about the state of the face. The key points only describe the general position and shape of the face, but do not provide any information, for example, about the presence of wrinkles and the shape of eyebrows (such features, although they may seem less important, play a significant role in the classification of facial expressions). FaceReader works in three stages:

1) Face detection. FaceReader uses the popular Viola-Jones algorithm [4] to detect the presence of a face in an image.

2) Accurate face modeling using an algorithmic approach based on the Active Appearance method [5] described by Cootes and Taylor [6]. At this stage, the model is trained using a database of annotated images. This method describes over 500 key points on the face and the facial texture, e.g.:

a) points that cover the face (the part of the face analyzed by FaceReader);

b) points on the face that are easy to recognize (lips, eyebrows, nose and eyes).

3) The actual classification of facial expressions is performed by training an artificial neural network [7]. More than 10,000 manually annotated images were used as training material.

Texture is also important as it gives additional information about the state of the face. The key points only describe the general position and shape of the face, but do not provide any information; for example, about the presence of wrinkles and the shape of

eyebrows (such features, although they may seem less important, play a significant role in the classification of facial expressions).

EmoDetect

This is the name of the program that helps to determine the psycho-emotional state of a person by a sample of images by identifying six emotions: happiness, surprise, sadness, anger, fear, disgust. EmoDetect program contains the principles of coding facial expressions FACS Action Units, developed by Paul Ekman.

During the study, a person is shown various stimuli on a monitor, and a camera is located in front of the test participant, which shoots video.

The system processes the video according to the following algorithm: the image of a person whose facial expression is to be classified is captured. Then the face is extracted from the image using the Jones method to determine the elements. Next, the algorithms extract the corresponding elements from the cropped image of the face. Trained classifiers match these features with the corresponding emotions.

The code is implemented using OpenCV, an open source computer vision library.

The final stage is the fixation of emotions with reference to the time of demonstration and visualization of the results in the form of graphs and tables.

Viola–Jones method

The Viola–Jones method is used to detect objects in images. This method can recognize different classes of images, but the main task in its creation is face recognition. The algorithm of this method is able to detect objects very reliably and quickly enough to work in real time.

The Viola–Jones detector is a cascade of classifiers.

It has the following features:

- Haar features are used;
- images are presented in a holistic form;
- boosting is used.

Haar feature is a rectangular primitive. From the very beginning, the authors proposed the main four primitives, but the OpenCV library uses additional primitives that improve the quality of recognition with additional tests from non-standard points of view.

Each feature is paired with a threshold, and the decision of the feature is determined by comparing its value with the threshold. The Viola–Jones method implements a fast way to calculate attribute values

that uses the integral representation of the image. The integral representation of an image is a matrix that has the same dimensions as the original image. The value of its elements is defined as the sum of ridge intensity peaks located on the left and top. In such a matrix, the sum of pixel values in an arbitrary rectangle is calculated for a constant time.

The detection process is performed by sliding the detection window over the entire image. A cascade solution is calculated for each window. In the case of a positive answer, it is assumed that the desired object is located inside the window. After one pass of the image is completed, the window size is increased. The window size is increased until the predefined size is reached. A smaller percentage of magnification improves the detection rate but increases the overall processing time [8].

Materials and methods

Nowadays, with the spread of artificial intelligence, such fields as machine learning and its branches, deep learning and neural networks have gained immense popularity. Learning requires software and tools such as classifiers that feed huge amounts of data, analyze it and perform useful functions. The goal of the classification process is to categorize all pixels of a digital image into one of several classes. Usually multispectral data is used to perform the classification. Indeed, the spectral pattern present in the data for each pixel is used as a numerical basis for categorization.

Image classification is probably the most important part of digital image analysis. Classification between objects is a complex process, so it is an important task for the field of computer vision. The essence of image classification is that an image is labeled and put into a number of predefined classes. There are potentially classes to which an image can belong. Manually checking and classifying images can be a tedious task, especially if there are a large number of them. Therefore, it is very useful if it is possible to automate this whole process using computer vision.

The algorithm for performing image classification consists of four stages.

- Image preprocessing, which aims to improve the image data (features). To do this, unwanted aspects are removed and some important features of the image are enhanced so that computer vision models can use this enhanced data to work. The image preprocessing steps include reading the image, resizing the image, and
-

enhancing the data (gray-scaling, reflection, Gaussian blur, histogram, alignment, and rotation).

- Object detection. This stage localizes the object, that is, the image is segmented and the position of the object is determined.

- Feature extraction and training. This is a crucial stage where statistical or deep learning techniques are used to identify the most interesting patterns in the image, features that may be unique to a particular class and that will later help the model to discriminate between different classes. (The process where a model learns features from a dataset is called model training.)

- Object classification. In this step, the detected objects are grouped using an appropriate classification technique that compares the image patterns with the target patterns.

In recent years, the application of machine learning techniques in everyday life has become commonplace. Many modern websites and devices use machine learning algorithms, ranging from automatic recommendations for watching movies, ordering food or buying groceries, to personalized online radio broadcasts and recognizing human emotions in an image.

Machine learning can be broadly defined as computational techniques that use experience to improve performance or to make accurate predictions. Since the success of a learning algorithm depends on the data used, machine learning is closely related to data analysis and statistics. In general, machine learning methods are data mining techniques that combine fundamental concepts of computer science with statistics, probability, and optimization [9]. It consists of the development of efficient and accurate prediction algorithms, and if more time is devoted and more input samples of documents are specified for training, then the predictions of such models will be better. As in other areas of computer science, some critical indicators of the quality of algorithms are their temporal and spatial complexity.

Neural networks are widely used to classify emotions from facial expressions. They employ both supervised (training with a teacher) and unsupervised (training without a teacher) neural architectures for emotion classification.

In this paper, a supervised learning algorithm is used to perform the image classification task.

Supervised (learning with a teacher) classification is based on the idea that the user can select samples of pixels in the image that are representative of certain classes, and then the software processes the image using training sites to classify all other pixels in the image.

The training sites (also known as test sets or input classes) are selected based on the user's knowledge. The user also sets boundaries on how similar other pixels must be to be grouped together. These bounds are often set based on the spectral characteristics of the training region. The user also defines the number of classes to which the image belongs.

The main goal of emotion recognition is to identify human emotional states based on face images. Deep convolutional neural network (DCNN) architectures have been successfully used for image classification [10, 11]. As the network deepens, the input information (or back propagation gradient) passes through many layers and may disappear by the time it reaches the end of the network. Several DCNN architectures have been proposed to address this problem. Highway Networks [12], ResNet (Residual Neural Network) [13] and DenseNet 169 traverse the signal from one layer to another by identifying intermediate connections between layers. FractalNets [14] repeatedly combines multiple parallel sequences with different numbers of convolutional units to obtain a large nominal depth while maintaining many short paths in the network. Although these methods differ in terms of network architecture and learning strategy, they all share a key characteristic: creating short paths from earlier layers to later layers. The DenseNet 169 model has slightly higher overall accuracy than the ResNet and FractalNets models. It will be used in this paper.

In 2017, DenseNets [15] was proposed at the CVPR 2017 conference. They started by trying to build a deeper network based on the idea that if a convolutional network has shorter connections between its layers close to the inputs and outputs, then this deep convolutional network can be more accurate and more efficient for training. Unlike ResNet, which adds a skip connection that bypasses the nonlinear transformation, DenseNet adds a direct connection from any layer to any subsequent layer. Hence, a l layer receives the feature maps of all former layers from x_0 to x_l

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]),$$

where $[x_0, x_1, \dots, x_{l-1}]$ belongs to the spectrum of the feature map taken from the layers $0, 1, 2, \dots, l-1$.

Fig. 1 shows the architecture of a five-layer tightly coupled block.

Table 1 shows the proposed architecture of the DenseNet 169 network.

