ISSN 2587-814X

BUSINESS INFORMATICS

БИЗНЕС-ИНФОРМАТИКА

Vol. 19 No 1 — 2025

RECOMMENDATION SYSTEM BASED ON TECHNICAL EVENTS

INTELLIGENT TRANSPORT SYSTEM

THE PROGRAM OF IMPROVING THE EFFICIENCY OF BUSINESS PROCESSES

INDICATORS OF AN ORGANIZATION'S INTELLECTUAL CAPITAL

ISSN 2587-814X (print), ISSN 2587-8158 (online)
Russian version: ISSN 1998-0663 (print), ISSN 2587-8166 (online)

Vol. 19 No. 1 – 2025

BUSINESS INFORMATICS

HSE Scientific Journal

CONTENTS

Publisher:
HSE University

The journal is published quarterly

The journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Russian Federation

Address:
26-28, Shabolovka Street,
Moscow 119049, Russia

Tel./fax: +7 (495) 772-9590 *28509
http://bijournal.hse.ru
E-mail: bijournal@hse.ru

Circulation:
English version – 100 copies,
Russian version – 100 copies,
online versions in English and Russian – open access

Printed in HSE Printing House
44, build.2, Izmaylovskoye Shosse,
Moscow, Russia

ABOUT THE JOURNAL

Business Informatics is a peer reviewed interdisciplinary academic journal published since 2007 by HSE University, Moscow, Russian Federation. The journal is administered by HSE Graduate School of Business. The journal is issued quarterly, in English and Russian.

The mission of the journal is to develop business informatics as a new field within both information technologies and management. It provides dissemination of latest technical and methodological developments, promotes new competences and provides a framework for discussion in the field of application of modern IT solutions in business, management and economics.

The journal publishes papers in the following areas: modeling of social and economic systems, digital transformation of business, innovation management, information systems and technologies in business, data analysis and business intelligence systems, mathematical methods and algorithms of business informatics, business processes modeling and analysis, decision support in management.

The journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Russian Federation.

The journal is included into Scopus, Web of Science Emerging Sources Citation Index (WoS ESCI), Russian Science Citation Index on the Web of Science platform (RSCI), EBSCO.

The journal is distributed both in printed and electronic forms.

EDITORIAL BOARD

ABOUT HSE UNIVERSITY

Consistently ranked as one of Russia's top universities, HSE University is a leader in Russian education and one of the preeminent economics and social sciences universities in Eastern Europe and Eurasia.

Having rapidly grown into a well-renowned research university over two decades, HSE sets itself apart with its international presence and cooperation.

Our faculty, researchers, and students represent over 50 countries, and are dedicated to maintaining the highest academic standards. Our newly adopted structural reforms support both HSE's drive to internationalize and the groundbreaking research of our faculty, researchers, and students.

Now a dynamic university with four campuses, HSE is a leader in combining Russian educational traditions with the best international teaching and research practices. HSE offers outstanding educational programs from secondary school to doctoral studies, with top departments and research centers in a number of international fields.

Since 2013, HSE has been a member of the 5-100 Russian Academic Excellence Project, a highly selective government program aimed at boosting the international competitiveness of Russian universities.

ABOUT HSE GRADUATE SCHOOL OF BUSINESS

HSE Graduate School of Business was created on September 1, 2020. The School will become a priority partner for leading Russian companies in the development of their personnel and management technologies.

The world-leading model of a ‘university business school’ has been chosen for the Graduate School of Business. This foresees an integrated portfolio of programmes, ranging from Bachelor’s to EMBA programmes, communities of experts and a vast network of research centres and laboratories for advanced management studies. Furthermore, HSE University’s integrative approach will allow the Graduate School of Business to develop as an interdisciplinary institution. The advancement of the Graduate School of Business through synergies with other faculties and institutes will serve as a key source of its competitive advantage. Moreover, the evolution and development of the Business School’s faculty involves the active engagement of three professional tracks at our University: research, practice-oriented and methodological.

What sets the Graduate School of Business apart is its focus on educating and developing globally competitive and socially responsible business leaders for Russia’s emerging digital economy.

The School’s educational model will focus on a project approach and other dynamic methods for skills training, integration of online and other digital technologies, as well as systematic internationalization of educational processes.

At its start, the Graduate School of Business will offer 22 Bachelor programmes (three of which will be fully taught in English) and over 200 retraining and continuing professional development programmes, serving over 9,000 students. In future, the integrated portfolio of academic and professional programmes will continue to expand with a particular emphasis on graduate programmes, which is in line with the principles guiding top business schools around the world. In addition, the School’s top quality and all-encompassing Bachelor degrees will continue to make valuable contributions to the achievement of the Business School’s goals and the development t of its business model.

The School’s plans include the establishment of a National Resource Center, which will offer case studies based on the experience of Russian companies. In addition, the Business School will assist in the provision of up-to-date management training at other Russian universities. Furthermore, the Graduate School of Business will become one of the leaders in promoting Russian education.

The Graduate School of Business’s unique ecosystem will be created through partnerships with leading global business schools, as well as in-depth cooperation with firms and companies during the entire life cycle of the school’s programmes. The success criteria for the Business School include professional recognition thanks to the stellar careers of its graduates, its international programmes and institutional accreditations, as well as its presence on global business school rankings.

DOI: 10.17323/2587-814X.2025.1.7.21

Recommendation system model based on technical events

Kirill I. Pashigorev
E-mail: kipashigorev@sberbank.ru

Andrei O. Reznikov
E-mail: aoreznikov@sberbank.ru

PJSC Sberbank, Moscow, Russia

Abstract

Recommendation systems are widely used in the commercial field. The algorithms and architectures of recommendation systems are similar in various fields of application and have proven their effectiveness. Recommendations are based on the user's profile, the manner of his behavior on various IT (Information Technology) resources, as well as on similar users. At the same time, the use of recommendation systems in specialized areas is not widespread. Technology divisions are a promising new area of application for recommendation systems, and IT experts themselves will be the users. The purpose of this article is to consider a combination of a recommendation system, machine learning (ML) and LLM (Large Language Model) and to design these tools in a single system. Data volumes are currently measured in petabytes (10^{15} bytes) and exabytes (10^{18} bytes). In order to process even technical information (metadata/technodata) from the surrounding IT landscape, from the IT systems used by experts, AI (Artificial Intelligence) agents are needed. This article provides a literature review regarding the use of recommendation systems in combination with LLM applications, and suggests an application architecture model that generates human-readable news from technical event logs. The system is designed for a group of users who work with big data (ML engineers, data analysts, and data researchers). It is a combination of recommendation system technologies, LLM, and machine learning models. The article also provides the first results of the research that was carried out.

Keywords: recommendation system, matrix factorization, prompt engineering, LLM, AI agents

Citation: Pashigorev K.I., Reznikov A.O. (2025) Recommendation system model based on technical events. *Business Informatics*, vol. 19, no. 1, pp. 7–21. DOI: 10.17323/2587-814X.2025.1.7.21

Introduction

Today we are witnessing the widespread creation and use of AI agents. Tools such as recommendation systems, expert systems and voice assistants have become commonplace. But to effectively solve practical problems, the isolated use of individual tools is no longer effective, so system architectures are becoming more complex, and organizations are designing integrated systems to solve complex problems. The cases of designing complex systems will be discussed later in section 1.

In terms of the combination of technologies, we can single out currently popular AI agents - digital assistants for performing a set of tasks: several AI agents will be able to interact with each other and autonomously perform complex tasks without human intervention [1, 2], and be his unconditional assistant. The implementation of AI agents in all spheres of life, including human processing of large amounts of disparate data using smart assistants, is an urgent task for the coming years.

Recommendation systems have become a universal tool used in various fields, including those unrelated to the Internet: healthcare, education, logistics, etc. The introduction of recommendation systems in Technology departments is promising. The emergence of powerful language models based on the transformer architecture (LLM) has opened up a new approach to solving the problem of processing large amounts of data. Using LLM, it became possible to extract relevant information from large amounts of technical data, which was previously a task requiring significant human resources; identify target audiences among employees of the Technology department who may be interested in specific events or information; work with various types of professional information, including metadata, technological events, news from disparate channels and chats, articles, professional meetings, etc., and combine them all in a single interface. By focusing on the professional interests and competencies of employees, the system so designed will be able to provide more accurate recommendations and focus on relevant technical events without distracting the employee with uninteresting events that are outside the scope of his professional activity.

The use of LLM for log analysis was discussed in [3] and [4]. The authors consider a model for reducing the number of anomalies by analyzing the logs. At the same time, in the approach considered by the authors, there is a limitation on the composition of fields – the so-called weak adaptivity property, and, as a result, the need to retrain LLM in case of changes in input data structures. That is, technical data were used for analysis and forecasting and have limitations on input data structures. With the development of LLM, new scenarios for working with technical data appear, namely, notifying the end user about relevant events in the IT landscape. The authors also explicitly point out another drawback of the models working with logs – this is insufficient interpretability. When log-based anomalies are detected, interpreted results are crucial so that the administrator and analysts can trust and act on the automated analysis. The system model proposed in this article will allow LLM to be integrated into the industrial development process, ensuring that information is not lost and will be delivered to the recipient in a timely manner, as well as unambiguously interpreted. This approach is a new way of applying LLM to existing technical data.

The proposed system model also assumes the formation of a user's portrait for the correct selection of recommendations. To do this, concepts such as a user's portrait or user profile are introduced into the terminology. The users of the system are data experts – ML engineers, data analysts, and data researchers who are located within a diverse IT landscape, hundreds and thousands of systems of which are regularly updated, and in which tens of thousands of events occur regularly every month that need to be monitored.

1. Recommendation systems in large dynamic IT landscapes for metadata filtering

Recommendation systems are used in various fields, which include both commercial activities – e–commerce, distance education and entertainment platforms – these systems contribute to the selection of goods, services and content that match individual preferences and behavioral characteristics of users. They are increasingly being used to solve technical problems in the fields of security and construction, as a tool generalization of large texts, etc.

The review article [5] states that the key element used in modern information technology is data analysis using artificial intelligence technologies. The author considers machine learning, neural networks and natural language processing as the main elements of artificial intelligence that are used for data analysis.

In the article [6], the author describes hybrid systems in which various approaches and algorithms are combined to achieve more accurate personalized recommendations within the framework of building recommendation systems. It also describes the advantage of this combination over collaborative filtering and content-based recommendation models. At the same time, the author claims that hybrid systems solve the problems of cold start and aggregation of information from various sources. In the article [7], the authors additionally classify hybrid systems into monolithic, mixed, and ensemble models, defining monolithic recommender systems as a set of parts of various types of recommendation algorithms, and mixed recommender systems as a combination of the results of all the recommender systems included in it. At the same time, the complexity of developing such systems is highlighted, since significant resources and efforts are required.

In the article [8], the authors also use generalization methods for the purposes of their research and apply them to the input stream of social network messages numbering in the millions. The authors consider the case that the length of a generalized topic is a controlled variable, and further express the opinion that automatic generalization is the task of creating a consistent abbreviated version of the document, which outlines its main provisions. In this case, depending on the chosen use case, the target length of the final result can be selected relative to the length of the input document or can be limited.

The authors of the article [9] consider the use of LLM applications in construction to compile automated reports based on technical information, and also cite resource optimization as one of the results obtained. The authors use the term "intellectualization of construction inspection" in their work, and also point out the shortcomings in research that currently construction inspection mainly relies on manual execution and analysis.

The authors of the article [10] consider the use of a combination of machine learning, LLM and generative AI technologies. Similar to previous works, the authors pursue the goal of optimizing the working hours of experts in a specific field, and achieve it using modern technologies. The authors conduct research using the GPT-4 model (Generative Pre-trained Transformer 4) and mention the risk of hallucinations as LLM responses, as well as identify non-determinism (different responses during different sessions) as an additional level of complexity and unpredictability in the process of generating responses.

The authors of the article [11] also consider a combination of LLM applications and RecSys (Recommender Systems) to solve their problem and especially focus on the problem of cold start, considering various options. The authors call their system A-LLM-Rec (All-round LLM-based Recommender system), because the main idea is to allow LLM to directly use the collective knowledge contained in a pre-trained modern recommendation system based on collaborative filtering (CF-RecSys, Collaborative Filtering), so that new LLM features can be shared, as well as high-quality representations of users and products that are already trained in the modern CF-RecSys system. But the experiment is still being conducted not on technical data, but with human-readable headings and descriptions.

The authors of the article [12] position their approach as innovative. They describe in detail the application of an LLM-based multi-agent architecture, which uses the following chain of agents: Perceiver – Learner – Performer – Critic – Thinker. The architecture uses a Learn – Act – Criticize cycle and a reflection mechanism to increase the effectiveness of user interaction. As in the previously reviewed articles, the authors of this article focus on cold start. At the same time, the focus is on the balance between the accuracy of recommendations and user satisfaction. The experiment was also conducted on human-readable data. In general, it is the LLM component (consisting of small agents/modules) that is presented as innovation.

In the sources considered, much attention is paid to large language models, the architecture of which can be found in detail in [13]. It should also be noted that the creation of intellectual assistants in various subject areas is actively considered not only by foreign authors and researchers, examples of which are given in sufficient detail in this section, but also by domestic authors and researchers [14, 15]. The difference between the system being designed in this study is that the system we designed does not have business goals such as audience retention or increased content consumption, but rather aims to notify the user of changes in the company's information infrastructure that are relevant to their specifics and objectives. For example, the release of a new data product related to the user's projects, changes in the data relevant to him, the release of releases of systems relevant to the user, changes in metadata, etc. Let's highlight two problems that were not fully solved in [4] – weak adaptability of models and insufficient interpretability of results, and further in this article we will consider ways to solve these problems, among others.

2. Architecture of the Techno Events Smart News feed system

The primary information about an event is technical data generated by other systems (event logs), which are difficult for humans to perceive. Our task is not only to find an event that is relevant to the user, but also to bring it to a form that is easily perceived by humans. Due to the wide range of events being processed, a universal solution for bringing event information to a user-friendly form is to implement LLM to summarize technical data in a news notification containing all the key aspects of the event. In addition, to better match users to a specific event, we will use tagging to highlight key properties.

To define the context of the study, we will introduce the term "technodata": These can be thousands of events taking place in hundreds of IT landscape systems. *Table 1* shows a simple example of the structure of such events.

To solve the research problem, an intelligent system for processing a large volume of events was designed, the architecture of which is shown in *Fig. 1.*

Let's consider the main elements of this system.

- Events generated in IT landscape systems are stored in the DBMS.

Table 1.

Example of an event table

id	Date	Event's type	Description	…	Author	Source
uuid	Date	String	Text	…	Varchar	Varchar

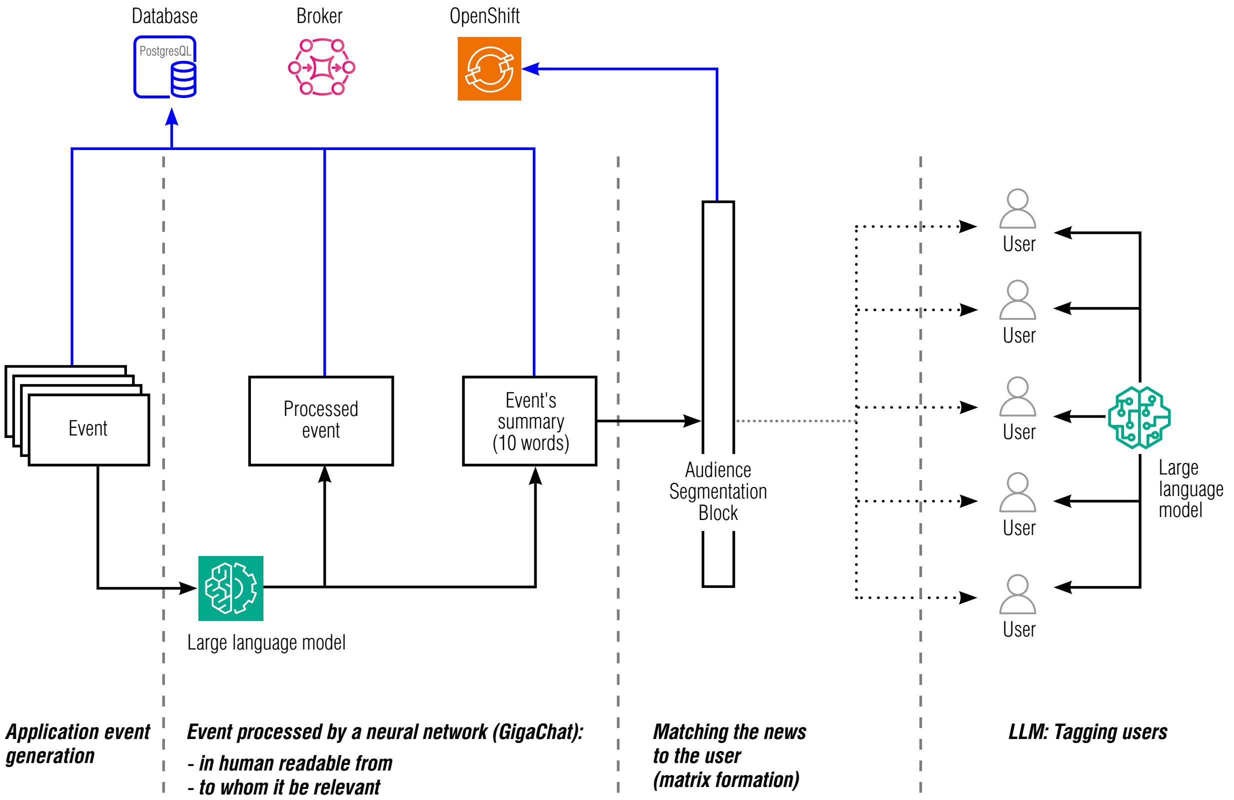


Fig. 1. The main elements of an intelligent system for processing a large volume of events.

- The Giga-intgr microservice sends technical event data to LLM. The response from LLM (processed events) in the form of generated news and assigned tags is recorded in the DBMS.
- The Jira-intgr microservice requests a list of user tasks in the Task Accounting system (Jira) and transmits it to Giga-intgr using a message broker. Giga-intgr transmits the received message to LLM. The response from LLM (user tagging) in the form of tags for each user is recorded in the DBMS.
- The SegmentOfAuditory-serv service generates a matrix of user-news correspondence using a model.
- The homepage service receives news for display in the feed from the DBMS via the API.

The overall result of using LLM in the presented architecture model is the ability to solve the following key tasks:

- Tagging users, i.e., assigning them appropriate tags reflecting their interests and preferences.
- News tagging, which allows you to classify news materials into various categories and topics.
- Forming a short news content based on technical information so that the user can quickly familiarize himself with the content of the message without having to read the entire text.
- The recommendation system has to solve the problem of selecting interesting news for each specific user based on his personal interests and preferences.
- The system architecture presented is able to take

into account a large number of the following parameters:

- A large number of information systems as sources, which can be measured in the range from one to several thousand.
- A large number of events with different structures, which can also vary depending on the event or the source system, as well as over time; the number of events can be measured in hundreds of thousands, and the number of attributes in the structure of these events can vary from two to several hundred.
- A large number of user roles and an even larger number of users.

The architecture of the system is shown in *Fig. 2*, and a description of the technical components is given in *Table 2*.

The architecture is designed in a micro-service style and assumes placement in a dynamic infrastructure on an OpenShift cluster in order to increase the level of horizontal scaling.

3. Optimizing work with technical events

3.1. Setting the task of recommending news

In order to preserve the coherence of the presentation and to formalize the problem, we present a

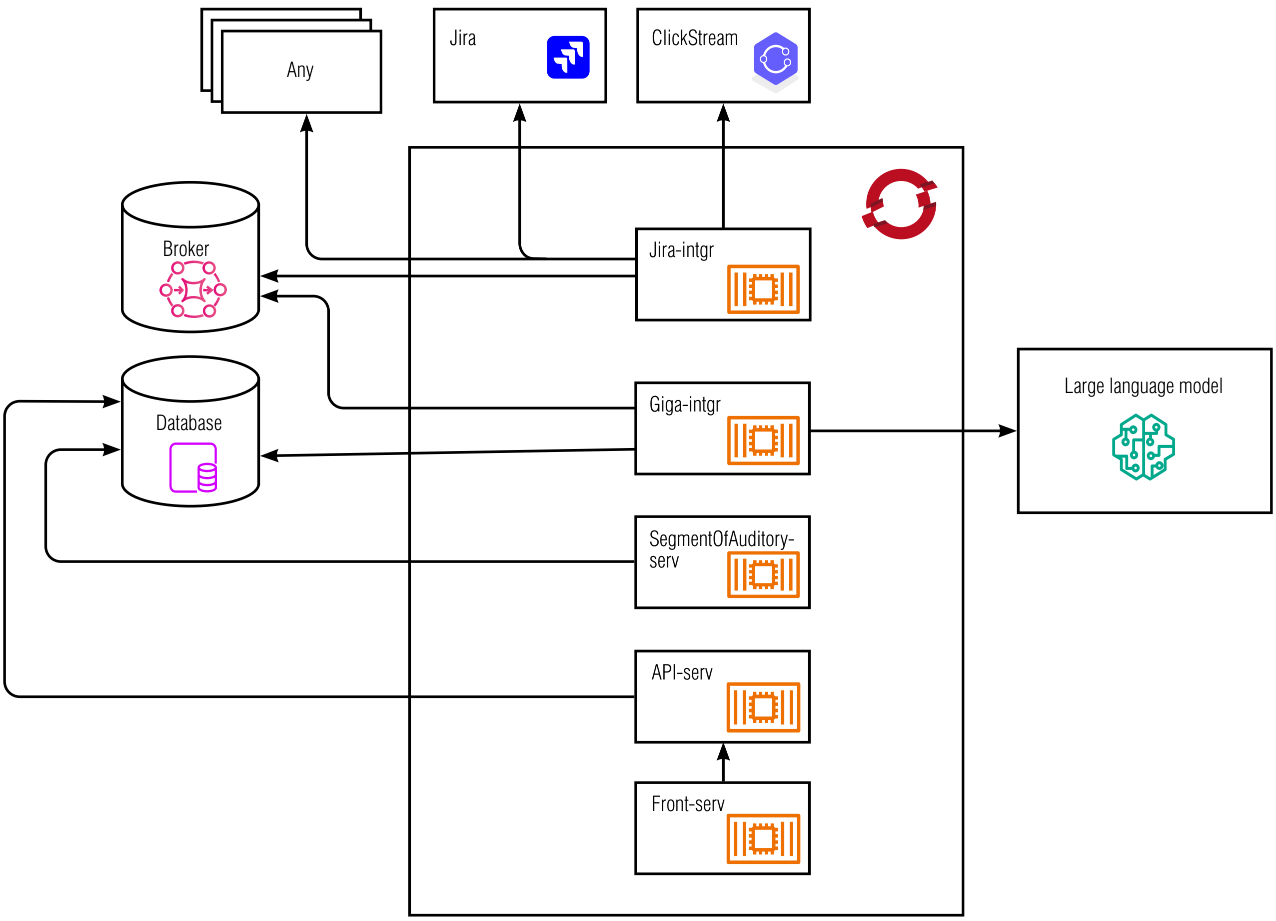


Fig. 2. Component diagram.

Table 2.

System components

Component	Description
Large language model	The Giga Chat API technology will be used as a means of accessing the large language model [16].
Any, Jira, ClickStream	These components are external data source systems. Any arbitrary system can act as a source.
DBMS	A stateful component for storing data in a structured form.
Kafka Broker	A stateful component for implementing interaction between microservices.
Giga-intgr	A component (microservice) for implementing interaction with a large language model. Implemented in Python.
Jira-intgr	A component (microservice) for implementing data acquisition from the surrounding landscape. Implemented on NodeJS.
SegmentOfAuditory-serv	This component (microservice) is used to select, transform, combine, and other ways to prepare data to find useful patterns, as well as to extract patterns from the data obtained.
API-serv	The API for the main page of the feed is implemented in NodeJS.
Front-serv	The main page of the feed (GUI) is implemented in JavaScript – React.

description of the matrix factorization method, which was repeatedly and in detail discussed in [17–19], and is a decomposition of the source matrix into the product of two other matrices of lower rank. In the study described in this paper, the final matrix R is decomposed into matrices A and B: a set of employees is considered $U = \{u_1, u_2, ..., u_m\}$, multiple tasks $T = \{t_1, t_2, ..., t_n\}$ and multiple news $N = \{n_1, n_2, ..., n_k\}$. Employee task matrix $A = (a_{ij}) \in \{0;1\}^{m \cdot n}$ is defined in such a way that $a_{ij} = 1$, if the u_i employee is engaged in the task t_j, and $a_{ij} = 0$ otherwise. News matrix $B = (b_{ij}) \in \mathbb{R}^{n \cdot k}$ represents the relevance of the n_i news to the t_j task. The purpose of the study is to find the recommendation matrix $R = (r_{ij}) \in \mathbb{R}^{m \cdot k}$, where r_{ij} – is the degree of relevance of n_j news to the u_i employee.

Assessment of the relevance of news for tasks is performed by calculating the degree of relevance of b_{ij} for each task t_j and news n_i using BERT text analysis methods. The relevance of news for a u_i employee is assessed by calculating the degree of relevance of r_{ij} for each u_i employee and n_j news using the employee's task matrix A and the news matrix B. The ranking of news for each u_i employee is based on the degree of relevance of r_{ij} in descending order.

This approach can be represented using the following formula:

$$r_{ij} = \sum_{l=1}^{n} a_{il} \cdot b_{jl},$$

where a_{il} is the degree of participation of the u_i employee in the t_l task, and b_{jl} is the degree of relevance of the n_j news to the t_l task.

Restrictions are introduced on the research task: a limit on the maximum number of recommendations L for each employee and the establishment of a minimum degree of relevance r_{min}, below which the news is not recommended to the user.

The following criteria are used to evaluate the model used: accuracy (Precision), completeness (Recall) and F1-score. The goal is to maximize these metrics to provide the most relevant recommendations for each employee. The combination of evaluation criteria is presented in *Table 3*.

This assessment will determine how boldly the model should select recipients. The initial assessment is first performed by a human.

It is important to note that when training the model in the course of the study, an assumption will be made about the confidence interval: if there are no errors of the first kind and there is an acceptable number of errors of the second kind, then the same picture will be preserved for the entire set (sample). Errors should be aggregated by the reviewer and a decision is made whether further training of the model and correction of the error are required. Errors of the second kind are corrected by promptness, errors of the first kind require additional training.

It is also necessary to take into account the need to solve the "cold start" problem [20], which occurs once when the system is started.

At the presented stage of the study, the assumption is made that technical news is offered to users randomly. To evaluate the quality of the predictions, the set of estimates V was divided into separate sets V_{train} for training and V_{test} for testing. The tests for the model will be performed on quantitative indicators, which are presented in *Table 4*.

The characteristics presented in the *Table 4* are relevant to the specifics of this work:

Table 3.

The criteria matrix

The news should be recommended and indeed got into the recommendation (TP)	The news should be recommended, but it doesn't actually make it into the recommendation (FP)
The news should not be recommended, but it actually gets into the recommendation (FN)	The news should not be recommended, and indeed does not get into the recommendation (TN)

Table 4.

Quantitative characteristics of the experiment

Characteristic	Value	Characteristic	Value
Number of "DP Creation" events	40000	Training sample	75%
Number of "MD Change" events	10000	Test sample	25%
Number of DM application releases	10–1500, including child tasks	The rank of the matrices	the value of *X* will be accepted
Number of DSM application releases	10–1500, including child tasks	Factorization steps	the value of *Y* will be accepted
Number of users	50	Range of values of the original matrix	{0, 1} – implicit feedback

- DP (data product) is a type of information asset that is a structured description of a set of industrial data (type and composition of data, data sources and delivery methods) available for ordering and receiving through industrial data distribution paths for use in the interests of the Bank's products and the Ecosystem.
- CAP (Corporate Analytical platform) is the general name of the solution responsible for obtaining data, processing it, and providing the processed data to interested parties. A centralized platform for data collection (loading), delivery (unloading), processing, integration, research and analysis. The CAP consists of a Core and a User Space.
- MD (Data Manager) is an employee of the Bank, assigned to the role in accordance with the Data Management Policy (depending on the destination).
- DM App (Data Map) is a system responsible for managing data products: their attributes, structure and accessibility to distribution in the Data Supermarket App.
- DSM App (Data Supermarket) is a single portal for interaction with the CAP which allows you to study and order data from the CAP. It is intended for distributing data in the CAP from replicas of industrial sources.

At the initial stage of the experiment, more than 95% of the cells of the matrix V will be unfilled, i.e. the matrix will be very sparse.

The standard deviation (*RMSE*) will be used as a metric [21]. *RMSE* will allow you to estimate the proportion of news that should have been shown to a specific group of users and were shown:

$$RMSE = \sqrt{\frac{1}{N-d}\sum_{i=1}^{N}\left(M_i - O_i\right)^2},$$

where

M_i is the predicted value for the i-th observation in the dataset;

O_i is the observed value for the i-th observation in the dataset;

N is the sample size;

d is the degree of freedom in the model configuration. For a linear fit $d = 2$.