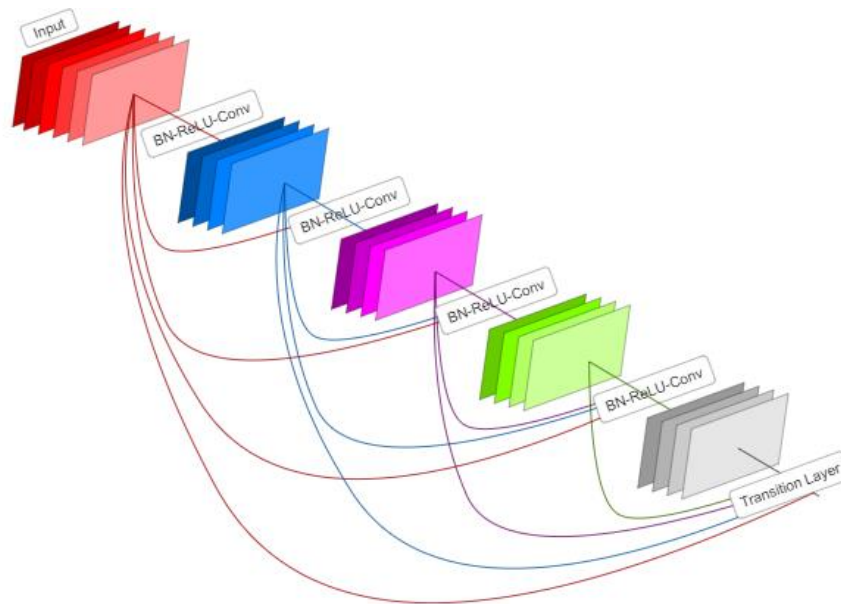


Fig. 1. Architecture of a five-layer densely connected unit of the neural network

Table 1. Architecture of DenseNet 169 network

Layers	Output size	DenseNet 169
Convolution	112×112	7×7 max. value, step 2
Association	56×56	3×3 convolution
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 6$
Transition layer (1)	56×56	1×1 convolution
	28×28	2×2 mean value, step 2
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 12$
Transition layer (2)	28×28	1×1 convolution
	14×14	2×2 mean value, step 2
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 32$
Transition layer (3)	14×14	1×1 convolution
	7×7	2×2 mean value, step 2
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 32$
Classification layer	1×1	7×7 total mean value, step 2
	1000	1000D full-coupled convolution

Let us build an algorithm for solving the problem of recognizing human emotions from a face image.

At the first stage, a set of data is collected for training and testing the classifier. In our case, we use FER2013 (Facial Expression Recognition 2013 Dataset) [16]. The dataset contains 35685 examples of 48×48 pixel grayscale face images divided into

training and testing sets. The images are classified based on the emotions manifested in facial expressions (happiness, neutrality, sadness, anger, surprise, disgust, fear).

The second stage involves data processing, i.e. splitting the dataset into training, validation and testing samples, as well as preprocessing the

images themselves: bringing them to the same size (48×48 pixels) and color model (RGB).

At the third stage, we build a neural network model. First, we use the previously trained on ImageNet dataset DenseNet169 for feature extraction. Next, we create a classifier, the set of layers of which is shown in fig. 2. We train the model on a previously prepared training set with the following parameters: optimization – stochastic gradient descent, loss function – cross entropy and

metric – accuracy (to assess the quality of the classifier). Also, in order to improve the accuracy of the model, we use one of the transfer learning methods – Fine Tuning ("pre-training", "fine tuning"). After all training processes, we evaluate the quality of the model using appropriate metrics.

As a result of this algorithm, a model will be built that will be able to recognize human emotions with high accuracy.

```
def classifier(inputs):
    x = tf.keras.layers.GlobalAveragePooling2D()(inputs)
    x = tf.keras.layers.Dense(256, activation="relu", kernel_regularizer = tf.keras.regularizers.l2(0.01))(x)
    x = tf.keras.layers.Dropout(0.3)(x)
    x = tf.keras.layers.Dense(1024, activation="relu", kernel_regularizer = tf.keras.regularizers.l2(0.01))(x)
    x = tf.keras.layers.Dropout(0.5)(x)
    x = tf.keras.layers.Dense(512, activation="relu", kernel_regularizer = tf.keras.regularizers.l2(0.01))(x)
    x = tf.keras.layers.Dropout(0.5)(x)
    x = tf.keras.layers.Dense(NUM_CLASSES, activation="softmax", name="classification")(x)

    return x
```

Fig. 2. Software implementation of the image classifier

Research results and their discussion

The program for solving problems of human emotion recognition is implemented in Python programming language. The code is completely written in one interactive web notebook with a runtime based on Google tensor processors on the Kaggle platform.

This application classifies human face images into seven different emotional states: happiness, neutrality, sadness, anger, surprise, disgust, fear. That is, the purpose of this program is to build and train a classifier that

would presumably tell us what emotion a person is depicted in the photo.

To train the classifier, we used the FER2013 dataset. We loaded it into the runtime environment, pre-processed the images, and generated training and test sets. We selected the optimal parameters, such as the number of training epochs – 30, the number of fine-tuning epochs – 20, the standard image size – 48×48 pixels and the batch size (number of objects in one iteration) – 64. We analyzed the training sample for distribution by classes, the visualization of the graph is shown in fig. 3.



Fig. 3. Distribution of images of the training sample by classes

Further in the program, based on the pre-trained neural network DenseNet169, a model for feature extraction was obtained. A classifier with tightly connected layers was built. To improve the accuracy of the model and reduce the probability of overfitting, the Fine Tuning method was used. After training the model, we evaluated the quality of classification on a test sample.

We used Pandas libraries for data processing, NumPy – for matrix computations, Matplotlib and Plotly – for visualization, scikit-learn and TensorFlow (Keras) – for deep learning model development.

To classify emotions, depending on the input data and compare the results, a deep learning model was built to which the faces of people in the image were applied. The model built for this study is based on a densely connected convolutional neural network (DCCN).

The quality of the classifier is evaluated by the following metrics: accuracy (fig. 4); confusion matrix (fig. 5); precision, recall, f1-score (fig. 6); ROC curve and AUC value (fig. 7).

The accuracy value is relatively high – 63%, provided that the dataset has unbalanced classes. AUC also has a high value – 89%.

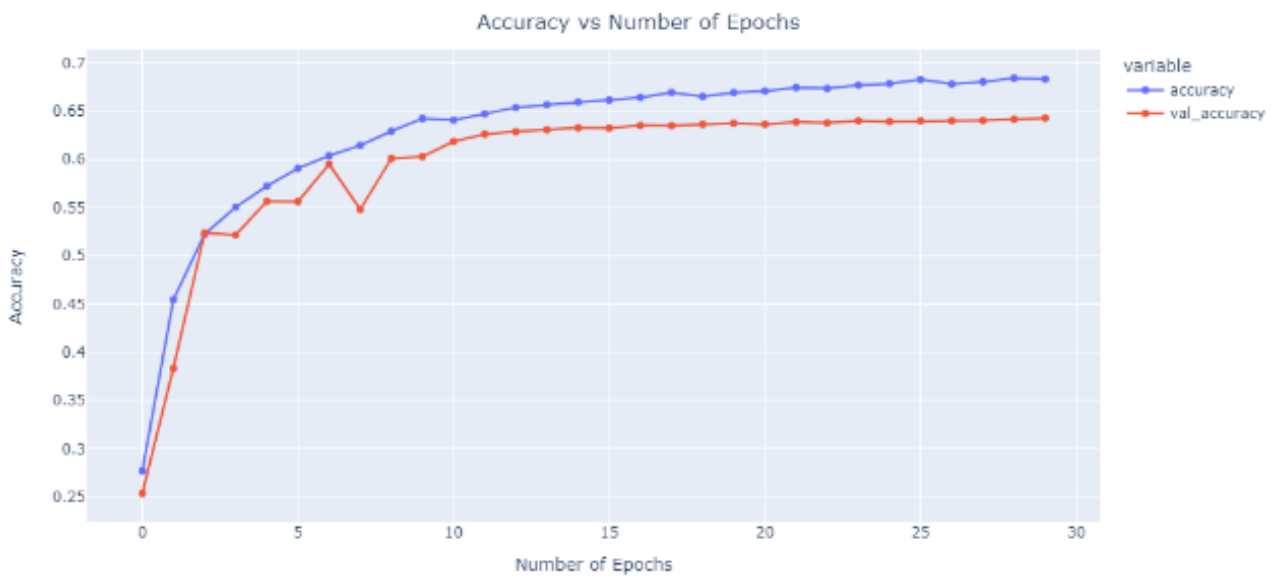


Fig. 4. Distribution of accuracy value by epochs

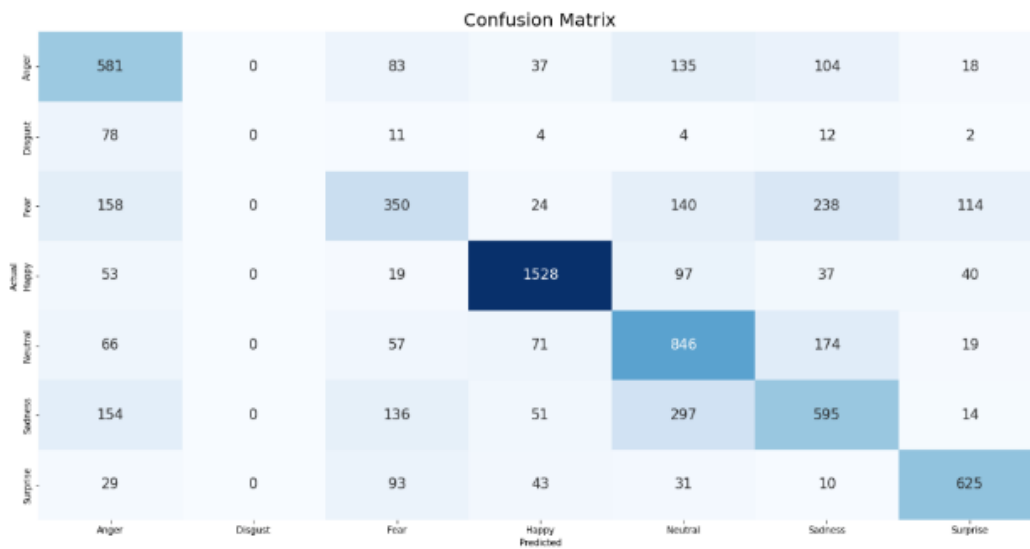


Fig. 5. Image of the matrix of inconsistencies

	precision	recall	f1-score	support
0	0.52	0.61	0.56	958
1	0.00	0.00	0.00	111
2	0.47	0.34	0.39	1024
3	0.87	0.86	0.87	1774
4	0.55	0.69	0.61	1233
5	0.51	0.48	0.49	1247
6	0.75	0.75	0.75	831
accuracy			0.63	7178
macro avg	0.52	0.53	0.52	7178
weighted avg	0.62	0.63	0.62	7178