3.2. LLM and prompt engineering

For the purposes of our research, we will use the already trained GigaChat language model [22], and focus on prompt engineering to achieve the desired result without further training the model. The meaning of these terms is presented in [23]: prompt engineering is the process of creating efficient and accurate tools for working with large language models (LLM). Prompt is a text description of a task that needs to be completed using an AI model. A query that is set by the model to generate text, images, code, or other types of content.

The following elements will be used to transfer it to a large language model:

- Summary type prompt – highlighting the main theses from the text, keywords for these types of prompt: reduce, summarize;
- Prompt with the type "Generation", keywords: write, compose, invent;
- Role – for acting as a certain character;
- The Zero-shot method does not contain any response examples, the response of a large language model will be received in free form.

A preliminary list of event log attributes to be transmitted to a large language model based on which the news will be generated:

- FB tribe;
- Tribe;
- DP name;
- Data product;
- Category;

- Data product type;
- Cluster;
- Schema Platform;
- Database schema;
- Number of tags;
- Tag name;
- Tub. MD DP number;
- UUID;
- Latest status;
- MD assignment date;
- Publication flag;
- Name of the receiving system;
- Distribution platform;
- SLA and others.

A tribe is a group of cross-functional teams working on a product or service area. Each team includes specialists of all the necessary profiles to create a turnkey product.

Thus, based on the attributes of the "Date Product Change" event:

- the news should appear: *"The W showcase for the online assistant has been launched in Block Bl"*;
- the following tags should be assigned to the news: *{"Focus Group": "Block Bl"}, {"Topic": "Showcase W for online Assistant"}, {"Category": "Block Bl News"}.*

When preparing prompt, it is important to make sure that there is no hallucination effect [16, 24]. Also, a mandatory point in the preparation of the prompt is the task of mandatory passage of the Censor's check. There may be cases when the Censor regards a correct request to a large language model as undesirable for discussion and returns a response about the inappropriateness of this topic with a request to change it to a correct one.

Based on the results of the primary study, the following results were obtained, presented in *Table 5.*

Table 5 shows a comparison of the distances between objects and the results of generation based on them.

- Prompt 1: *"Generate a clear and concise news article summarizing the key details of the event described in the technical data provided. Focus on presenting factual information in a human-readable format, avoiding any speculation, assumptions or excessive embellishments. Make sure that the article has a clear and logical structure, uses the correct grammar and syntax of sentences. Use a neutral tone and avoid sensationalistic language.*

Table 5.

Results of the primary research

The distance between the prompts	Prompt 2	Prompt 3
Prompt 1	0.03732394628917002	0.095708435241878
Prompt 2	0	0.0855076122459778
The average distance between the news generated using these prompts	AI-news of prompt 2	AI-news of prompt 3
AI-news of prompt 1	0.15939979460493073	0.1789409953210833
AI-news of prompt 2	0	0.13067957637939245

The goal is to provide an informative and objective overview of the event, making it easy for readers to understand.”

- Prompt 2: *“Generate a short and informative news article summarizing the key details of the event described in the provided technical data. Focus on presenting factual information in a neutral tone, avoiding any speculation, speculation or emotional language. Use a clear and logical structure and avoid sensational or attention-grabbing language. The goal is to provide an objective summary of the event so that readers can quickly understand what happened.”*
- Prompt 3: *“As an expert in making informative notifications, make up a short message no more than 10–12 words long that will provide key information about the event.”*

From the data in *Table 5*, it can be seen that even products that are similar in structure and content can produce results with significant deviations from each other. This behavior can be compared with the behavior of a rigid system of ordinary differential equations (ODES). Based on the fact that the project uses an already trained LLM without the global need for further training, as well as on the basis of the results obtained, it is concluded that a clear formulation of the plan is necessary to obtain a stable result.

Based on the study of the behavior of the model with various prompts, three prompts were selected, the data on which were given in *Table 5*. With the help of these data, a test sample of news was generated, which was provided for markup on the TagMe platform [25] to the target group of users of this product, in order to select the optimal informative and easy-to-understand version of the generated news. As a result, the second prompt was selected for use in the pilot version of the product.

3.3. Cold start and first results

Considering that the users of this service will be employees of our company, it is advisable to think about how to “warm up” the cold start for new users in order to reduce the funnel of convergence of the system to the recommendation of the most relevant news. Since our target audience exists in the same IT landscape, it is possible, in addition to collecting technical data, to also collect digital traces of user work. The first and most obvious trace is the tasks that were started by the user, and which the user started himself. Based on this, we can immediately hypothesize that the tasks associated with the user reflect the type of his professional activity, and therefore, comparing the news in proximity to the user’s tasks in Jira, we can assume how relevant they are to the user. Having received the user’s tasks and having constructed embeddings for the tasks that the user was engaged in, and embeddings for news and technical data, we can build a matrix of proximity (cosine distances) between tasks and news. This way we can get primary information about which news is most likely to overlap with the user’s professional activities.

After constructing the correlation matrix, the task arises of displaying it in a ranked list of news for recommendation to the user. The simplest approach would be to calculate the average correlation for each news item and sort it by the values obtained. However, in the course of the study, another method was chosen, the algorithm of which looks like this: for each news item, the top N $(N = 5)$ news items are selected for each task, each news item is assigned a weight equal to $(N - i)$, where i is the number of news in the top for each of the tasks; next, all the weights for each news item are summed up and sorted according to the values obtained. The peculiarity of this approach is that not all news gets into the final ranking, however, the distribution of news occurs in a more “honest” way due to the fact that some tasks are poorly correlated with both relevant news for the user and other user tasks. This distribution allows, firstly, to reduce the impact of outliers and make news ranking more resistant to abnormal user activity, while maintaining relevant news for abnormal activity, and, secondly, to exclude irrelevant news from the ranked sample, which in the future will allow more complex models to be used more confidently on such a filtered sample news. But at the moment of the cold start, we can simply output K (the number of news recommended to a particular user) of the first news from this list.

Thus, when the user logs in for the first time, the system already has primary data on the specifics of the user's work, based on his tasks obtained from the task accounting system. The system calculates embedding based on the user's tasks which will later be used to find the cosine distance with news embeddings. Next, the user is invited to read the recommendations that the proposed model has calculated for him in the news feed.

The advantage of using the proposed system model is the exclusion of the use of classical conditional algorithms to determine the target groups of users. In the proposed system model, this function will be performed by the recommendation system using the cosine similarity algorithm and the overlap matrix. The way a large number of event types are handled by classical methods using conditional operators does not produce the desired result. Meanwhile, using a language model to categorize events and users allows you to process these types of events quickly and optimally. It is also worth considering that the same type of event, but with different input characteristics, such as, for example, a logging object, may be relevant for a specific user in one case, but in another case it will no longer be relevant with a different logging object. Processing all possible combinations of input parameters using conditional operators is an expensive operation from a technical point of view.

Conclusion

This article presents the concept of an event management system architecture and a model for routing events in the news format to specific users, as well as defines criteria for evaluating the application of the model. This approach uses LLM to convert raw technical data into short news, which is then delivered to users through a recommendation system. Such an intelligent system design combines neural network technologies, recommendation systems and machine learning to minimize the effect of spam and notify users in a timely manner. The result of automating the news generation process will be a reduction in the time spent by an expert searching for information, and as a result, the risk of critical incidents is minimized. The proposed architecture of the software system implements the interaction of unrelated components, combining them into a single AI agent, minimizing the flow of news to a single user. The proposed architecture also allows for further development of the system at the lowest cost by integrating speech recognition components into the system, which will make the system a full-fledged AI assistant. The combination of these technologies is an indispensable assistant to support experts in their daily data-related tasks.

Based on the results of the first application of the proposed model to the technical data that had been accumulated historically, the first result was obtained: usually the problem of incident detection was solved on such data. But in the new realities, when the rate of change in the IT landscape is regularly increasing, we have to look for new ways to apply LLM to technical data.

Using the example of the system model presented in this article, more than 0.5% of useful data was extracted from the total volume of messages, and more than 1% of relevant consumers of this data were found. This result also reduces the time spent by employees on tracking changes in the IT landscape. Unlike classical approaches to solving the problem of detecting anomalies based on logs presented in [3] and [4], the approach proposed in this article focuses not on log analysis, but on anticipating potential incidents due to sensitive changes in the IT landscape, in products consumed by user systems, due to notifying users about this in a timely manner. The correctness of our chosen direction is also confirmed by the hypothesis from [4] that a high percentage of false positives can lead to missing important failures in the system, and a high percentage of false passes can lead to a waste of developers' efforts. Our proposed approach allows us to prevent failures and costs for developers' efforts.

We also obtained a metric of the final relevant amount of information in the amount of 0.05–0.15% for the test sample of employees and a reduction in the volume of targeted information while maintaining informativeness to no more than 10% of the initial volume. The average volume of incoming material decreased from 8629 characters to 189–538 characters, which reduced the amount of information consumed by more than 23 times while maintaining 96% of the semantic load. ■

Acknowledgments

The authors would like to thank the Sber Department of Research and Innovation for hosting the Techno-Idea pitch session and the opportunity to conduct research, the Sber Department of Data Dissemination and the Department of Artificial Intelligence and Machine Learning Technology Development for providing resources for research, as well as Associate Professor of the HSE Faculty of Computer Science, Candidate of Pedagogical Sciences S.A. Videnin for methodological recommendations when writing the article.

References

1. Microsoft (2024) *What are AI agents?* Available at: https://learn.microsoft.com/ru-ru/azure/cloud-adoption-framework/innovate/best-practices/conversational-ai (accessed 20 July 2024).
2. Sber (2024) *A platform for launching autonomous AI agents is presented*. Available at: https://ai.sber.ru/en/post/predstavlena-platforma-dlya-zapuska-avtonomnyh-ai-agentov (accessed 20 July 2024).
3. Shah A., Pasha D., Zadeh E., Konur S. (2022) Automated log analysis and anomaly detection using machine learning. *Frontiers in Artificial Intelligence and Applications*, vol. 358: Fuzzy Systems and Data Mining, pp. 137–147. https://doi.org/10.3233/FAIA220378
4. Chen Z., Liu J., Gu W., et al. (2021) Experience report: Deep learning-based system log analysis for anomaly detection. *arXiv:2107.05908*. https://doi.org/10.48550/arXiv.2107.05908
5. Mokshanov M.V. (2024) The use of artificial intelligence in data analysis: an overview of the current state and future directions. *Universum: technical sciences: electronic scientific journal*, no. 5(122) (in Russian). https://doi.org/10.32743/UniTech.2024.122.5.17513
6. Eremin O.Y. (2023) Methods of implementation of hybrid recommendation systems. *E-Scio*, no. 3(78) (in Russian).
7. Kurennykh A.E., Sudakov V.A. (2022) Approach to the development of hybrid recommendation systems. *Bulletin of Science and Practice*, vol. 8, no. 11 (in Russian).
8. Völske M., Potthast M., Syed S., Stein B. (2017) TL;DR: Mining Reddit to Learn Automatic Summarization. Proceedings of the *Workshop on New Frontiers in Summarization, Copenhagen, Denmark, 2017*, pp. 59–63. Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-4508
9. Pu H., Yang X., Li J., Guo R. (2024) AutoRepo: A general framework for multimodal LLM-based automated construction reporting. *Expert Systems with Applications*, vol. 255, part B, article 124601. https://doi.org/10.1016/j.eswa.2024.124601
10. Sivakumar M., Belle A.B., Shan J., Shahandashti K.K. (2024) Prompting GPT–4 to support automatic safety case generation. *Expert Systems with Applications*, vol. 255, part C, article 124653. https://doi.org/10.1016/j.eswa.2024.124653

11. Kim S., Kang H., Choi S., et al. (2024) Large Language Models meet Collaborative Filtering: An efficient all-round LLM-based recommender system. *arXiv:2404.11343*. https://doi.org/10.48550/arXiv.2404.11343

12. Shu Y., Zhang H., Gu H., et al. (2023) RAH! RecSys-Assistant-Human: A human-centered recommendation framework with LLM agents. *arXiv:2308.09904*. https://doi.org/10.48550/arXiv.2308.09904

13. Vaswani A., Shazeer N., Parmar N., et al. (2017) Attention is all you need. *arXiv:1706.03762*. https://doi.org/10.48550/arXiv.1706.03762

14. Morozevich E.S., Korotkov V.S., Kuznetsova E.A. (2022) Development of a model for the formation of individual educational trajectories using machine learning methods. *Business Informatics*, vol. 16, no. 2, pp. 21–35. https://doi.org/10.17323/2587-814X.2022.2.21.35

15. Palchunov D.E., Yakobson A.A. (2024) Development of an intelligent assistant for the selection of goods in the process of dialogue with the user. *Business Informatics*, vol. 18, no. 1, pp. 7–21. https://doi.org/10.17323/2587-814X.2024.1.7.21

16. Amenitsky A.V., Rukhovich I.V., Amenitskaya L.A., et al. (2024) Side effects of hallucinations of artificial intelligence. *Science, innovation, education: current issues and modern aspects*, pp. 224–235. Penza, 2024 (in Russian).

17. Strömqvist Z. (2018) *Matrix factorization in recommender systems: How sensitive are matrix factorization models to sparsity?* Uppsala University Publications. Available at: https://uu.diva-portal.org/smash/get/diva2:1214390/FULLTEXT01.pdf (accessed 22 July 2024).

18. Moisyuk-Dranko P.A., Revotyuk M.P. (2020) Methods of matrix factorization for recommendation systems. Proceedings of the international scientific conference *Information technologies and systems 2020 (ITS 2020)*, pp. 193–194. Minsk: BGUIR (in Russian). Available at: https://libeldoc.bsuir.by/bitstream/123456789/41339/1/Moysyuk_Dranko_Metody.pdf (accessed 22 July 2024).

19. Kuznetsov I.A. (2019) *Methods and algorithms of machine learning for preprocessing and classification of weakly structured text data in scientific recommendation systems*. Moscow: MEPhI (in Russian). Available at: https://ds.mephi.ru/documents/90/Кузнецов_И_А_Текст_диссертации.pdf (accessed 22 July 2024).

20 Yuan M., Lin H.-T., Boyd-Graber J. (2020) Cold-start active learning through self-supervised language modeling. *arXiv:2010.09535*. https://doi.org/10.48550/arXiv.2010.09535

21. Liemohn M.W., Shane A.D., Azari A.R., et al. (2021) RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics. *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 218, article 105624. https://doi.org/10.1016/j.jastp.2021.105624

22. Sber (2024) *GigaChat API* (in Russian). Available at: https://developers.sber.ru/portal/products/gigachat-api (accessed 22 July 2024).

23. Sber (2024) *Prompt engineering* (in Russian). Available at: https://developers.sber.ru/docs/ru/gigachat/prompt-engineering (accessed 26 December 2024).

24. Amenitsky A.V., Rukhovich I.V., Amenitskaya L.A., Amenitsky D.A. (2024) Causes, ethical problems and prevention of hallucination LLM. *Intelligence. Collection of articles of the International Competition of Young Scientists*. Penza, pp. 12–15(in Russian).

25. Sber (2024) *TagMe Data Markup Platform* (in Russian). Available at: https://developers.sber.ru/portal/products/tagme (accessed 09 December 2024).

About the authors

Kirill I. Pashigorev

Head of the Department, SberData, PJSC Sberbank, 25A bld. 6, Warsaw Highway, Moscow 117105, Russia;

E-mail: kipashigorev@sberbank.ru

ORCID: 0009-0008-3478-4874

Andrei O. Reznikov

Chief Development Engineer, SberData, PJSC Sberbank, 25A bld. 6, Warsaw Highway, Moscow 117105, Russia;

E-mail: aoreznikov@sberbank.ru

DOI: 10.17323/2587-814X.2025.1.22.33

Determining the sequence of project implementation for the program of improving the efficiency of business processes

Arkadiy I. Maron [a]
E-mail: amaron@hse.ru

Grigoriy D. Fainburg [b*]
E-mail: fain2000@mail.ru

[a] HSE University, Moscow, Russia
[b] Moscow State Technical University of Civil Aviation, Moscow, Russia

Abstract

This article is devoted to the problem of determining the sequence of project implementation in a program of improving business processes of an organization. The relevance of the study is related to current conditions, where the quality of business processes is essential not just for the success, but also for the survival of an organization. Improvement of business processes is a costly program that involves certain projects. The projects of the program cannot be started at the same time due to limited budget and human

* Corresponding author

resources. Thus, we face the task of determining the sequence of stages of program implementations. The solution of this task is one of the most important problems of business informatics. This paper proposes a new criterion for prioritizing projects. It takes into account the fact that funds for projects are generated during the implementation of business processes. The criterion also takes into account the pace of spending the project budget and the need for participation of key employees of the organization. The implementation of the program is divided into a few stages. At each stage, the problem is solved by determining a set of projects whose sum of priorities is maximum and whose resource requirements do not exceed the constraints developed at that stage. The relevance of the article is initiated by looking at the need of enterprises that ensure the airworthiness of civil aviation airplanes. This work is of interest for project program managers of production and service companies, as well as for a wide range of researchers.

Keywords: business process, project, program, priority, criterion, sequence of projects implementation

Citation: Maron A.I., Fainburg G.D. (2025) Determining the sequence of project implementation for the program of improving the efficiency of business processes. *Business Informatics*, vol. 19, no. 1, pp. 22–33.
DOI: 10.17323/2587-814X.2025.1.22.33

Introduction

Business processes (BPs) are essential for Enterprise Architecture because organizational and IT structures should be built upon them [1]. BPs need continuous improvement. Moreover, in modern conditions, the period of invariability of BPs is decreasing.

Two approaches to BPs improvement can be distinguished: efficiency improvement (optimization) and reengineering. In the first case, individual BP elements are improved [2]. Based on qualitative and mathematical analysis, "as is" (AS-IS) and "to be" (TO-BE) models are created [3–5]. Reengineering involves creating a TO-BE model without regard to the current state [6–8]. BP modeling notations are an evolution of the network approach to modeling workflows that has been widely used in project management. They allow one to explicitly indicate the possibility of different BP implementation scenarios. This is not the case in the network project model. All work specified in the project plan must be completed. This is the basis of the critical path method [9]. The project is unique by definition, and the manager must plan it without scenarios. This is an obvious contradiction that causes difficulties. The ability to create and store multiple plans does not solve the problem [10]. The GERT method [11, 12] also did not lead to a practical result. The problem of scenarios in project management remains open.

In this article, the BP optimization program is considered. The development of a methodology for the formation of such a program and its management is a pressing issue for enterprises in various sectors of the Russian national economy. In connection with the current situation, this issue is particularly relevant for enterprises ensuring the maintenance of airworthiness of civil aviation aircraft [13]. For each BP, the implementation of the transition from AS-IS to TO-BE is a separate project. Together, these projects form a program aimed at achieving the goal of optimization of the company's BPs. The following two main questions arise:

- Which projects should be included in the program?
- In what order should the projects be executed?

These issues are related to financial and resource constraints. Even if we decide to optimize only the main BPs that generate revenue, there may not be enough resources to launch all the initiated projects at the same time. It is necessary to prioritize the projects. In this case, it is necessary to take into account the importance of the BPs for the company. It is also necessary to consider that the financial resources for the implementation of projects come from the income generated by the BPs. Let us clarify whether this problem has been solved so far.

Project management is the most regulated area of management [14]. Qualification requirements for project managers are regulated [15]. Project management activities are also regulated [16–19]. These and earlier project management standards indicate that projects should be aligned with the strategic goals of the organization. In accordance with this, the work [20] proposes to determine the priorities of projects by the degree of their relevance to the strategic goals of the organization. For this purpose, first of all, it is necessary to determine the priorities of goals. This approach is based on Saaty's Analytic Hierarchy Process (AHP) method [21]. The source data for determining the priorities of goals is the matrix of pairwise comparisons. The comparison is made on a qualitative linguistic scale, where each linguistic value corresponds to a specific numerical value. In this way, the matrix of pairwise comparisons can be translated into digital form. The priorities of the goals are calculated as an eigenvector corresponding to the principal eigenvalue of the resulting matrix. The AHP method does not take into account the links between the goals. At the same time, these connections are clearly displayed on the strategy map. The strategy map is an element of the Balanced Scorecard concept [22, 23]. Its use is recommended to ensure consistency when filling in the matrix of pairwise comparisons.

The approach proposed by Vargas [20] is implemented in MS Project Server [24]. The method for calculating the priority vector is not specified by the developers. However, its simple approximate calculation proposed in [21] can lead to errors in calculating priorities affecting the ordering of goals. It is necessary to adhere to a highly accurate procedure for determining the main eigenvalue of the matrix and the corresponding eigenvector [25].

Later, a number of modifications of the method proposed by Vargas [20] were proposed. Thus, in the work [26] it was proposed to take into account the risk level of projects when comparing them. Risk is an integral element of the project. This follows from the uniqueness of the project. Works devoted to project risks can be divided into two groups. The first proposes methods that allow you to determine, or better yet, predict the moment when deviations in terms and/or finances become irreversible [27–29]. The second group of works is devoted to organizing timely control of the correctness of the work [30–32]. Both areas are called project diagnostics [33]. In technical diagnostics, a direction similar to the first of them is related to failure prediction [34]. A direction similar to the second is the choice of control points [35].

The method proposed by Vargas [20] does not explicitly use numerical characteristics of either goals or projects. The work [36] suggests assessing the priority of a project based on the sum of weighted points. When calculating the points, "project sustainability assessment indicators" are used. They are divided into groups: "Economic indicators," "Ecological indicators," "Social indicators." The economic indicators include, in particular, the NPV (net present value) and IRR (internal rate of return) of the project. The group of ecological indicators includes, among others, specific energy consumption in kind and the use of energy from renewable sources. The social indicators include: creation of new jobs, level of expenditure on labor protection. Then a portfolio (set) of projects is sought, in which the sum of weighted points is maximum, and the restrictions on allocated resources and the number of projects are met. In this case, the simplex method is used.

The approaches outlined in [20] and [36] can be considered the main ones when solving the problem of forming a portfolio or program of projects based on their priorities. At the same time, [20] deals only with determining project priorities. [36] not only proposes a method for calculating project priorities, but also solves the problem of initial portfolio formation. Initial in the sense that it determines which of the initiated projects should be included in the project and which should not due to existing restrictions.

The problem that our article is devoted to postulates that all projects for improving the selected BPs must be implemented, albeit in several stages. Both at the initial (first stage) and at subsequent stages, the funds (budget) for implementing the projects are formed from the income generated by the BPs. This fact must be taken into account explicitly when determining project priorities. There is no such thing in the literature today.

Thus, the following points are relevant:

- development of a method for calculating the priorities of projects aimed at improving the organization's business process, taking into account that the project budget is formed from the business process's income;
- determining the sequence of implementation of projects of the BP optimization program taking into account the existing financial and resource constraints, provided that all projects must be completed.

As a starting point, let us consider the case where there are no interrelations between the projects that are carried out to optimize the BP. Such a situation is quite typical for enterprises that ensure the maintenance of airworthiness of civil aviation aircraft [13].

1. Statement of the problem

The company has n BPs that directly generate income. Currently, the income generation intensity of BP with number i ($i = 1, 2, \ldots, n$) is v_i monetary units per unit of time. For the sake of concreteness, we will assume that we are talking about thousands of rubles per day. An initial budget of B^1. This budget is formed from the income that these BPs previously brought in. It was decided to improve them (optimize them). For this purpose, BP modeling was conducted. As a result, it was determined what needs to be done to transition each BP from the AS-IS state to the TO-BE state. For each i-th BP, it was calculated how much the intensity of income generation will increase as a result of the transition from AS-IS to TO-BE. Let us denote this value as d_i. We will also assume that we are talking about thousands of rubles per day.

For each BP with number i, a project P_i initiated, the implementation of which will ensure the required transition to a new improved state. Accordingly, the value d_i can be considered as an indicator of the effectiveness of project P_i. A total of n projects is initiated. It has been established that there are no dependencies between the projects. The cost (budget) of project P_i is equal to C_i. The planned time of its implementation is equal to T_i. The company employs m key specialists. Each project requires the participation of at least one of them. Key specialists are engaged in BP. Key specialist j ($j = 1, 2, \ldots, m$) can allocate only a certain share s_j of his daily working time to work on projects. In what follows, we will call s_j the project resource of key specialist j or simply a resource. The resource can be defined both in shares and in percentages of the total working time of the specialist. It is known that project i requires a resource of s_{ij} from key specialist j. In this case, $s_{ij} \leqslant s_j$.

It is decided that all projects should be implemented. If at the first stage the initial budget and/or resources are not enough to simultaneously launch all n projects, then the remaining projects will be launched as income from the above BPs accumulates and resources are freed up key specialists.

It is necessary to determine the order in which projects will be launched.

2. Solution

2.1. Priorities of projects to improve BP

When developing performance management systems, it is necessary to create special indicators [37]. As indicated in the work [37], this is due to the fact that the indicators of economic benefits used in investment analysis are indirect and implicit. We will synthesize an expression for calculating project priorities that is adequate to the formulated conditions. In this case, we will proceed from the following.

1. The higher the efficiency of a project, the higher its priority.
2. If the cost of a project is high, then its inclusion in the portfolio reduces the chances of other projects entering this portfolio.
3. Of two projects of equal cost, all other things being equal, the one with a higher intensity of budget development is less preferable.
4. All other things being equal, a project that requires fewer total key personnel resources is preferable.

Let's construct an expression for calculating the project priority as a dimensionless value. The numerator should reflect positive factors. The denominator should reflect negative factors. Following point 1, we put d_i in the numerator project efficiency. Let's consider what should be in the denominator. In accordance with points 2 and 3, the denominator should contain the intensity of the project budget development:

$$c_i = C_i / T_i. \tag{1}$$

In accordance with paragraph 4, it is necessary to take into account the resource intensity of projects. It is logical to define the resource intensity of a project as the sum of the resources of key specialists required for its implementation:

$$R_i = \sum_{j \in M_i} s_{ij}. \tag{2}$$

where M_i – a set of numbers of key employees whose resources are needed by the project P_i.

Let us agree to characterize the relative resource intensity of the project P_i the ratio of the total volume of resources of key specialists required for its implementation to the total volume of resources that can be used to implement projects:

$$r_i = \frac{R_i}{\sum_{j=1}^{n} s_j}. \tag{3}$$

The relative resource intensity of a project determined in this way shows what share the given project requires from the total volume of resources of key specialists allocated to all n projects. The relative resource intensity can also be expressed as a percentage. To do this, it is sufficient to multiply the right-hand side in formula (3) by 100%.

Finally, we obtain the following expression for the project priority:

$$w_i = \frac{d_i}{c_i r_i}. \tag{4}$$

Thus, if the goal of the project is to improve the BP, then it is proposed to adopt as its priority the ratio of the project efficiency to the product of the project's budget development rate and its relative resource intensity. Let us recall that the efficiency of the project is understood as an increase in the intensity of income generation of the BP, the improvement of which the project is aimed at.

2.2. Determining the order of launching projects

Since, according to the condition, the budget is less than the total cost of the projects, at the first stage it is proposed to begin implementing a portfolio of $n_1 < n$ projects, the sum of the priorities of which is maximum, and the following restrictions are met: the total cost does not exceed the allocated budget, and the total resource intensity of these projects does not exceed the project resource of any key specialist involved in them.

To solve this problem, we can use the linear programming method. Note that if there are dependencies between projects, the problem is no longer a linear programming problem. Let us introduce n binary variables X_i. The value of each X_i is 1 if project i is included in the portfolio, and 0 otherwise. The problem is reduced to maximizing the objective function

$$\sum_{i=1}^{n} w_i X_i \to \max \tag{5}$$

under one budget constraint

$$\sum_{i=1}^{n} C_i \le B^1 \tag{6}$$

and m restrictions on the workload of key specialists

$$\sum_{i=1}^{n} s_{ij} X_i \le s_j. \quad (7)$$

This is a linear programming problem that is solved by the simplex method. As a result of the solution, we will obtain a certain set of projects that are subject to launch at the first stage. Let us designate the set of their numbers as N_1. Their total cost is equal to

$$C^1 = \sum_{i \in N_1} C_i. \quad (8)$$

Let us agree to consider the possible beginning of the second stage of project implementation as the moment when at least one project is completed. Let us also agree to consider that if one project is completed and there is a small time lag left until the end of another project, then the beginning of the stage is postponed until the end of the second project. The size of the lag is determined by the program manager, based on the condition that the lag should not exceed the number of working days set by the program curator.

The remainder of the initial budget for the second stage will be

$$\Delta B^1 = B^1 - C^1. \quad (9)$$

Obviously, this budget will not be enough. The company's management must decide to allocate an additional budget. The additional budget is taken from the funds that the BP will generate during the time t_{12} from the beginning of the first stage to the beginning of the second stage.

Note that the funds are taken from the income that the initially profitable BP bring in. Therefore, the fact that the effect of optimization does not occur immediately and does not jump affect only the volume of allocated funds.

It is reasonable to assume that this budget will be proportional to time with a certain coefficient q_2:

$$B_e^1 = q_2 t_{12}. \quad (10)$$

As a result, we will find that the budget for the second stage will be

$$B^2 = \Delta B^1 + B_e^1. \quad (11)$$

Reasoning similarly, we can write that the budget of stage number k is determined by the formula

$$B^k = \Delta B^{k-1} + B_e^{k-1}. \quad (12)$$

Here

$$B_e^{k-1} = q_\kappa t_{(k-1)k}, \quad (13)$$

where

$t_{(k-1)k}$ – time from the beginning of stage $(k-1)$ to the beginning of stage k.