Fig. 6. Values of precision, recall, f1-score metrics

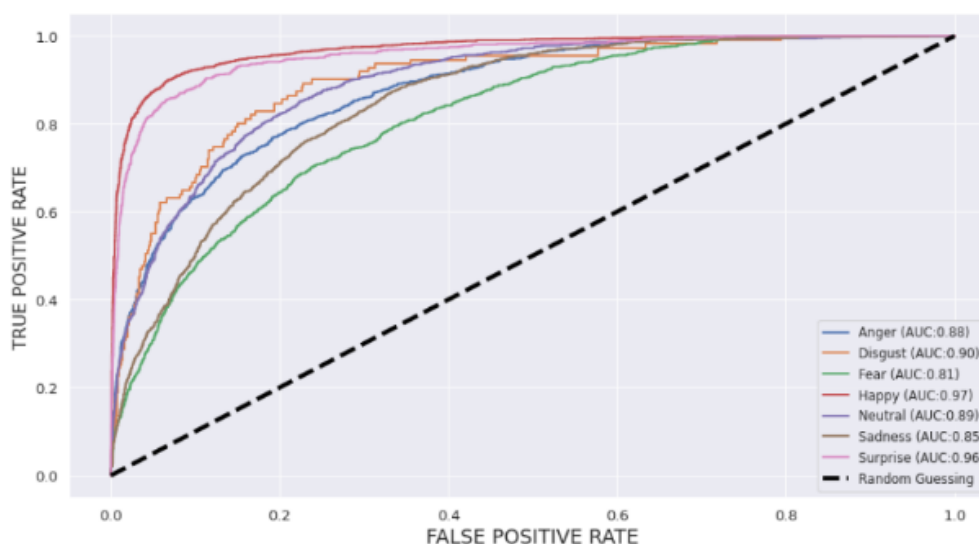


Fig. 7. Graph of the ROC curve

Conclusions and prospects for further development

In this paper, the methods of emotion classification are considered in detail and the best method for the purpose is chosen. Using the FER2013 dataset, the software implementation of the selected neural network was successfully trained and tested. The quality of the classifier was tested on a large number of different metrics: accuracy, precision, recall, f1-score, confusion matrix, ROC-AUC. Given the indicators, we can say that this work corresponds to the current level of scientific and technical knowledge in the field of deep learning.

References

- Hess, U., Thibault, P. (2009), "Darwin and Emotion Expression", *American Psychologist*, Vol. 64, No. 2, P. 120–129. DOI: <https://doi.org/10.1037/a0013386>

The relevance of the topic is growing every year and the solution of such problems is becoming important in many fields.

The results obtained can be used to improve the work of robots and bots to make their conversation with a person look more emotional. Emotion recognition can also be used to collect user feedback on any project or service in business. Thanks to the created system, it becomes possible to analyze in detail what emotions a particular product and service evokes in consumers and make more accurate forecasts for future sales and changes in the popularity of products and services.

2. Turabzadeh, S. (2018), "Facial expression emotion detection for real-time embedded systems", *Technologies*, No. 6 (1), 17 p. DOI: <https://doi.org/10.3390/technologies6010017>
3. Ekman, P. (1970), "Universal facial expressions of emotion", *California Mental Health Research Digest*, Vol. 8, No. 4, P. 151–158.
4. Viola, P., Jones, M. (2001), "Rapid object detection using a boosted cascade of simple features", *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 8–14 December, P. 1–9.
5. Coots, T., Edwards, G., Taylor, Ch. (2001), "Active Appearance Model", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, P. 681–685.
6. Coots, T., Taylor, K. (2004), "Statistical appearance models for computer vision", *Technical Report, University of Manchester, Wolfson Image Analysis Group, Imaging Science and Biomedical Engineering*, Manchester M13 9PT, United Kingdom, 125 p.
7. Van Dyke, Parunak H. (1995), *Neural Networks for Pattern Recognition*. New York: Christopher M. Bishop, Oxford University Press, 482 p.
8. Yashina, E., Artiukh, R., Pan, N., Zelensky, A. (2019), "Information technology for recognition of road signs using a neural network", *Innovative technologies and scientific solutions for industries*, No. 2 (8). DOI: <https://doi.org/10.30837/2522-9818.2019.8.130>
9. Mehryar, M., Rostamizadeh, A., Talwalkar, A. (2018), *Foundations of Machine Learning*, Cambridge: The MIT Press, 505 p.
10. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017), "Image Netclassification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60, No. 6, P. 84–90.
11. Huang, G., Sun, Y., Liu, Z. (2016), "Deep networks with stochastic depth", *European Conf. on Computer Vision (ECCV)*, Amsterdam, Netherlands, October, P. 646–661.
12. Srivastava, R. K., Greff, K., Schmidhuber, J. (2015), "Training very deep networks", *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, Montreal, Canada, December, P. 2377–2385.
13. He, K., Zhang, X., Ren, S. (2016), "Deep residual learning for image recognition", *The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 26 June – 1 July, P. 1–9.
14. Larsson, G., Maire, M., Shakhnarovich, G. (2017), "Fractalnet: ultra-deepneural networks without residuals", *5th International Conference on Learning Representations*, Toulon, France, April, P. 1–11.
15. Gao, H., Zhuang, L., Laurens van der Maaten, Kilian, W. (2017), "Weinberger Densely Connected Convolutional Networks", *The IEEE on Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July, P. 1–9.
16. FER2013 (Facial Expression Recognition 2013 Dataset), available at: <https://paperswithcode.com/dataset/fer2013> (last accessed 12.05.2022).

Received 26.09.2022

Відомості про авторів / Сведения об авторах / About the Authors

Ювченко Катерина Сергіївна – бакалавр зі спеціальності "Прикладна математика", Харківський національний університет радіоелектроніки, студентка кафедри прикладної математики, Харків, Україна, e-mail: kateryna.iuvchenko@nure.ua; ORCID ID: <https://orcid.org/my-orcid?orcid=0000-0002-5921-3049>

Ювченко Екатерина Сергеевна – бакалавр по специальности "Прикладная математика", Харьковский национальный университет радиоэлектроники, студентка кафедры прикладной математики, Харьков, Украина.

Yuvchenko Kateryna – Bachelor of Applied Mathematics, Kharkiv National University of Radioelectronics, Student of Department of Applied Mathematics, Kharkiv, Ukraine.

Єсілевський Валентин Семенович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри прикладної математики, Харків, Україна; e-mail: valentyn.yesilevskiy@nure.ua; ORCID ID: <https://orcid.org/0000-0002-5935-1505>

Єсілевский Валентин Семенович – кандидат технических наук, доцент, Харьковский национальный университет радиоэлектроники, доцент кафедры прикладной математики, Харьков, Украина.

Yesilevskiy Valentyn – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Applied Mathematics, Kharkiv, Ukraine.

Середя Олена Григорівна – Харківський національний університет радіоелектроніки, старший викладач кафедри прикладної математики, Харків, Україна; e-mail: olena.sereda@nure.ua; ORCID ID: <https://orcid.org/0000-0003-3579-0092>

Середя Елена Григорьевна – Харьковский национальный университет радиоэлектроники, старший преподаватель кафедры прикладной математики, Харьков, Украина.

Sereda Olena – Kharkiv National University of Radio Electronics, Senior Lecturer at the Department of Applied Mathematics, Kharkiv, Ukraine.