Let us determine what the resource constraints will be at the second stage.

At the first stage, the workload of specialist j is equal to the total demand for his resource for all projects launched at the first stage:

$$S_j^1 = \sum_{i \in N_1} s_{ij}. \quad (14)$$

Let us assume that the second stage begins after the completion of project number f, from among the projects of the first stage. The completion of this project will free up the resource s_{fj} from specialist j. Accordingly, at the second stage the available resource of this specialist will be equal to

$$S_{2j} = s_j - S_j^1 + s_{fj}. \quad (15)$$

If the second stage begins after the completion of several projects, the set of numbers of which is N_{1F}, then the available resource of specialist j will be equal

$$S_{2j} = s_j - S_j^1 + \sum_{i \in N_{1F}} s_{ij}. \quad (16)$$

Reasoning in a similar way, we obtain the following. If at the beginning of stage k the projects with a set of numbers of , and the stage begins upon completion of the projects with a set of numbers of N_{kF}, then the available resource of specialist j will be

$$S_{kj} = s_j - \sum_{i \in N_k^*} s_{ij} + \sum_{i \in N_{kF}} s_{ij}. \quad (17)$$

The problem of determining projects that were not started at previous stages and which will be launched at stage $k > 1$ is reduced to a linear programming problem, similar to how it was done earlier for stage 1. In this case, the summation in the objective function should be carried out only over the set of numbers N^k of those projects that were not started at previous $(k - 1)$ stages:

$$\sum_{i \in N^k} w_i X_i \to \max. \tag{18}$$

The budget constraint for stage k is

$$\sum_{i \in N^k} C_i \le B^k, \tag{19}$$

where

B^k – budget of stage k, determined by formula (12).

At all stages, the number of resource constraints is equal to m. For specialist j at stage k, it has the form

$$\sum_{i \in N^k} s_{ij} \le S_{kj}, \tag{20}$$

where

S_{kj} – the resource of specialist j available at the beginning of stage k, determined by formula (17).

Note that it is possible that an attempt to start the next stage after completing one or even several projects may be impossible due to insufficient resources. In this case, it is necessary to wait for the completion of the following projects. Let us apply the obtained results to solving a real problem.

3. Example

The company's strategic goal is "To optimize the main BPs." Ten such BPs have been identified. Ten projects have been initiated to achieve the goal. Five key specialists should be involved in their implementation. Following the style adopted in the company, we will say that the allocated resource of each specialist is 100%. Accordingly, the resource intensity of the projects will be set as a percentage, as is done in *Table 1.*

The characteristics of the projects that determine their priorities are given in *Table 2.*

Initially allocated budget B^1 = 10 00 000 rubles is three times less than the total cost of the projects that need to be implemented. In addition, the total resource requirements of the projects exceed 100% for each of the specialists. Therefore, the projects will be launched

Table 1.

Projects' need for specialist resources (%)

Projects	Specialists				
	1	2	3	4	5
1	29	8	31	41	50
2	44	11	1	39	22
3	36	38	27	36	43
4	1	18	21	8	7
5	15	16	2	2	27
6	24	7	21	18	33
7	15	45	48	45	5
8	50	7	43	41	33
9	18	29	13	32	4
10	10	19	28	48	17

Table 2.

Project characteristics

Projects	Relative resource intensity	Planned implementation time (days)	Cost (thousand rubles)	Budget utilization rate (thousand rubles/day)	Efficiency (thousand rubles/day)	Priority
	r_i	T_i	T_i	c_i	d_i	w_i
1	0.318	159	3 665	23	22.5	3.07
2	0.234	74	648	9	3.9	1.90
3	0.36	126	6 333	50	32.5	1.80
4	0.11	144	6 896	48	50	9.49
5	0.124	73	1 071	15	5.2	2.86
6	0.206	52	5 590	108	32.9	1.49
7	0.316	129	691	5	6.6	3.90
8	0.348	36	1 418	39	10.4	0.76
9	0.192	84	3 822	46	20.1	2.30
10	0.244	66	4 076	62	20.2	1.34

in stages. It was decided that for each stage $k > 1$ to the remaining budget will be added an amount proportional to the duration of the previous stage with a constant coefficient $q = 100\,000$ rubles.

Since all projects are aimed at achieving one strategic goal, it can be said that they form a program of projects. It is necessary to determine the order of implementation of this program.

Following the proposed approach, we define the projects that should be launched at the first stage using the simplex method. We find that these are projects with numbers $N_1 = \{2; 4; 5; 7\}$. Their total cost is $C^1 = 9\,306\,000$ rubles. We proceed to the second stage. It can be started after the completion of project 5. However, given that project 2 should be completed only one day later, a decision is made to start the second stage after its completion. Thus, the duration $t_{12} = 74$ days. The budget of the second stage will be in accordance with formula (11) $B^2 = (10\,000\,000 - 9\,306\,000) + 100\,000 \cdot 74 = 8\,094\,000$ rubles.

We determine the resources of specialists using formula (17). We again solve the linear programming problem with binary variables whose numbers belong to the set $N^2 = \{1; 3; 6; 8; 9; 10\}$. We find that project 1 should be launched at the second stage.

The program will require six stages to be implemented, details of which are provided in *Table 3*.

Figure 1 shows the roadmap of the considered program for optimizing the company's business process.

Table 3.

Stages of program implementation

Stage	Budget (million rubles)	Specialist resources (%)					Launched projects	Projects, after the completion of which the stage begins	Duration of stage (days)
		1	2	3	4	5			
k	B^k	$S_1^{\ k}$	$S_2^{\ k}$	$S_3^{\ k}$	$S_4^{\ k}$	$S_5^{\ k}$	N_k	N_{kF}	$t_{(k-1)}k$
1	10	100	100	100	100	100	2; 4; 5; 7		74
2	8.094	84	37	31	47	88	1	2.5	55
3	9.929	70	74	48	51	43	6; 9	7	84
4	8.917	53	63	56	27	46	3	4; 6; 9	20
5	4.584	71	92	69	59	50	10	1	66
6	7.108	64	62	73	64	57	8	10	40

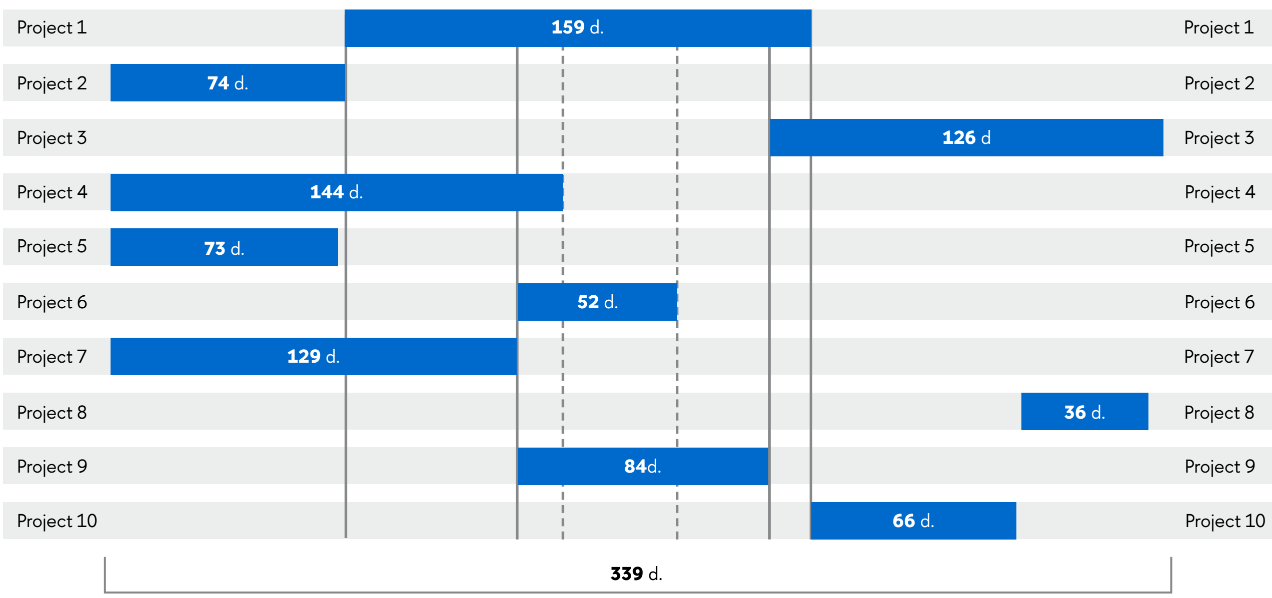


Fig. 1. Roadmap of the company's business process optimization program.

Conclusion

According to the authors, the following are new in this work:

- expression for calculating the priority of projects, reflecting the fact that the projects are aimed at optimizing the business processes that generate the budget for their implementation;
- a method for step-by-step determination of the order of implementation of the entire program aimed at optimizing the business processes that generate the company's income.

The work was initiated to meet the current need of enterprises that ensure the airworthiness of civil aviation aircraft. The results obtained are also applicable in other sectors of the national economy.

The article is of interest to managers of project programs of manufacturing and service companies, as well as to a wide range of researchers and postgraduate students of technical and economic specialties. The further development of the approach is the consideration of the situation of the presence of dependencies between the projects of the program. ■

References

1. International Institute of Business Analysis (2015) *A guide to the business analysis body of knowledge (Babok Guide). Version 3.0.*
2. Gromov A.I., Fleishman A., Schmidt V. (2016) *Management business processes. Modern methods*. Moscow: Yurait (in Russian).
3. Lapkin A., Piatkov V. (2018) Optimization of business processes based on system analysis. *IOP Conference Series: Materials Science and Engineering*, no. 451, pp. 12–36.
4. Davis R. (2004) *Business process modeling with ARIS. A practical guide*. London: Springer.
5. Jahangirian M., Eldabi T., Naseer A., et al. (2010) Simulation in manufacturing and business: A review. *European Journal of Operational Research*, vol. 203, no. 1, pp. 1–13. https://doi.org/10.1016/j.ejor.2009.06.004
6. Hammer M., Champy J. (1997) *Reengineering the corporation: A manifesto for business revolution*. St. Petersburg: St. Petersburg State University (in Russian).
7. Cheremnykh O.S., Cheremnykh S.V. (2005) *Strategic corporate reengineering: process-cost approach to business management*. Moscow: Finance and Statistics (in Russian).
8. Ishmuradova I.I., Shagaipov D.R. (2016) Reengineering of business processes at enterprises producing bakery products. *Economy, entrepreneurship and law*, no. 1, pp. 37–48 (in Russian).
9. Barkalov S., Voropaev V.I., Sekletova G.I (2005) *Mathematical foundations of project management*. Moscow: Higher School (in Russian).
10. Trofimov V.V., et al. (2006) *Project management with Primavera*. St. Petersburg: St. Petersburg State University of Economics and Finance (in Russian).

11 Pritsker A.A.B., Happ W.W. (1966) Gert: Graphical Evaluation and Review Technique. Part 1. Fundamentals. *Journal of Industrial Engineering*, vol. 17, no. 5, pp. 267–274.

12. Pritsker A.A.B., Whitehouse G.E. (1966) Gert: Graphical Evaluation and Review Technique. Part 2. Applications. *Journal of Industrial Engineering*, vol. 17, no. 5, pp. 293–301.

13. Itskovich A.A., Fainburg G.D., Fainburg I.A., Chernov A.O. (2018) System of processes and projects for maintaining airworthiness of aircraft. *Scientific Bulletin of Moscow State Technical University of Civil Aviation*, vol. 21, no. 1, pp. 164–173 (in Russian).

14. Tsipes G.L., Tovb A.S. (2009) *Projects and project management in modern companies*. Moscow: Olimp-Business (in Russian).

15. Global Alliance for Project Performance Standards (2006) *A framework for performance-based competency standards for global level 1 and 2 project managers*.

16. Project Management Institute (2021) *A guide to the project management body of knowledge (PMBOK Guide) – 7th edition*.

17. Standartinform (2015) *GOST R ISO 21500-2014 Guidelines for project management.* Moscow: Standartinform (in Russian).

18. Project Management Institute (2017) *The standard for portfolio management – Fourth edition*.

19. Office of Government Commerce (2009) *Managing successful projects with PRINCE2.* London: The Stationery Office.

20. Vargas R.V. (2010) Using the analytic hierarchy process (AHP) to select and prioritize projects in a portfolio. *PMI Global Congress*, vol. 32(3), pp. 1–22.

21. Saaty T.L. (1990) *The analytic hierarchy process: Planning, priority setting, resource allocation*. RWS Publications.

22. Kaplan R.S., Norton D.P. (1992) *The balanced scorecard – measures that drive performance*. Harvard Business Review.

23. Kaplan R., Norton D. (2008) *Balanced Scorecard. From Strategy to Action*. Moscow: Olimp-Business (in Russian).

24. Kuznetsova E.V. (2024) *Project portfolio management as a tool for implementing corporate strategy: A textbook for universities*. Moscow: Yurait (in Russian).

25. Demidovich B.P., Maron I.A. (2006) *Fundamentals of computational mathematics*. St. Petersburg: Lan (in Russian).

26. Tikhomirova A.N., Sidorenko E.V. (2012) Modification of the method of analysis of hierarchies of T. Saaty for calculating the weights of criteria in the evaluation of innovative projects. *Modern Problems of Science and Education*, no. 2 (in Russian).

27. Williams T., Eden C., Ackermann F., Tait A. (1995) The effects of design changes and delays on project costs. *Journal of the Operational Research Society*, vol. 46, no. 7, pp. 809–818. https://doi.org/10.1057/jors.1995.114

28. Zuikov K.A. (2012) Project sustainability. An approach based on system dynamics. *Project and program management*, no. 3(31), pp. 178–187 (in Russian).

29. Kuzmin E.A. (2012) Identification of risks in project management using the method of analyzing balances of factors and deviations. *Financial risk management*, no. 3, pp. 200–214 (in Russian).

30. Lesnykh V.V., Litvin Yu.V. (2013) On the assessment of the significance of the implementation of project work. *Audit and financial analysis*, no. 4, pp. 254–260 (in Russian).

31. Maron M.A. (2016) The choice of control points of projects taking into account possible change of structure of works. *Business Informatics*, no. 2(36), pp. 57–61. https://doi.org/10.17323/1998-0663.2016.2.57.62

32. Maron M.A. (2018) Diagnostics of projects. *European Research Studies Journal*, vol. 21, no. 1, pp. 18–30.

33. Jaafari A., Jabari N. (2008) Project diagnostics as a means of accelerating its implementation. *Project and program management*, no. 2, pp. 128–139 (in Russian).