СИСТЕМА РОЗПІЗНАВАННЯ ЕМОЦІЙ ЛЮДИНИ З ВИКОРИСТАННЯМ АЛГОРИТМІВ ГЛИБОКОГО НАВЧАННЯ

Предметом дослідження є програмна реалізація нейронного класифікатора зображень. У статті розглянуто емоції як особливий вид психічних процесів, що виражають переживання людини, її ставлення до навколишнього світу та до себе. Емоційна сфера може виражатися різними способами: мімікою, позою, руховими реакціями, голосом. Проте найбільшу виразність має обличчя людини. Технології для розпізнавання людських емоцій використовуються для поліпшення обслуговування клієнтів, прийняття рішення після співбесіди з кандидатами, для збільшення емоційного впливу реклами тощо. Тому **метою** роботи є знаходження та оптимізація найбільш задовільного, з погляду точності, алгоритму класифікації емоцій людини із зображенням обличчя. У статті вирішуються такі **завдання**: огляд і аналіз сучасного стану "розпізнавання емоцій"; розгляд методів класифікації; вибір найкращого методу для поставленого завдання; розроблення програмної реалізації для класифікації емоцій; проведення аналізу роботи класифікатора, формулювання висновків про проведену роботу на основі отриманих даних. У статті також використовується метод класифікації зображень, оснований на щільно зв'язаній згортковій нейронній мережі. Отримані **результати** довели, що для вирішення завдань розпізнавання емоцій добре підходить метод класифікації зображень, оснований на щільно зв'язаній згортковій нейронній мережі, оскільки зазначений метод має досить високу точність. Оцінена якість класифікатора за такими метриками: accuracy; confusion matrix; precision, recall, f1-score; ROC-крива та значення AUC. Значення accuracy відносно високе – 63%, за умови, що набір даних має незбалансовані класи. AUC також має високе значення – 89%. **Висновки.** Отримана модель має високі показники розпізнавання емоцій людини та надалі може успішно використовуватися за призначенням.

Ключові слова: виявлення об'єктів; класифікація об'єктів; навчання з учителем; розпізнавання емоцій.

СИСТЕМА РАСПОЗНАВАНИЯ ЭМОЦИЙ ЧЕЛОВЕКА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ ГЛУБОКОГО ОБУЧЕНИЯ

Предметом исследования является программная реализация нейронного классификатора изображений. В работе рассмотрены эмоции как особый вид психических процессов, раскрывающих переживания человека, его отношения к окружающему миру и самому себе. Эмоции могут быть выражены разными способами: мимикой, позой, двигательными реакциями, голосом. Однако наибольшую выразительность имеет лицо человека. Технологии для распознавания человеческих эмоций используются фирмами для улучшения обслуживания клиентов, принятия решений о собеседовании с кандидатами и для оптимизации эмоционального воздействия рекламы. Поэтому **целью** работы является нахождение и оптимизация наиболее удовлетворительного, с точки зрения точности, алгоритма классификации эмоций человека с изображением лица. В статье решаются следующие **задачи**: обзор и анализ современного состояния задачи распознавания эмоций; рассмотрение методов классификации; выбор наилучшего способа для поставленной задачи; разработка программной реализации для классификации эмоций; проведение анализа работы классификатора, формулирование выводов о проделанной работе на основе полученных данных. В статье также используется метод классификации изображений, основанный на плотно связанной сверточной нейронной сети. **Результаты** показали, что для решения задач распознавания эмоций хорошо подходит метод классификации изображений, основанный на плотно связанной сверточной нейронной сети, поскольку он имеет достаточно высокую точность. Произведена оценка качества классификатора по следующим метрикам: accuracy; confusion matrix; precision, recall, f1-score; ROC-кривая и AUC. Значение accuracy относительно высокое – 63%, при условии, что набор данных имеет несбалансированные классы. AUC также имеет высокое значение – 89%. **Выводы.** Полученная модель имеет высокие показатели распознавания эмоций человека и в дальнейшем может успешно использоваться по назначению.

Ключевые слова: обнаружение объектов; классификация объектов; обучение с учителем; распознавание эмоций.

Бібліографічні опису / Bibliographic descriptions

Ювченко К. С., Єсілевський В. С., Серєда О. Г. Система розпізнавання емоцій людини з використанням алгоритмів глибокого навчання. *Сучасний стан наукових досліджень та технологій в промисловості*. 2022. № 3 (21). С. 60–69. DOI: <https://doi.org/10.30837/ITSSI.2022.21.060>

Yuvchenko, K., Yesilevskiy, V., Sereda, O. (2022), "Human emotion recognition system using deep learning algorithms", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (21), P. 60–69. DOI: <https://doi.org/10.30837/ITSSI.2022.21.060>

A. ANDRUSEVICH, S. OMAROV, N. STARODUBCEV, V. NEVLIUDOVA

STUDY OF THE INFLUENCE OF THE MODERN ROBOTISATION LEVEL ON THE CHALLENGES AND RISKS OF THE ECONOMIC SECURITY OF AN INDUSTRIAL ENTERPRISE

Industry 4.0 envisages a comprehensive transformation of all industries by combining digital technologies and the Internet with traditional industries. It is important to emphasize that Industry 4.0 relies heavily on robotics. Robotization opens up new opportunities for an industrial enterprise, completely transforming the technologies and organization of its production process and business model. Many of today's threats to the economic security of an industrial enterprise can be mitigated due to digitalization and flexible robotization of production and promotion of goods. But at the same time, processes related to the robotization of the economy create new threats to economic and social security. The consequences of their influence and the mechanisms of their leveling are the **subject** of research. The **goal** of the work is to study the impact of robotics on the economic security of an industrial enterprise. The following **tasks** are solved in the article: an overview of the areas of application and trends in the development of industrial robots; overview of the robotics market; analysis of risks and threats to the economic security of industrial enterprises arising as a result of robotization. The **methods** of system analysis, graphic generalization and classification are used. The **results** of the work include an analysis of the current state of robotics and the introduction of industrial robots in production, a brief analysis of the robotics market and a study of external challenges, uncertainties and risks arising from robotics: political-economic, technological, financial and environmental. **Conclusions.** The development of robotics is a priority direction for innovative industrialization not only in the long term, but also in the short term. Robotics can also become a locomotive for general economic growth, given that the most advanced developed countries are currently creating their own robotic enterprises and transferring production from developing countries to them. However, in the context of robotization, it should be taken into account that these processes, along with potential benefits for the industrial enterprise, can also create new threats to economic and social security.

Keywords: Industry 4.0; robotics; digitization; risks; economic security; industrial enterprise.

Introduction

Today, the creation and implementation of robotics is recognized as one of the priority areas of industrial development. More than a thousand modern companies develop and produce industrial work. All large companies are beginning to invest in the production of industrial robots. New companies specializing in these products are being created, as well as intermediary companies for the introduction of industrial robots. All developed countries have created national associations for industrial robotics, and in some countries, activities in this area are the priority of the state program.

The term Industry 4.0 means the fourth industrial revolution, which involves a comprehensive transformation of all industries by combining digital technology and the Internet with traditional manufacturing. All production links (suppliers, manufacturing companies, distributors, the product itself) are combined, based on the digitalization of processes, into an integrated value chain. Industry 4.0 envisions: a change in the production process (integration of machines, warehouse systems and production facilities into cyber-physical systems,

a significant increase in flexibility, efficiency and productivity in "smart" enterprises), and a change in the nature of work in enterprises (workers are expected to be freed from routine tasks and perform work that requires creativity) [1, 2]

The implementation of Industry 4.0 is based on the development of the following areas: information and communication technologies (ICT), providing digitalization of all stages of product creation and use both inside and outside companies; cyber-physical systems for control and management of physical processes and systems (in particular, integrated sensors, smart works, 3D-printed devices, etc.); network communications (wireless and Internet technologies for communication of machines, products, systems and people); modeling and virtualization in product development and production processes; systems for collecting, processing and using large amounts of information and cloud computing; augmented reality tools and intelligent systems [3].

As a result, it is expected to increase productivity by significantly reducing the time between the development of a new product and its delivery to the

consumer, increasing efficiency, saving energy, ensuring competitiveness in the global market, etc.

Central to Industry 4.0 is cooperation and work in the industries of the future, where new generations of robots with a high degree of artificial intelligence and humans will become equal partners. Industrial work will help workers with various tasks, and the quality of production and production processes will improve. In this case, different products can be produced on the same production equipment, making it possible to produce small batches thanks to the ability to quickly reconfigure the equipment [4].

However, with the current level of robotization of enterprises, it is necessary to take into account that these processes, along with the potential benefits for industry, can cause new threats to economic and social security.

Industrial robots and the advantages of their introduction into production.

The relevance of robotics

The number of robots installed in industry worldwide has more than tripled in the last ten years. Robotics is a new tool for the comprehensive mechanization and automation of production, a technique of the latest generation that gives high efficiency.

The industrial robot is another step in the development of flexible automation to optimize production with the ability not only to constantly repeat the same operations with guaranteed accuracy, but also with the possibility of easy reprogramming when the user's production program changes.

The concept begins with simple workstations where the robot is equipped with a positioner to place devices and position parts at two or more stations for an entire robotic production line, where the function of the devices, in particular the loading and unloading of parts, is handled by robots. Important helpers in the world of modern automation are auxiliary systems that are widely used, such as imaging systems or cameras that allow robots to remove and manipulate large parts.

However, the reliability of robots, their software, high productivity and ease of operation are necessary prerequisites for the proper functioning of these devices and systems.