34. Itskovich A.A., Fainburg I.A., Fainburg G.D. (2019) Methodological aspects for controlling the processes that secure the reliability of aviation engineering. *Journal of Machinery Manufacture and Reliability*, vol. 48, no. 5, pp. 446–455.

35. Dmitrenko I.E., Sapozhnikov V.V., Dyakov D.V. (1994) *Measurements and diagnostics in railway automation, telemechanic and communication systems*. Moscow: Transport (in Russian).

36. Anshin V.M., Manaykina E.S. (2015) Formation of a company's project portfolio based on the principles of sustainable development. *Bulletin of the Institute of Economics of the Russian Academy of Sciences*, no. 1, pp. 126–140 (in Russian).

37. Isaev D.V. (2020) Evaluation of performance management systems development programs based on the maturity model. *Applied Informatics*, vol. 15, no. 3, pp. 5–18 (in Russian).

About the authors

Arkadiy I. Maron

Candidate of Technical Sciences, Senior Researcher;

Associate Professor, School of Applied Mathematics, HSE Tikhonov Moscow Institute of Electronics and Mathematics (MIEM HSE), 34, Tallinskaya St., Moscow 123458, Russia;

E-mail: amaron@hse.ru

ORCID: 0000-0003-4443-3329

Grigoriy D. Fainburg

Candidate of Technical Sciences, Associate Professor;

Associate Professor, Department of Technical Operation of Aircraft and Aircraft Engines, Moscow State Technical University of Civil Aviation, 20, Kronstadtsky Blv., Moscow 125993, Russia;

E-mail: fain2000@mail.ru

ORCID: 0000-0001-8004-1969

DOI: 10.17323/2587-814X.2025.1.34.49

Modeling and optimization of the characteristics of intelligent transport systems for "smart cities" using hybrid evolutionary algorithms

Evgeniy A. Zaripov
E-mail: e.a.zaripov@ya.ru

Andranik S. Akopov
E-mail: akopovas@umail.ru

Central Economic Mathematical Institute, Moscow, Russia
MIREA – Russian Technological University, Moscow, Russia

Abstract

Modern cities are facing increasing traffic congestion, necessitating the implementation of intelligent traffic management systems. One of the key areas in this field is adaptive traffic signal control, which can adjust to changing traffic conditions. However, existing methods for optimizing traffic signal cycle parameters have several limitations, such as high computational complexity, the risk of premature convergence of algorithms and the difficulty of accounting for traffic dynamics. This study proposes an approach to optimizing the characteristics of an intelligent transportation system using hybrid evolutionary algorithms. The methods we developed combine the principles of genetic algorithms (GA) and particle swarm optimization (PSO), enabling a balance between global and local search for optimal parameters. The research examines six different hybridization schemes, including modified versions of basic algorithms, as well as their integration with HDBSCAN clustering methods for adaptive optimization frequency tuning. To evaluate the effectiveness of the proposed algorithms, a simulation model was developed in the AnyLogic environment, replicating real urban traffic conditions. Numerical experiments conducted on a local section of the road network in Moscow demonstrated that the hybrid SlipToBest algorithm achieves the best results in reducing average travel time and fuel consumption, while the Alternating algorithm (structured switching between GA and PSO) ensures high solution stability. The results of this study confirm the feasibility of using hybrid evolutionary methods for traffic flow management tasks. The proposed algorithms not only enhance the efficiency of traffic signal control but also establish a foundation for the further development of adaptive urban traffic management systems.

Keywords: intelligent transport systems, traffic infrastructure management, smart city, hybrid evolutionary algorithms, simulation modeling, traffic management, AnyLogic

Citation: Zaripov E.A., Akopov A.S. (2025) Modeling and optimization of the characteristics of intelligent transport systems for "smart cities" using hybrid evolutionary algorithms. *Business Informatics*, vol. 19, no. 1, pp. 34–49. DOI: 10.17323/2587-814X.2025.1.34.49

Introduction

In the context of growing urbanization, adaptive traffic flow management has become an essential tool for improving road network efficiency. One of the key areas in this regard is traffic signal optimization, which enables dynamic phase adjustments based on traffic intensity. However, determining the optimal parameters for traffic signal cycles presents a complex optimization problem characterized by high dimensionality, nonlinearity and stochastic variations in traffic flows.

Traditional analytical methods and heuristic algorithms often prove ineffective in solving such problems due to their inability to adapt to changing traffic conditions and their high computational complexity. As a result, increasing attention has been given in recent years to evolutionary algorithms such as the Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). These methods allow for approximate optimal solutions without requiring gradient calculations, making them particularly attractive for optimization problems in complex systems.

Despite their proven effectiveness, both approaches have limitations. GA provides good solution diversification but often converges slowly. PSO, on the other hand, demonstrates rapid convergence but is prone to premature convergence to local minima. To overcome these shortcomings, hybrid methods that combine the strengths of GA and PSO are actively being developed. Hybridization allows for a balance between global and local search processes, which is particularly important for dynamic systems such as traffic signal control.

This study proposes an improved hybrid evolutionary optimization method for traffic signal regulation. Various integration schemes of GA and PSO are considered, aimed at enhancing the accuracy and stability of solutions. A key element of the proposed approach is a modified swarm algorithm, "SlipToBest," which introduces a directed shift of particles toward the best-found solution to accelerate the optimization process. Additionally, the "Alternating" method is analyzed, which involves alternating GA and PSO to achieve a balance between global and local search strategies.

Furthermore, the study explores the potential application of the HDBSCAN clustering algorithm for analyzing traffic flow density and dynamically adjusting optimization frequency. This reduces computational costs and enhances the adaptability of the algorithms to changing traffic conditions.

The effectiveness of the methods we developed is evaluated using a traffic flow simulation model built in the AnyLogic environment. The experiments are conducted using real-world data from the road network of Moscow, allowing for an objective assessment of the proposed algorithms' impact on average travel time and fuel consumption.

The results demonstrate that the proposed hybrid methods significantly outperform classical algorithms. SlipToBest shows the best performance in terms of convergence speed and travel time reduction, while Alternating ensures solution stability even under high traffic variability. The inclusion of the HDBSCAN clustering method further enhances the adaptability of the control system.

Thus, the study confirms the effectiveness of hybrid evolutionary algorithms for intelligent traffic flow management and highlights the necessity for their further improvement to enhance the quality of urban transportation infrastructure.

1. Formulation of the optimization problem for an intelligent transportation system

The optimization of traffic signal control parameters is complicated by high dimensionality, nonlinearity and the stochastic nature of traffic flows, as confirmed by research findings [1–5]. Under such conditions, classical methods, including gradient-based algorithms and heuristics, often require significant model simplifications, reducing the accuracy of the solutions obrained. At the same time, evolutionary algorithms demonstrate the ability to perform global search and adapt to changing conditions [4, 7, 9], making them preferable for solving such problems. The most widely used evolutionary optimization methods are Genetic Algorithms (GA), which utilize selection, crossover, and mutation mechanisms to search for optimal solutions, and Particle Swarm Optimization (PSO), which simulates the collective behavior of agents in the search space [3–6]. GA effectively explores the solution space and provides high diversity but may exhibit slow convergence. In contrast, PSO achieves high search speed but is prone to premature convergence to local minima [7, 10–12].

To overcome these limitations, this study proposes a hybridization of GA and PSO that combines their strengths. Several hybridization schemes are considered, including **Alternating** (alternating GA and PSO stages), **SlipToBest** (an additional shift towards the best-found solution in PSO), **MixIntegrate** (integration of genetic operators into the swarm search process), and **Mix** (a simple combination without a rigid structure). As demonstrated in several studies, hybrid methods improve search efficiency, enhancing both convergence speed and solution quality [7, 10–12].

The optimization of traffic signal control parameters in a multi-agent transportation system is formulated as a minimization problem of an objective function that defines the quality of the transportation process. Let $X = (x_1, x_2, \ldots, x_n)$ be the parameter vector, including the durations of red and green traffic light phases. Constraints on phase parameters are defined in accordance with technical regulations, determining the search space: $x_i \in [L_i, U_i]$, $i = 1, 2, \ldots, n$, where L_i, U_i are the lower and upper allowable durations of traffic signals, and n is the total number of controlled signals.

The objective function is defined as the average travel time of vehicles passing through the studied road section during a given simulation period. Let this value be denoted as $f(x)$. Its computation is performed using a simulation model in the AnyLogic environment, which replicates vehicle movement for each parameter set and records travel time and fuel consumption indicators. The average travel time serves as an integral characteristic of traffic flow quality, as it reflects the impact of traffic intensity, intersection throughput and traffic signal control logic [3–5].

In addition to average travel time, additional indicators such as fuel consumption and other economic and environmental factors are considered. However, the primary optimization goal is to minimize travel time, as it is the key criterion for evaluating the efficiency of traffic signal control. This formulation allows for an objective assessment of the impact of different algorithms on traffic processes and ensures a correct comparison of their effectiveness. Thus, the problem is formally defined as:

$$\min_{x_i \in X} f(x_i). \quad (1)$$

subject to the constraints:

$$x_i \in [L_i, U_i],\ i = 1, 2, \ldots, n.$$

This problem formulation is widely used in transportation system studies [1, 2] since minimizing average travel time directly correlates with improving network throughput, reducing congestion and optimizing fuel consumption. The use of this objective function is justified by fundamental studies on evolutionary adaptation [4] and research in transportation optimization [7, 9]. It has been proven that evolutionary algorithms, including GA and PSO, efficiently minimize such functions without requiring analytical gradients, allowing for the modeling of complex dynamic processes [3–6, 9]. Thus, the minimization of this function serves as the primary criterion for selecting the optimal set of traffic signal control parameters, ensuring shorter travel times and lower transportation costs.

The problem (1) can, in particular, be solved using the **PSO algorithm**, due to its ability to efficiently search for solutions without requiring gradient information [3, 13]. This approach simulates the collective behavior of particles that exchange information about discovered improvements and move toward the global optimum.

Each particle in PSO has a current position, a velocity that defines the rate of parameter change, a personal best position, and a best global position. At each iteration, particles adjust their velocity and position based on acquired information using update procedures [5, 6, 13, 14]. One of the algorithms used in this study (**Swarm**) is also implemented based on this classical PSO framework. This algorithm serves as a baseline for comparing the effectiveness of hybrid modifications. Shi and Eberhart [6] proposed a strategy of dynamically changing the inertia weight (variable w) to balance between exploration of the solution space and exploitation of already found solutions. Clerc and Kennedy [5] introduced the concept of constriction coefficients, ensuring stable convergence. Recent studies continue to explore optimal parameter settings for various problem classes, including transportation optimization [1, 2, 9, 14, 15]. PSO has proven to be an effective global search method, demonstrating successful applications in transportation optimization tasks.

Additionally, problem (1) can be solved using the **Genetic Algorithm**, an evolutionary optimization method based on selection, crossover and mutation mechanisms that mimic the natural evolution process. Initially proposed by Holland [4], GA has demonstrated efficiency in optimizing complex, high-dimensional search spaces [7, 16–18]. In this study, GA is used to optimize traffic signal control parameters, where each individual represents a parameter vector $x = (x_1, x_2, \ldots, x_n)$, defining the durations of traffic light phases.

The initial population is generated randomly or using heuristics, followed by iterative selection, crossover and mutation processes to improve solutions. The crossover operation involves combining parameters from two parent individuals to create offspring. Mutation represents a random modification of individual parameters, preventing premature convergence of the algorithm. Mutation plays a crucial role in exploring new regions of the search space and enhancing solution quality. GA ensures a global search capability, allowing for effective exploration of complex search spaces. In this study, a classical GA implementation (**Parallel**) is considered, which employs basic crossover and mutation mechanisms. This algorithm serves as a foundation for further hybrid modifications that combine GA with PSO.

The third and most promising approach to solving problem (1) is based on the **hybridization of PSO and GA**. This hybridization leverages the strengths of both methods: GA's ability to perform global search and overcome local minima, along with PSO's capability to rapidly refine solutions and move toward optimal points in the search space. Such a combination improves both convergence speed and solution quality, especially in complex dynamic systems such as transportation infrastructure [1, 2, 9, 10].

This study considers four implemented hybrid schemes, differing in the principle of operator combination and optimization strategy.

- **Alternating (GA and PSO alternation).** This approach involves the sequential application of GA and PSO at different iterations of the algorithm. Let the number of iterations be defined as ITERATIONS. On even steps, GA is applied, while on odd steps, PSO is used (or vice versa, depending on the initial conditions):

$$\begin{aligned} iter mod 2 = 0 &\Rightarrow apply\ GA, \\ iter mod 2 = 1 &\Rightarrow apply\ PSO, \end{aligned} \quad (2)$$

Such combination strategies have been mentioned in reviews on hybrid methods [9, 19–21]. Alternating helps maintain a balance between global and local search: GA periodically updates solutions, increasing diversity, while PSO efficiently refines candidate solutions.

- **SlipToBest (PSO + alpha).** This is a proposed PSO modification in which, after updating their positions, each particle is additionally shifted toward the globally best solution by a parameter α. That is, the value of the j-th target variable of the i-th particle can be computed as:

$$x_{i,j}(t+1) = x_{i,j}(t) + \alpha\left(g_j^{best}(t) - x_{i,j}(t)\right), \quad (3)$$

which defines a directed shift toward the best solution. This method accelerates convergence but may reduce solution diversity if the parameter α is set too high [6, 11]. Here,

$x_{i,j}$ is the value of the j-th target variable of the i-th particle at iteration t;

$g_j^{best}(t)$ is the globally best value of the j-th target variable found by all swarm particles up to iteration t.

- **MixIntegrate (Genetic Operators within PSO).** This approach **integrates genetic operators** (crossover and mutation) directly into the swarm update stage. After executing the standard PSO step, genetic operators are applied to **selected particle pairs**:

$$x_{j,new} = crossover(x_{i,j}, \bar{x}_{\bar{i},j}),\ x'_{j,new} = mutate(x_{i,new}), \quad (4)$$

where

$\{x_{j,new}, x'_{j,new}\}$ are the new values of the j-th target variable obtained from the crossover and mutation operations;

$\{x_{i,j}, \bar{x}_{\bar{i},j}\}$ are the values of the j-th target variable for the i-th and $\bar{i}$-th particles in the swarm.

  - **The integration of genetic operators into PSO** helps maintain solution diversity and prevents premature convergence of the swarm. Similar strategies have been discussed in hybrid optimization research [7, 9].

- **Mix (Simple Combination of PSO and GA).** This is a less structured hybridization approach in which PSO and GA are applied without fixed rules or systematic use of operators: Mix: Apply PSO, then GA, or vice versa, without a fixed rule.

  - **A simple Mix scheme**. This approach has been mentioned in the literature as an example of unstructured hybridization, which does not always outperform pure PSO or GA [10, 12, 20].