The level and methods of production automation significantly depend on its type and scale, and if in mass and large-scale production the use of automatic lines is most justified, then in medium-scale, small-scale and single-piece production complex automation is implemented

by means of computer technology, machines with numerical program control. (NPC) and industrial robots.

On the basis of NPC technological equipment and industrial robots, multi-nomenclature lines, sections, workshops are composed, which are called flexible automated production. The basic principle of such flexible production is modularity. The automation of flexible production develops from simple to complex – first, flexible production modules (FPM) are created and implemented, on their basis flexible production complexes (FPC) are built, and, finally, flexible automated production (FAP). However, the reliability of robots, their software, high productivity and ease of operation are necessary prerequisites for the proper functioning of these devices and systems.

Their further development is practically unmanned automatic production, where flexible automated production is supplemented by systems of automated product design (APD) and technological preparation of their production, planning and dispatcher control (AEMS). The basic structural unit of flexible production modules of any complexity are robotic technological complexes (RTC), which can be formed on the basis of one industrial robot providing individual or group maintenance of the equipment connected to it or a complete product processing cycle (for example, welding), or on the basis of several industrial robots performing interrelated operations.

Versatility of the majority of industrial robots allows their wide application as a part of robotic technological complexes for different kinds of production.

Modern industrial robots can be used in many machining operations such as milling, grinding, polishing, trimming, deburring and cutting. They can be used in all industries and are also available to small and medium-sized businesses, allowing them to grow with a small investment.

Today, the most common area where industrial robots are used is in repetitive operations on production lines: welding, moving parts, assembly, painting, etc. On these lines, robots operate in a cyclic manner and carry out the same operations, replacing routine human labor. This makes it possible to increase labor productivity, reduce the factor of human error and injuries and automate the process as much as possible.

When creating parts, industrial robotic systems (IBS) can save a significant amount of materials and raw materials (with a rational organization of the work process) [5]. But this is not the limit: more expensive robotic systems can perform multi-axis movements along the desired path thanks to the six-step freedom of the

robot. This makes it possible to perform any machining operations that were previously only performed on specialized machines.

When properly organized, all of these operations are controlled by a single operator. The IBS makes it possible to carry out work where it is difficult and dangerous for humans. They are precise, error-free and fast. They can work 24 hours a day with the lowest lighting and the lowest temperature in the shop, thereby saving on utility costs [6].

Thus, the intensive use of industrial robotics at present and in the future is due to a number of reasons.

First of all, creation and wide introduction of industrial robots and manipulators, allowing intensifying various technological processes and operations, makes it impossible to use manual low-skilled and monotonous labor, especially in heavy, dangerous and hazardous conditions for humans.

In the coming years, a significant increase in output in industry must be ensured through the introduction of new types of machinery and progressive technologies. Although the share of manual labor in industry has declined, millions of people in the world are still engaged in manual labor today.

Numerous surveys of working conditions show that about 30% of workers are adversely affected by noise, 30% have to work according to a regimented schedule, 25% are exposed to damp, heat or cold, 20% work in physically uncomfortable positions or in conditions of smoke and fumes. % have to expend a lot of physical effort, and 15% work at night.

These stressors often act in combination, so about 40% of workers are simultaneously affected by two and about 25% by three or more factors. Accordingly, the use of robotics contributes to a significant reduction in the proportion of manual, heavy, harmful and tedious work (social factor). In addition, industrial robots have very high positioning accuracy, and repeatability allows achieving the desired level of product processing while reducing production defects. By avoiding the human factor in technological processes, the percentage of work errors is significantly reduced. This approach has a positive effect on the growth of enterprise performance and overall industry performance [7].

In addition, the nature of production has changed – about 80% of products are produced in small batches. Production automation is one of the significant levers to increase labor productivity in small batch production (economic factor).

Robotics makes it possible to solve the problems of two- and three-shift work, increase the equipment load factor and rhythm of its work, improve the quality of products and reduce their cost, especially in small-scale production.

Robotics creates prerequisites for transition to flexible automated production systems that allow for quick changeover for operations with a different sequence and nature of actions and with minimal human involvement.

The main technical characteristics of industrial robots are related to the sphere of application and production conditions for which they are designed.

Research shows that when used in individual operations, one industrial robot, depending on the variability of the work, replaces 1–3 workers, increases labor productivity by 60–80% and reduces the cost of production preparation by 45–50%.

When using industrial robots in groups, the efficiency of industrial robots increases dramatically: the productivity increases at least 3–5 times, and in some cases 8–10 times, capital investment and maintenance costs decrease, the intensity and rhythm of production, shiftiness, product quality increase, the number of defects decreases.

Among the directions of reduction of manual and heavy physical labor, in addition to robots, an important place also belongs to the simplest devices – manipulators as a means of complex mechanization of production.

In the production, where there is no need for human protection from the environment and frequent loading and unloading of equipment, manipulators with command control, characterized by the fact that a man-operator sequentially turns on the drives of each link, have become widespread.

Such robots-manipulators are the simplest in design relatively cheap and their use does not change the technological process, because they easily fit into the existing technology. Versatility, low cost and high efficiency on loading and unloading operations are their special qualities.

Many types of work, in particular, mechanical assembly, construction and finishing, lifting and transporting, storage and repair, can be mechanized in the near future just with the help of manipulators.

According to calculations, the satisfaction of industrial needs in the work of manipulators will reduce the number of manual labor in more than 30 professions: mechanics by 4%, repairers – by 3%, packers – by 5%, storekeepers – by 2.5%, transport operators – by 3% and loaders – by 5%.

As a result of innovations in networking technology, many sectors in manufacturing have only recently begun to embrace automation. Examples include the food and beverage, textile, wood and plastics industries.

Until recently, a number of safety and accident prevention measures accompanied the use of standard types of industrial robots. Therefore, the use of "collaborative" robots or "cobots" is a justified solution. Research and development of "collaborative" robots initially focused on the safety of use and at the same time the possibility of their integration into working lines with human operators.

The design of the robot has limited power and a function that immediately stops work when a collision is detected, which can be done in several ways. In many programs, this robot can be used without a guardrail.

Modern robotics manufacturers can offer their customers a type of robot that meets a clear requirement for its characteristics and at the same time has a number of other advantages, among which can be highlighted the following:

- simultaneous operation with a human operator;
- space saving;
- easy setup;
- high productivity
- accuracy;
- reliability.

Instead of a worker performing the same task for many hours, it is relatively easy to implement a "collaborative" job in its place due to simple programming and setup and without all of the safeguards required for traditional industrial robots. For the same reasons, the "collaborative" robot is also much more affordable (not only the cost of the job is lower, but also the installation cost, which is achieved by eliminating the time required for technical support and configuration of the "robotic cell" equipment), and therefore it is easier to justify financially.

The best examples of successful implementations of such robots are manufacturing facilities where there are multiple stations performing the same type of process, such as manufacturing processes with multiple NPC machines.

Robotics market survey

Recent growth trends in the robotics market are due to its significant penetration into the economy, particularly in the service sector; technical improvements

in "collaborative" robots designed to interact with humans in a shared workspace; and cheaper manufacturing of robots. As robotization accompanies the digital transformation of the enterprise, the digitalization of the economy also stimulates the development of robotics. At the same time, robotization and the widespread use of robots in the enterprise entails new risks to its security – human, technical-technological, financial, etc. The economic security of an industrial enterprise is understood as such a state of its production, labor, management, information, creative structures, which provides a timely and adequate response of the enterprise to the emergence and development of external and internal risks and threats.

Depending on the field of application, it is customary to distinguish between industrial and service robotics. At present, this division is rather arbitrary due to the mutual penetration of industrial and service technologies. To assess the dynamics of the industrial robotics market, the indicator of the number of installed industrial robots over a certain period of time is used. According to estimates of the International Federation of Robotics, 422,000 industrial robots were installed worldwide in [8], 2018, which is 6% more than in 2017. Although there has been an increase in this number since 2012, its rate is not as high and steady as many analysts expected. Experts are unanimous that the market for industrial robots will grow; only the estimates of the rate of this growth as well as its structure are different. For example, experts' expectations that one of the key robotics trends will be wide implementation of "collaborative" robots ("cobots") did not come true.

Asia is the largest and fastest growing market for industrial robots. The share of this region in global demand is about 68%, while Europe's share is 18% and America's is 14% (fig. 1).

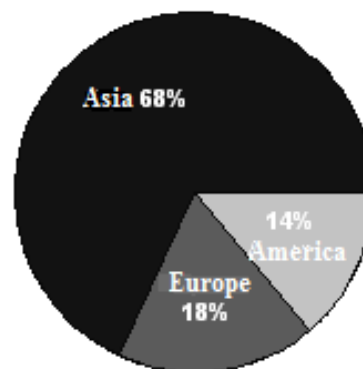


Fig. 1. Structure of the industrial robots market

According to forecasts by experts from the International Federation of Robotics [9], further growth in demand for industrial robots will be provided by the Asian region, primarily China. Regarding the breakdown by country, 74% of industrial robotics sales in 2018 came from five countries: China (154,000 units), Japan (55,200 units), the United States (40,400 units), South Korea (37,800 units) and Germany (26,700 units). Taiwan, Italy, France, Mexico, Spain, India, Singapore, Canada, Thailand and the Czech Republic are followed by a large margin (fig. 2).