To assess traffic flow density, this study employs the **HDBSCAN** (Hierarchical Density-Based Spatial Clustering of Applications with Noise) hybrid clustering algorithm [21], which extends the DBSCAN method hierarchically. This algorithm identifies groups of points (clusters) based on their density and automatically determines the optimal number of clusters, making it more flexible compared to classical methods such as k-means.

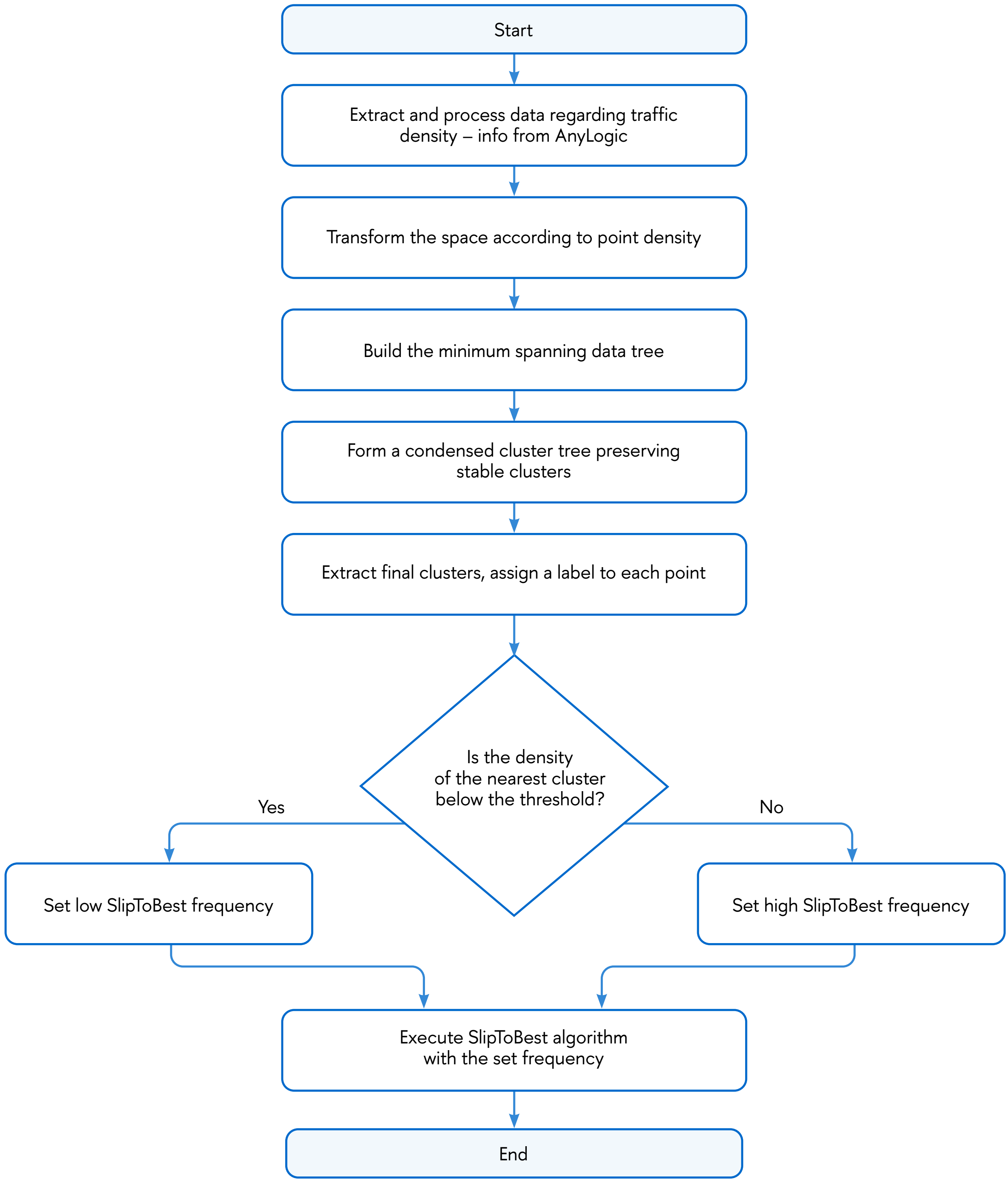


Fig. 1. High-level workflow diagram of the HDBSCAN algorithm integrated with the AnyLogic simulation model.

The key advantage of HDBSCAN is its ability to account for the dynamic structure of traffic flows, automatically detecting areas of high and low traffic density. This is particularly important for traffic analysis, where movement intensity can fluctuate significantly depending on the time of day, road infrastructure and unforeseen factors (e.g., accidents, roadworks). The algorithm also identifies noise points, which may correspond to anomalous traffic situations such as congestion or sudden traffic pattern changes.

To integrate HDBSCAN with the AnyLogic simulation model, a workflow diagram of the algorithm has been developed (*Fig. 1*). At each simulation step (at time *t*), the following stages are executed:

1. Collecting and processing traffic density data obtained from the "smart" traffic light system.

2. Analyzing the spatial distribution of vehicles around the traffic light, considering their density.

3. Constructing a minimum spanning tree to represent hierarchical cluster structures.

4. Evaluating cluster stability (analyzing density and cluster duration over time).

5. Forming a condensed tree to identify stable clusters.

6. Determining the final cluster structure and assigning appropriate labels.

HDBSCAN enables the adaptive adjustment of the activation frequency of the SlipToBest hybrid algorithm:

- If the nearest cluster to the traffic light has low density, the system reduces the frequency of SlipToBest application (minimizing computational load).
- If the cluster density exceeds a threshold value, SlipToBest is triggered more frequently, ensuring a prompt response to changing traffic conditions.

Thus, the HDBSCAN algorithm dynamically adapts traffic signal phase control, allowing for more efficient resource allocation and a more precise consideration of fluctuating road conditions.

2. Implementation of the simulation model in AnyLogic

For the listed hybrid algorithms, parameters such as the coefficient, the frequency of GA operator application, and the number of PSO iterations between genetic stages are selected empirically or based on recommendations from scientific literature [6, 7, 9–12, 19–22]. To ensure an objective comparison of their effectiveness, a unified simulation environment (AnyLogic) and a common objective function (see (1)) are used, guaranteeing the correctness of the analysis. The use of AnyLogic in adaptive traffic management tasks has previously demonstrated its effectiveness [23].

The AnyLogic model so developed extends the capabilities of classical PSO and GA, providing a flexible and efficient search for optimal traffic signal parameters. The following sections present the results of applying hybrid algorithms and a comparative analysis of their effectiveness.

During the refinement of the simulation model, improvements were made to bring it closer to real-world traffic conditions and to simplify integration with external optimization algorithms. Vehicles now switch lanes considering lane priority, the speed of adjacent vehicles, and the distance to stop lines. These modifications have increased the realism of the simulation, making the behavior of transport agents closely resemble real urban environments.

The model's variables and parameters, including the durations of red and green traffic light phases, were structured in a way that allows them to be easily read, modified and analyzed from external code. This significantly simplified integration with evolutionary algorithms and enabled mass iterations to assess the quality of traffic management decisions.

The model's structure and software implementation were optimized so that adding or removing traffic signals does not require significant changes to the source code. This flexibility allows the model to be adapted to different scenarios, modify the road network topology and adjust the optimization area without substantial time costs.

The developed model is prepared for integration with external optimization code. Its modular architecture allows AnyLogic to be used in a “black box” mode, where algorithms receive traffic signal parameters, process them, and return optimized values that minimize average travel time and reduce fuel consumption. This approach makes the implementation of new optimization algorithms convenient and ensures an objective comparison of their performance.

Figure 2 presents the structure of the updated model of an urban road network segment created in AnyLogic. This model includes the placement of traffic signals and phase timing parameters, allowing for flexible testing of various adaptive control algorithms.

The simulation model allows for variations in key traffic signal control parameters that directly affect network capacity and traffic dynamics. *Figure 3* illustrates the parameter panel of a typical “smart” traffic light (trafficLight4).

The model defines phase durations (*redTime*, *greenTime*), links to stop lines and the traffic light’s operating mode. Depending on traffic intensity and intersection configuration, the model adjusts signal switching logic, determining optimal green and red phase intervals.

During experiments, the algorithm generates control decisions, determining the optimal $redTime_i$ and $greenTime_i$ for all traffic signals. These parameters are loaded into the model before the simulation starts. Then, AnyLogic simulates vehicle flow, recording average travel time and fuel consumption. Thus, the evolutionary algorithm receives an objective evaluation of control quality, analyzing the impact of parameters on the objective function.

In subsequent iterations, the algorithm updates traffic signal parameters, minimizing the average travel time.

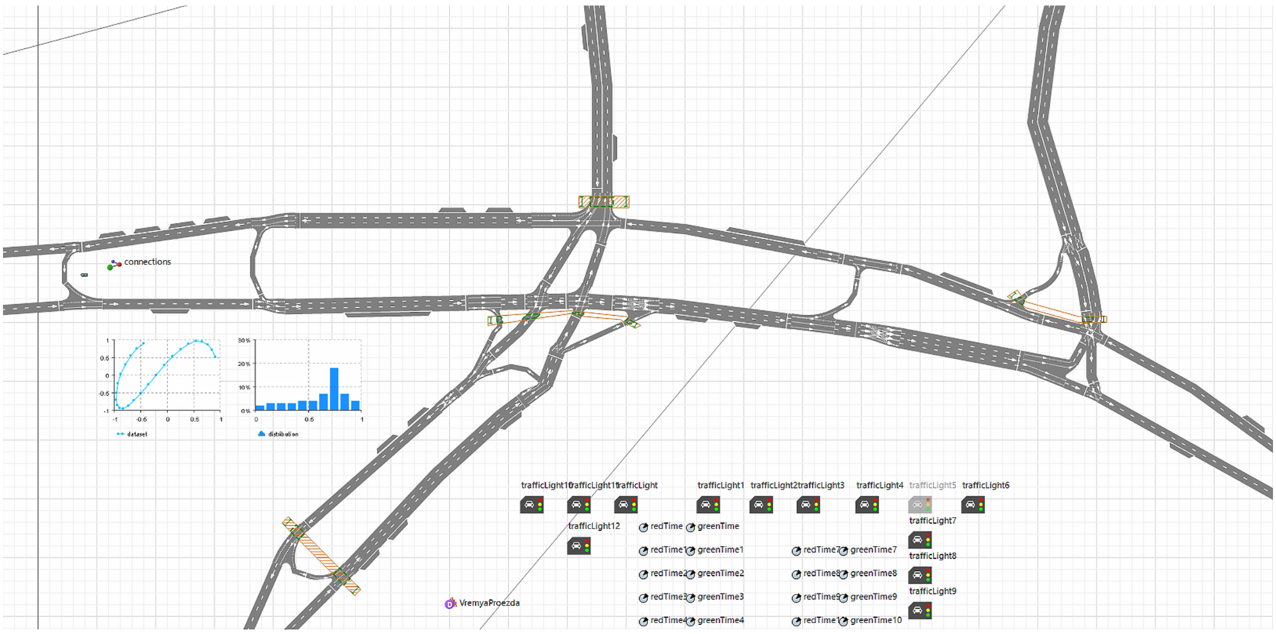


Fig. 2. Structure of the updated simulated urban road network segment in AnyLogic with placed traffic signals and phase timing parameters.

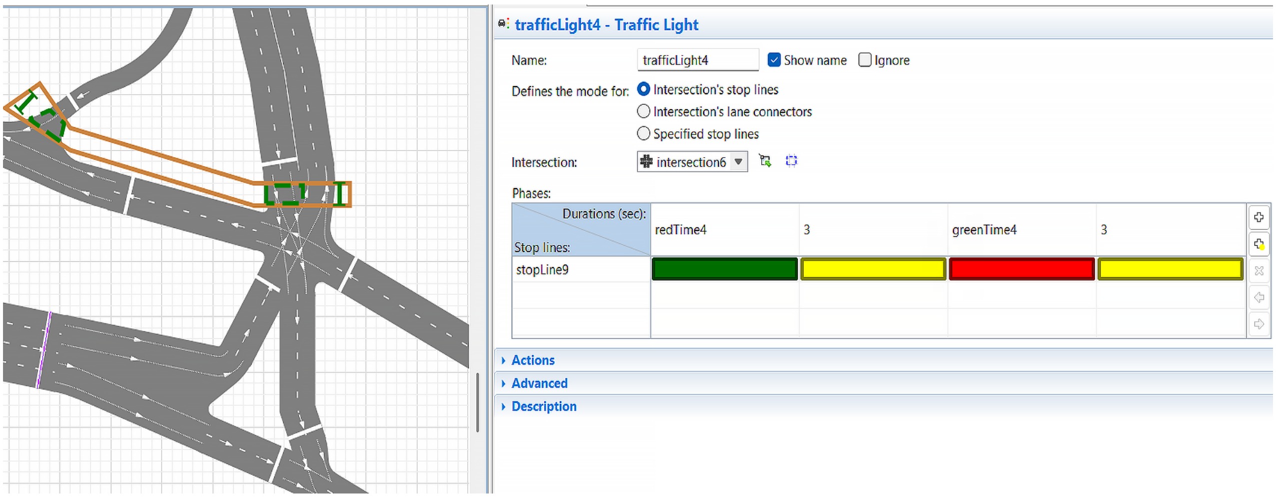


Fig. 3. Parameter panel of the "smart" traffic light (trafficLight4) in AnyLogic with phase durations and an attached stop line.

The model automatically adapts to parameter changes, allowing for dynamic traffic signal adjustments to various traffic conditions, including light traffic and heavy congestion.

3. Results of numerical experiments and analysis of algorithm efficiency

The traffic flow simulation model developed in this study for a local urban road network segment in AnyLogic was previously described in detail in the authors' works [1, 2]. The model considers the characteristics of road infrastructure, the placement of traffic lights and pedestrian crossings, and is based on an agent-based and discrete-event approach. The simulation parameters, including simulation duration, road network configuration and traffic flow intensity, were determined based on real traffic data near the Yugo-Zapadnaya metro station in Moscow [1].

During the experiments, the variable parameters were the durations of the red and green traffic light phases, as these characteristics have a critical impact on the network's capacity. Each phase configuration was evaluated based on the average travel time of vehicles and fuel consumption indicators, which were computed using simulation results.

For each of the examined algorithms (Alternating, SlipToBest, Swarm, Parallel, MixIntegrate, Mix), multiple simulation runs were conducted to reduce the influence of stochastic factors and assess solution stability. The stopping criterion was most often a fixed number of iterations or generations determined by computational resource constraints and empirical observations. Additionally, a threshold for lack of improvement was considered, beyond which the algorithm was terminated.