Most of them are also leaders in the indicator characterizing the intensity of the use of industrial robots – robot density, that is, the number of robots per 10,000 workers. In 2018, the highest value of this indicator came from Singapore and South Korea (831 and 774 robots per 10,000 workers, respectively). Germany (338 robots per 10,000 workers), Japan (327), Sweden (247), Denmark (240), Taiwan (221), the United States (217), and Italy (200) followed by a wide margin (Table 1). At the bottom of the ten leaders

is Belgium with 188 robots per 10,000 workers, and in China this indicator reached 140. Note that the global average is 99 robots per 10,000 workers.

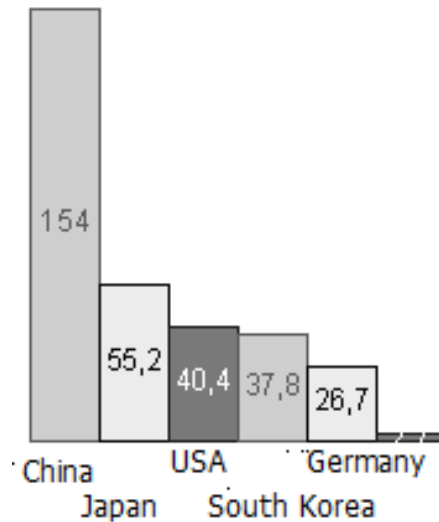


Fig. 2. Volumes of industrial robotics sales by leading countries

Table 1. "Top ten" countries in the intensive use of industrial robots

Ranking	Country	Number of robots per 10,000 workers
1	Singapore	831
2	South Korea	774
3	Germany	338
4	Japan	327
5	Sweden	247
6	Denmark	240
7	Taiwan	221
8	USA	217
9	Italy	200
10	Belgium	188

Among the main areas of the economy in which industrial robots are in demand are the automotive industry, which accounts for about 30% of global demand, electronics manufacturing (25%), machine building, metallurgy, chemical and pharmaceutical industries, and the food industry, which has shown a steady growth in its share in recent years. The industrial sector has high potential for further automation and robotization, as a number of industries still have a high share of manual labor. However, the ratio of the average cost of manual to robotic labor reduces the incentive to robotize industry.

Risks and threats to the economic security of industrial enterprises caused by robotization

Along with the increasing opportunities arising from robotization, there are also growing threats, primarily to the economic security of the industrial enterprise [10]. The introduction of intelligent robots is a crucial part of the digitalization of industry, which faces many external challenges related to uncertainty and risk:

- political-economic,
- technological,
- financial,
- environmental.

These risks are becoming pervasive, many of them unpredictable. As an example, the consequences of a pandemic are changing the social and economic landscape of today's world. The mass suspension of production, the shift to telecommuting, the disruption of supply chains (especially international supply chains), and the movement of some business processes into the digital world have demonstrated the benefits of robotization.

While experts call sustainable competitive advantage the main factor determining the state of a firm's economic security, and nowadays successful enterprises are those that in these conditions preserve production and ensure the safety of workers – online services, autonomous production and warehouse complexes, etc. [11].

The service robotics segment has received a new impetus for development. Many countries have faced a shortage of personnel providing care for the sick and elderly, as well as problems with ensuring the safety of workers. Compensating for their shortage by replacing them with robots offers benefits for the safety of staff and people in need of care. Robots can deliver food, medications, and take temperature and pulse. Their use provides safety for medical personnel by limiting the possibility of spreading a virus, since robot bodies are easier to disinfect and cannot be carriers of disease. In other areas of the economy, the demand for non-contact service delivery by service robots is also increasing. Businesses that are unable to adapt to the new environment face the risk of economic loss.

Consequently, one of the key competitive advantages of the industrial enterprise in modern conditions is its adaptability. Flexible automation of production based on the introduction of industrial robots contributes to the adaptability of enterprises through rapid retooling and changing programs of the means of production. Industrial automation uses both robots that have a material shell corresponding to their functions and those that do not.

The processes of digitalization, automation and robotization of industry cover all stages of production, from obtaining and analyzing data on trends emerging in the markets, automated design and engineering, robotization of logistics and the production process itself, to digital positioning and sales tools. These processes are associated with an increasing number and variety of external and internal risks and threats to the industrial enterprise, and the effective robotization and digitalization of foreign competitors generate threats

to the domestic real sector. Many challenges and threats to national and economic security can be timely detected and neutralized at the enterprise level. Threats to the economic security of various industrial enterprises, acquiring a stable, regularly recurring nature, it is legitimate to consider as potential threats to the economic security of the industry, and in conditions of a developed system of division and cooperation of labor – and as potential threats to the economic security of the country.

The key problem in the context of robotics economy scientists and experts call the impact of new technologies on labor markets. The transformation of labor relations and changes in the structure of employment as a result of mass automation and robotization occupy a significant place in the scientific economic discourse [12]. On the one hand, robots, replacing human labor, leave a large number of workers without work, forcing them to change their profession or place of work. The literature on this point notes that "if the use of robots continues to expand and automation increases, given advances in artificial intelligence research, worker participation is expected to be limited to making decisions about plant prospects within a human-designed strategy" [12]. Among the advantages of replacing manual labor with robotic labor the most frequently cited is that robots can work virtually without interruption, do not need vacations, sick leave, cannot quit, require higher wages, comfortable working conditions (temperature, lighting, convenient location, etc.), and are flexible in terms of the volume of products produced.

In addition, robotization is stimulating the emergence of a significant number of new professions and jobs in science, engineering, and analytics. The growing dynamism and flexibility of labor markets requires the workforce to adapt to new conditions. In order to maintain their competitiveness a person must improve their professional competencies throughout their life. Moreover, high qualifications do not guarantee sustainable competitive advantages. Along with the developed professional skills, a modern worker is expected to have a set of personal qualities, which allow effective interaction with colleagues and partners in the process of labor activity. All these requirements of the labor market necessitate the configuration of approaches to personnel training.

Mass robotization entails a likely polarization of jobs, with medium-skilled workers at risk of being displaced. Robotization of functions requiring low-skilled labor is less economically advantageous,

because the average cost of manual labor is lower or at the same level (depending on the country and industry) as the average cost of robotic labor. Highly skilled labor is still difficult to replace with robots and artificial intelligence. In the end, this creates risks of increasing property differentiation in society. "The likely result of robotization could be an economy where high wages go to a few people with exceptional talent, while most other workers receive low wages" [13]. In terms of the evolution of economies and labor markets, changes in employment patterns can be expected. "Opportunities to reduce risk and maximize profits as a result of technological advances depend on accurate forecasting of what skills will be needed in the future, as well as the alignment of educational policies and strategies" [13]. However, at the macro- and micro-levels, transformations in the structure of employment will be accompanied by social instability and changes in income distribution. In this regard, C. Webster and S. Ivanov rightly emphasize the presence of reasonable concerns "among researchers, politicians and representatives of the real sector about how people, companies, economies, governments and society should adapt to the new technological, economic, social and political realities that robonomics will create" [14]. This leads to social risks related to the ethical limitations of the use of robots, the increase of social class differentiation, the competition for jobs not "man with man" but "man with robot".

Changes in the structure of employment at the micro level, consisting in the transformation of the organizational structures of the enterprise, leads to the emergence of personnel risks. Personnel risks appear at two stages: first, at the stage of providing enterprises with specialists possessing modern competences that meet the modernization policy of the enterprise; second, at the stage of retention of such specialists. Ensuring personnel security is one of the most important components of the economic security of an industrial enterprise.

Robotization and digitalization transform not only economic, industrial and labor relations, but also the entire set of social relations, namely consumer behavior, which is becoming more varied due to the influence of new social and functional innovations, reducing the effectiveness of traditional mechanisms of promotion, in particular advertising. Complication of the management mechanisms of consumer choice requires the most adequate response from the producers. With the demand now being personalized, the manufacturer is forced to determine the features of

the manufactured product in direct coordination with the consumer. Information service providers, who create platforms for contact between producers and consumers, are beginning to play a role no less than traditional trade intermediaries do. IT-companies create new structures in the society and in the economy, in which users are grouped on different grounds. In such structures new nodes of relations have taken an obvious form around which the global economic networks have developed [15, 16]. Such platforms are electronic platforms for the promotion and sale of goods, social networks, virtual systems, which allow partially automating the process of consumer choice. This makes it necessary to take measures to integrate virtual elements into real production and business processes. Information technology is a tool to optimize their efficiency by increasing flexibility and responsiveness to consumer demands and changing trends (behavioral, technological, raw material trends).

The digitalization and robotization of the industry, with the accelerated development of information and communication technologies and the integration of virtual and real business processes, cause a shift in the centers of profit capitalization in value chains. If in the second half of XX century observed their movement from the direct material production to design, marketing and logistics, then they are shifting to the creation and maintenance of computer systems to manage business processes and process control systems use of big data. The acceleration of production cycles necessitates the intensification of product development processes, management decision-making, raw material renewal, innovation activity, interaction with partners and improvement of information exchange within the enterprise and with the external environment. Cost-effective shortening of production cycles and increasing the flexibility of production systems based on digitalization, automation and flexible robotics will enable the transition to small batch production and frequent renewal of the assortment.

At present, the undermining of the competitiveness of industrial enterprises, the disruption of the normal economic reproduction cycle can be carried out without explicitly violating the laws. In the conditions of expansion of information space, elimination of economic relations in the Internet and transformation of economic competition with the help of socially functional technologies aimed at the task of the enterprise-competitor both economic and reputational losses, the risks of assumption of various forms of unfair

competition (discredit, delusion, incorrect comparison, illegal use of results of intellectual activity, creating displacement, illegal actions with secrets protected by law) increase. Along with the listed forms of unfair competition prohibited by law, in practice there are common situations when: firstly, formally the actions of a competitor do not relate to prohibited forms of competition or in practice it is not proved (for example, implicitly discrediting a competitor, business intelligence); secondly, it is difficult to identify the source of aggression (for example, many negative reviews of "consumers" of goods / services, spread negative opinion about the company's products, particularly by bots forming public opinion) and, consequently

Digitalization and robotization of industry entail the emergence of fundamentally new risks associated with the spread of the industrial Internet of Things, which increases the risk of cyber attacks aimed at disrupting production processes and illegally obtaining commercial information (industrial espionage) [17, 18]. Additional risks arise from the use of industrial works, mainly imported production. In the process of using foreign software solutions in control systems there is a risk of vulnerabilities in them that can be used for cyber attacks, which entails economic losses.

Consequently, the current level of robotization poses a threat to the emergence of such social risks affecting the economic security of the enterprise:

- changes in labor markets as a result of new technologies;
- increase of property differentiation in society;
- social risks associated with moral constraints, with an increase in social class differentiation and competition;

References

1. Popkova, E. G., Ragulina, Y. V., Bogoviz, A. V. (2019), *Industry 4.0: Industrial Revolution of the 21st Century*, Springer International Publishing, Cham, 249 p.
2. Nevliudov, I., Omarov, M., Yevsieiev, V., Bronnikov, A., Lyashenko, V. (2020), "Method of Algorithms for Cyber-Physical Production Systems Functioning Synthesis", *International Journal of Emerging Trends in Engineering Research*, Vol. 8 (10), P. 7465–7473.
3. Skitsko, V. I. (2016), "Industry 4.0 as industrial production of the future" ["Industriia 4.0 yak promyslove vyrobnytstvo maibutnoho"], *Investment: practice and experience*, No. 5, P. 33–40.
4. Stoun, D. "Robotization of production in the world: scope, examples, pros and cons" ["Robotizatsiya proizvodstva v mire: sfera primeneniya. primery. plyusy i minusy"], available at: <https://fb.ru/article/406125/robotizatsiya-proizvodstva-v-mire-sfera-primeneniyaprimeryi-plyusy-i-minusyi>
5. Khafizov, I. I., Nurullin, I. G. (2019), "Influence of application of robotic systems on prospects and directions of development of electroerosive processings", *IOP Conference Series: Materials Science and Engineering*, P. 570. DOI: 10.1088/1757-899X/570/1/012049
6. Kashapov, N. F., Khafizov, I. I., Nurullin, I. G. (2018), "Influence of introduction of robotics on increase in efficiency of electrochemical production", *IOP Conference Series: Materials Science and Engineering*, Vol. 412, P. 436. DOI: 10.1088/1757-899X/412/1/012034

- personnel risks caused by the transformation of the organizational structures of the enterprise;
- risks of assumption of different forms of unfair competition;
- risks of cyber-attacks.

Conclusion

The development of robotics is a priority for innovative industrialization not only in the long term, but also in the short term. Robotics can also become an engine for overall economic growth, given that the most advanced developed countries are now creating robotic enterprises and relocating production from developing countries.

Many threats to the economic security of the industrial enterprise can be mitigated by digitalization and the flexible robotization of production and promotion. However, in a robotized economy, it must be borne in mind that these processes, along with the potential benefits of the industrial enterprise, may cause new threats to economic and social security.

The article analyzes the current state of robotics and the introduction of industrial robots into production. A systematic review of the current world market of robotics is made. External challenges, uncertainties and risks arising from robotization are investigated. Also the identification of partial socio-economic risks, their causes and their impact on the economic security of the industrial enterprise was carried out.

In further research, a qualitative and quantitative analysis of the risks caused by robotization will be conducted and recommendations for their elimination will be given.

7. International Federation of Robotics, available at: <https://ifr.org/>
8. Khoros, V. (2017), "Different consequences of robotization" ["Razlichnyye posledstviya robotizatsii"], *World economy and international relations*, Vol. 61, No. 12, P. 82–88.
9. Shirayayev, A. (2018), "About the effects of robotization production" ["Ob effektakh robotizatsii proizvodstva"], *Society and Economics*, No. 5, P. 59–67.
10. Lakshmia, V., Bahlib, B. (2020), "Understanding the robotization landscape transformation: A centering resonance analysis", *Journal of Innovation & Knowledge*, No. 5, P. 59–67.
11. Webster, C., Ivanov, S. (2020), "Robotics, artificial intelligence, and the evolving nature of work", *Digital transformation in business and society theory and cases*, ed. by B. George, J. Paul. Palgrave-MacMillan, P. 127–130. DOI: 10.1007/978-3-030-08277-2_8
12. Topishko, N. P., Topishko, I. I., Galetskaya, T. I. (2018), "Network economics as a form of socio-economic organization in global coordinates" ["Merezhiva ekonomika yak forma sotsialno-ekonomichnoï organizatsii v globalnikh k,ordinatakh"], *Scientific notes of the National University "Ostroh Academy"*, No. 8 (36), P. 33–38.
13. Zaluzhnyi, A. L. (2020), "Information and network measurements of the modern economy" ["Informatsiyno-merezhivi vimiri suchasnoï ekonomiki"], *Black Sea economic studios*, Issue 50-1, P. 59–63.
14. Andreyev, Yu. S., Dergachev, A. M., Zharov, F. A., Sadyrin D. S. (2019), "Information security of automated process control systems" ["Informatsionnaya bezopasnost avtomatizirovannykh sistem upravleniya tekhnologicheskimi protsessami"], *Ed. universities. Instrumentation*, Vol. 62, No. 4, P. 331–339.
15. Seksembayeva, M. A. (2019), "Industrial Internet of Things Security Features" ["Osobennosti obespecheniya bezopasnosti v promyshlennom Internete veshchey"], *E-Scio*, No. 5 (32), P. 13.

Received 26.09.2022

Відомості про авторів / Сведения об авторах / About the Authors

Андрусевич Анатолій Олександрович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри комп'ютерно-інтегрованих технологій, автоматизації та мехатроніки, Харків, Україна; e-mail: anatolii.andrusevych@nure.ua; ORCID ID: <https://orcid.org/0000-0002-3142-635X>

Андрусевич Анатолій Александрович – доктор технических наук, профессор, Харьковский национальный университет радиоэлектроники, профессор кафедры компьютерно-интегрированных технологий, автоматизации и мехатроники, Харьков, Украина.

Andrusevich Anatoliy – Doctor of Engineering Science, Professor, Kharkiv National University of Radio Electronics, Professor of Department of Computer-Integrated Technologies, Automation and Mechatronics, Kharkiv, Ukraine.

Омаров Шахин Анвер оглы – доктор економічних наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри комп'ютерно-інтегрованих технологій, автоматизації та мехатроніки, Харків, Україна; e-mail: shakhin.omarov@nure.ua; ORCID ID: <https://orcid.org/0000-0002-2887-9083>

Омаров Шахин Анвер оглы – доктор экономических наук, доцент, Харьковский национальный университет радиоэлектроники, профессор кафедры компьютерно-интегрированных технологий, автоматизации и мехатроники, Харьков, Украина.

Omarov Shakhin – Doctor of Economics, Associate Professor, Kharkiv National University of Radio Electronics, Professor of Department of Computer-Integrated Technologies, Automation and Mechatronics, Kharkiv, Ukraine.

Стародубцев Микола Григорович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри комп'ютерно-інтегрованих технологій, автоматизації та мехатроніки, Харків, Україна; e-mail: nikolaj.starodubcev@nure.ua; ORCID ID: <https://orcid.org/0000-0001-7856-5771>

Стародубцев Николай Григорьевич – кандидат технических наук, доцент, Харьковский национальный университет радиоэлектроники, доцент кафедры компьютерно-интегрированных технологий, автоматизации и мехатроники, Харьков, Украина.

Starodubcev Nikolaj – Candidate of Engineering Science, Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor of Computer-Integrated Technologies, Automation and Mechatronics Department, Kharkiv, Ukraine.

Невлюдова Вікторія Валеріївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри комп'ютерно-інтегрованих технологій, автоматизації та мехатроніки, Харків, Україна; e-mail: viktoriaa.nevliudova@nure.ua; ORCID ID: <https://orcid.org/0000-0002-1158-5089>

Невлюдова Виктория Валерьевна – кандидат технических наук, доцент, Харьковский национальный университет радиоэлектроники, доцент кафедры компьютерно-интегрированных технологий, автоматизации и мехатроники, Харьков, Украина.

Nevliudova Viktoriia – Candidate of Engineering Science, Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor of Computer-Integrated Technologies, Automation and Mechatronics Department, Kharkiv, Ukraine.

ДОСЛІДЖЕННЯ ВПЛИВУ РІВНЯ СУЧАСНОЇ РОБОТИЗАЦІЇ НА ВИКЛИКИ ТА РИЗИКИ ЕКОНОМІЧНОЇ БЕЗПЕКИ ПРОМИСЛОВОГО ПІДПРИЄМСТВА

Industry 4.0 передбачає комплексну трансформацію всіх галузей промисловості шляхом поєднання цифрових технологій та інтернету з традиційними виробництвами. Важливо наголосити, що Industry 4.0 багато в чому спирається на роботизацію. Роботизація відкриває нові можливості для промислового підприємства, повністю трансформуючи технології та організацію його виробничого процесу й бізнес-модель. Багато з наявних загроз для економічної безпеки промислового підприємства можуть бути пом'якшені завдяки цифровізації та гнучкої роботизації виробництва та просування товарів. Але одночасно процеси, пов'язані з роботизацією економіки, спричиняють нові загрози для економічної та соціальної безпеки. Наслідки їхнього впливу й механізми їхнього нівелювання є предметом дослідження. Метою роботи є дослідження впливу роботизації на економічну безпеку промислового підприємства. У статті вирішуються такі завдання: огляд сфер застосування й тенденцій розвитку промислових роботів; огляд ринку робототехніки; аналіз ризиків та загроз для економічної безпеки промислових підприємств, що виникають унаслідок роботизації. Застосовуються методи системного аналізу, графічного узагальнення та класифікації. Результати дослідження містять: аналіз сучасного стану робототехніки та впровадження промислових роботів на виробництві, стислий аналіз ринку робототехніки й вивчення зовнішніх викликів, невизначеностей і ризиків, що виникають унаслідок роботизації, – політико-економічних, технологічних, фінансових та екологічних. Висновки. Розвиток робототехніки є пріоритетним напрямом для інноваційної індустріалізації не тільки в довгостроковій, але й у короткостроковій перспективі. Робототехніка може стати локомотивом і для загального економічного зростання, урахувавши, що найбільш передові та розвинені країни нині створюють у себе роботизовані підприємства й переводять до себе виробництва з країн, що розвиваються. Однак в умовах роботизації необхідно враховувати, що ці процеси, одночасно з потенційними вигодами для промислового підприємства, можуть спричинити й нові загрози для економічної та соціальної безпеки.

Ключові слова: Industry 4.0; роботизація; цифровізація; ризики; економічна безпека; промислове підприємство.

ИССЛЕДОВАНИЕ ВЛИЯНИЯ УРОВНЯ СОВРЕМЕННОЙ РОБОТИЗАЦИИ НА ВЫЗОВЫ И РИСКИ ЭКОНОМИЧЕСКОЙ БЕЗОПАСНОСТИ ПРОМЫШЛЕННОГО ПРЕДПРИЯТИЯ

Industry 4.0 подразумевает комплексную трансформацию всех отраслей промышленности путем сочетания цифровых технологий и интернета с традиционными производствами. Важно подчеркнуть, что Industry 4.0 во многом опирается на роботизацию. Роботизация открывает новые возможности для промышленного предприятия, полностью трансформируя технологии и организацию его производственного процесса и бизнес-модель. Многие из существующих сегодня угроз экономической безопасности промышленного предприятия могут быть смягчены за счет цифровизации и гибкой роботизации производства и продвижения товаров. Но одновременно процессы, связанные с роботизацией экономики, порождают новые угрозы экономической и социальной безопасности. Последствия их влияния и механизмы их нивелирования являются предметом исследования. Цель работы – исследование влияния роботизации на экономическую безопасность промышленного предприятия. В статье решаются следующие задачи: обзор областей применения и тенденций развития промышленных роботов; обзор рынка робототехники; анализ рисков и угроз экономической безопасности промышленных предприятий, возникающих в результате роботизации. Используются методы системного анализа, графического обобщения и классификации. Результаты включают в себя анализ современного состояния робототехники и внедрение промышленных роботов на производстве, краткий анализ рынка робототехники и исследования внешних вызовов, неопределенностей и рисков, возникающих вследствие роботизации, – политико-экономических, технологических, финансовых и экологических. Выводы. Развитие робототехники является приоритетным направлением для инновационной индустриализации не только в долгосрочной, но и в краткосрочной перспективе. Робототехника может стать локомотивом и для общего экономического роста, учитывая, что наиболее передовые и развитые страны в настоящее время создают у себя роботизированные предприятия и переводят к себе производства из развивающихся стран. Однако в условиях роботизации следует учитывать, что эти процессы, одновременно с потенциальными выгодами для промышленного предприятия, могут породить и новые угрозы экономической и социальной безопасности.

Ключевые слова: Industry 4.0; роботизация; цифровизация; риски; экономическая безопасность; промышленное предприятие.

Бібліографічні описи / Bibliographic descriptions

Андруевич А. О., Омаров Ш. А., Стародубцев М. Г., Невлюдова В. В. Дослідження впливу рівня сучасної роботизації на виклики та ризики економічної безпеки промислового підприємства. *Сучасний стан наукових досліджень та технологій в промисловості*. 2022. № 3 (21). С. 70–79. DOI: <https://doi.org/10.30837/ITSSI.2022.21.070>

Andrusevich, A., Omarov, S., Starodubcev, N., Nevliudova, V. (2022), "Study of the influence of the modern robotisation level on the challenges and risks of the economic security of an industrial enterprise", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (21), P. 70–79. DOI: <https://doi.org/10.30837/ITSSI.2022.21.070>

АЛФАВІТНИЙ ПОКАЖЧИК

Андрусевич А. О.	70
Гобов Д. А.	5
Єсілевський В. С.	16, 60
Замірець Я. О.	51
Колядін А. В.	16
Корабльов М. М.	26
Косенко В. В.	51
Лутай Л. М.	51
Луцький С. В.	26
Невлюдова В. В.	70
Омаров Ш. А.	70
Середа О. Г.	16, 60
Соловей О. Л.	39
Стародубцев М. Г.	70
Федорович О. Є.	51
Ювченко К. С.	60

ALPHABETICAL INDEX

Andrusevich Anatoliy	70
Gobov Denys	5
Yesilevskiy Valentyn	16, 60
Zamirets Iaroslav	51
Koliadin Anton	16
Korablyov Mykola	26
Kosenko Viktor	51
Lutai Liudmyla	51
Lutskyy Sergey	26
Nevliudova Viktoriia	70
Omarov Shakhin	70
Sereda Olena	16, 60
Solovei Olga	39
Starodubcev Nikolaj	70
Fedorovich Oleg	51
Yuvchenko Kateryna	60

НАУКОВЕ ВИДАННЯ

**СУЧАСНИЙ СТАН
НАУКОВИХ ДОСЛІДЖЕНЬ ТА ТЕХНОЛОГІЙ
В ПРОМИСЛОВОСТІ**

Щоквартальний науковий журнал

№ 3 (21), 2022

Відповідальний секретар журналу *І. Г. Перова*
Відповідальний за випуск *А. А. Коваленко*
Комп'ютерна верстка *Л. Ю. Светайло*

АДРЕСА РЕДАКЦІЇ:

буд. 14, проспект Науки, м. Харків, Україна, 61166

Тел.: +38 (057) 704-10-51

Веб-сайт: <http://itssi-journal.com>

E-mail: journal.itssi@gmail.com

Формат 60×84/8. Умов. друк. арк. 12,83. Тираж 150 прим.

Віддруковано з готових оригінал-макетів в типографії ФОП Андреев К.В.
Єдиний державний реєстр юридичних осіб та фізичних осіб-підприємців.
Запис №24800170000045020 від 30.05.2003.

61157, Харків, вул. Акад. Богомольця, 9, кв. 50, тел. +38 (063) 993-62-73
e-mail: ep.zakaz@gmail.com