The analyzed algorithms differ in their structural composition and the way they combine swarm-based (PSO) and genetic (GA) approaches, which affects their effectiveness. To ensure an objective comparison, the following metrics were used:

- Average travel time, which characterizes the road segment's capacity and the level of convenience for drivers.
- Fuel consumption (AllLostGas), reflecting the economic and environmental efficiency of the system, as reducing unnecessary vehicle idling leads to lower fuel consumption.

Table 1.

Comparative results for average travel time and fuel consumption

Algorithm	Approach characteristics	Best average travel time (min)	Fuel consumption (L)
SlipToBest	Swarm algorithm with an additional shift toward the best solution (alpha parameter)	3.2213	288.5
Alternating	Alternating GA and PSO	3.2483	295.4
Swarm	Pure particle swarm optimization (PSO)	3.2246	296.2
Mix	Simple combination of PSO and GA, without strict alternation	3.2650	308.7
MixIntegrate	Swarm algorithm with GA operators (mutation, crossover)	3.2798	312.9
Parallel	Pure genetic algorithm (GA)	3.3009	318.5

The results of numerical experiments are summarized in *Table 1*, which presents the best average travel times achieved by each algorithm. *Figure 4* provides a graphical visualization of the effectiveness of the examined approaches.

Fuel consumption refers to the total excess fuel consumption (liters) for a given set of vehicles (e.g., 10 000 vehicles) participating in multiple simulation runs. The numerical values were averaged to account for stochastic fluctuations in road traffic, which arise from variations in acceleration, braking, and traffic density.

The analysis of data from *Table 1* shows that SlipToBest demonstrates the best results, achieving the lowest average travel time with minimal fuel consump-

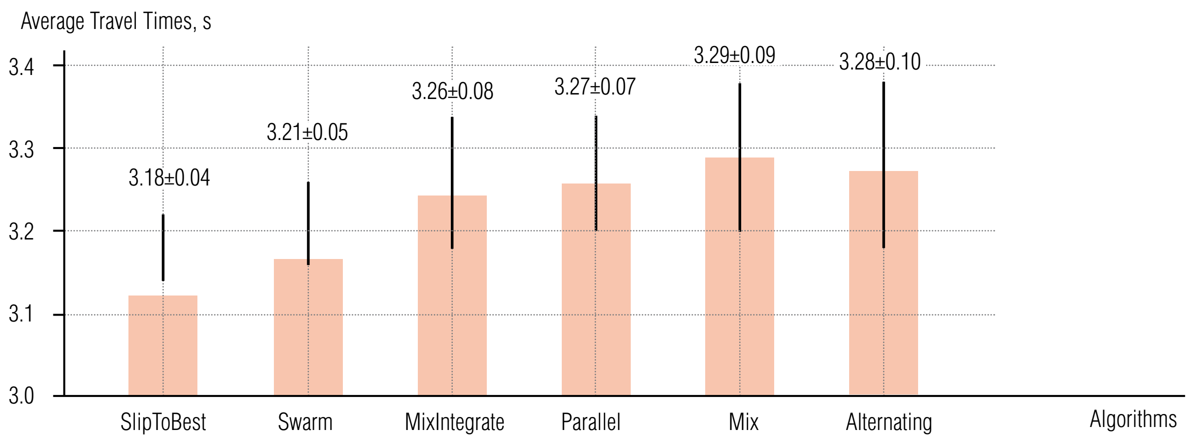


Fig. 4. Comparison of algorithm efficiency based on average travel time.

tion. The next most effective algorithms are Alternating and Swarm. Notably, Alternating, due to its alternating application of GA and PSO operators, outperforms both pure PSO (Swarm) and less structured hybrid approaches (Mix, MixIntegrate). The use of a pure genetic algorithm (Parallel) leads to the highest time and fuel consumption, indicating its insufficient efficiency without additional modifications.

Thus, targeted improvements to the particle swarm algorithm (such as introducing the alpha parameter in SlipToBest) or rational alternation of evolutionary operators (as in Alternating) provide the greatest gains in travel speed and fuel cost reduction. At the same time, unstructured hybridization schemes (Mix, MixIntegrate) and the pure genetic algorithm (Parallel) lag behind more organized hybrid architectures, confirming the importance of a systematic approach when designing evolutionary algorithms.

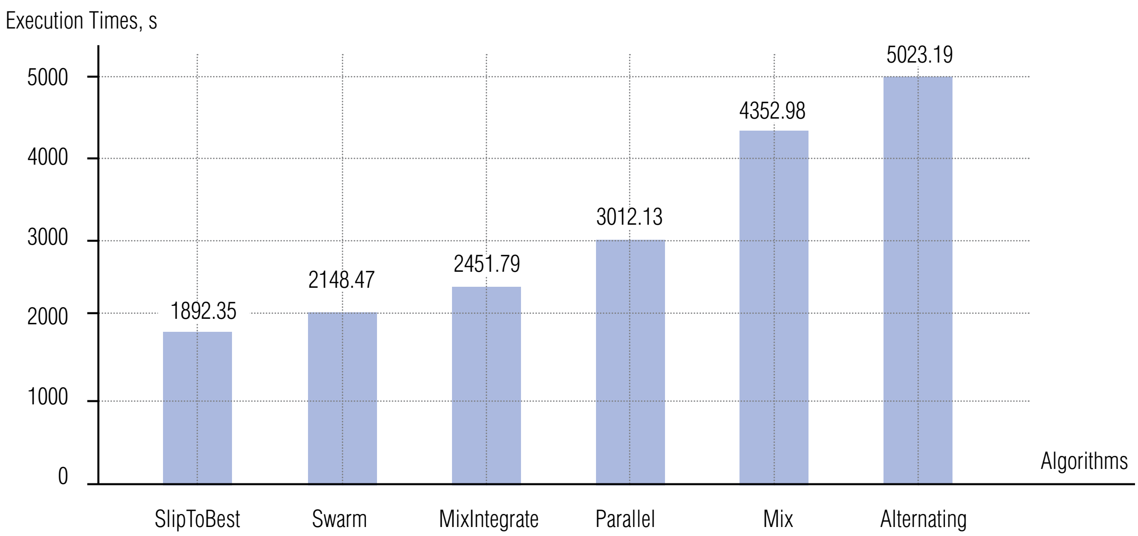
To provide a more detailed comparison, an assessment of the execution time of different algorithms for traffic signal cycle control on the studied road segment was conducted. The results are presented in numerical format, allowing for a comparison of the computational complexity of the methods and their applicability in practical modeling scenarios.

The experiment analyzed six optimization algorithms:

- **SlipToBest** – a modified particle swarm optimization algorithm with the *alpha* parameter.
- **Parallel** – a pure genetic algorithm with parallel data processing.
- **Swarm** – a standard implementation of PSO.
- **MixIntegrate** – a combined method integrating PSO with genetic operators (mutation, crossover).
- **Mix** – a hybrid algorithm combining PSO and GA without a strict operator alternation scheme.
- **Alternating** – a method in which PSO and GA stages are alternated to balance global and local search.

The primary goal of the analysis was to determine the execution time of each algorithm under fixed input parameters in the AnyLogic environment.

Figure 5 presents the experiment results, illustrating the execution time of the algorithms in seconds based on numerical simulations.



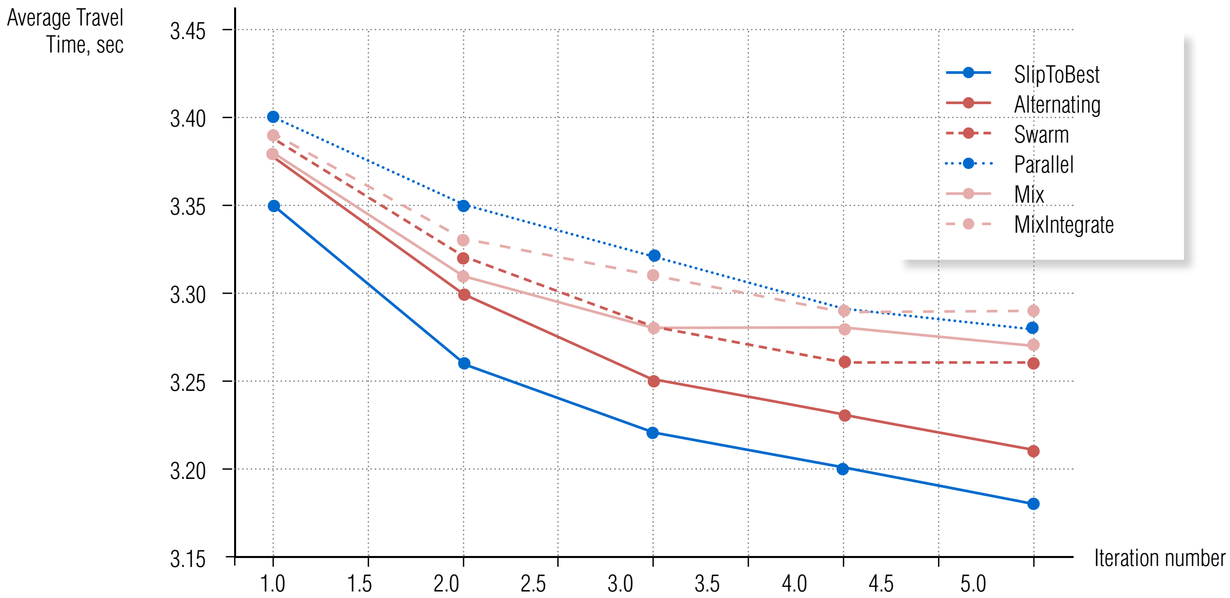


Fig. 6. Convergence dynamics of the average travel time.

The analysis of the data obtained shows that SlipToBest has the shortest execution time – 1992.78 seconds, indicating the high computational efficiency of the algorithm. Optimization of the α parameter allowed for accelerated convergence and reduced overall computation time. Parallel demonstrated 2415.93 seconds due to parallel data processing, significantly speeding up calculations compared to sequential methods. Swarm (classical PSO) recorded 3008.97 seconds, confirming its stable but not the fastest convergence compared to the modified SlipToBest.

The MixIntegrate and Mix algorithms required 4591.87 and 6351.96 seconds, respectively, which can be attributed to their complex hybrid structure and the additional iterations required for mutation and crossover operations. Alternating, which alternates between PSO and GA, exhibited the longest execution time – 6542.57 seconds, due to its high computational complexity, despite the high accuracy of the obtained solutions.

Analyzing the results, SlipToBest stands out as the most efficient in terms of computational speed, making it the optimal choice for real-time traffic flow management. Parallel and Swarm provide balanced results and can be used for practical optimization problems, where a trade-off between computational complexity and solution quality is needed. MixIntegrate and Mix, while delivering more accurate solutions, require computational load optimization to enhance their practical applicability. Alternating, despite producing the most precise solutions, demands significant computational resources, making it more suitable for tasks where accuracy is prioritized over execution speed.

These findings highlight the importance of choosing an appropriate algorithm depending on the modeling scenario. Future research should focus on enhancing hybrid methods to achieve an optimal balance between computational efficiency and solution quality.

During the experiments, each algorithm was tested under different initial conditions, including randomly initialized swarms for PSO and varied initial populations for GA. The results indicate that SlipToBest not only achieves the best average travel time but also exhibits high solution stability, showing minimal variability under different initial parameter settings.

Figure 6 illustrates the convergence dynamics of the average travel time $\overline{T}$. As shown in the graph, SlipToBest reaches the minimum value of $\overline{T}$ faster than other algorithms, thanks to the directed shift of swarm particles toward the globally best solution. Alternating also demonstrates high solution quality; however, in most repeated experiments, SlipToBest shows a faster approach to the optimum.

The remaining algorithms (pure PSO – Swarm, pure GA – Parallel, as well as unstructured hybrids Mix and MixIntegrate) could not outperform SlipToBest in terms of convergence speed or solution stability.

To formally validate these differences, a statistical analysis was conducted, including:

- Comparison of mean values $\overline{T}$.
- Analysis of variance (ANOVA).
- Nonparametric tests with a significance level of $p < 0.05$.

The results of the statistical analysis confirm the superiority of SlipToBest in both average travel time and solution stability.

When developing hybrid evolutionary algorithms, the tuning of key parameters plays a crucial role, including:

- PSO coefficients (w, c_1, c_2) and swarm size.
- The parameter α in SlipToBest, which determines the intensity of the swarm's attraction to the best-found solution.
- GA parameters (*mutationRate, crossoverRate*) and population size in hybrid schemes.

Experimental data show that pure PSO and GA often underperform in certain cases, and improper operator combination (as seen in Mix and MixIntegrate) can lead to inefficient allocation of computational resources. The SlipToBest modification addresses this issue by enhancing the swarm's ability for fast and stable search.

The experimental results lead to the following key conclusions:

- SlipToBest is the optimal choice for optimization in complex dynamic systems, where fast convergence to a high-quality solution is critical.
- Tuning the α parameter in SlipToBest (typically within the range 0.1–0.3) allows for a flexible trade-off between search speed and accuracy.
- With limited computational resources, it is essential to consider that SlipToBest demonstrates one of the best execution speeds (see *Fig. 4*), ensuring minimum travel time compared to Alternating.
- A targeted modification of the swarm algorithm leads to superior performance in the average metric $\overline{T}$ and enhances solution stability.
- Simple hybridization (Mix, MixIntegrate) does not guarantee superiority over pure algorithms, whereas a well-structured alternation strategy (Alternating) or a directed step toward the best solution (SlipToBest) significantly improves method efficiency.

Thus, properly structured hybrid algorithms can outperform pure PSO and GA, providing an optimal balance between convergence speed and solution quality. However, an irrational combination of PSO and GA operators can reduce effectiveness, as observed in the Mix and MixIntegrate cases.

Conclusion

In this study, a new optimization method for intelligent transportation systems (ITS) has been developed and tested, utilizing hybrid evolutionary algorithms aimed at improving traffic signal control efficiency in local urban road network segments. To conduct the experiments, an agent-based model was created in AnyLogic, simulating real traffic conditions. The Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) were used as the baseline methods, forming the foundation for the proposed hybrid schemes.

Based on these methods, six algorithmic frameworks were designed and analyzed: Alternating, SlipToBest, MixIntegrate, Mix, as well as pure PSO (Swarm) and pure GA (Parallel). The primary evaluation metric was average travel time, supplemented by fuel consumption analysis, allowing for the consideration of both economic and environmental factors.

The results of comparative experiments lead to the following conclusions:

1. SlipToBest, a modified PSO with an additional "pulling" mechanism that directs particles toward the best solution, demonstrated the highest performance in both convergence speed and solution quality.

2. Alternating, which systematically alternates GA and PSO, performed slightly worse than SlipToBest but provided more stable solutions due to its balanced use of global and local search.

3. Swarm (baseline PSO) outperformed Parallel (pure GA) but fell behind the leading methods (SlipToBest and Alternating).

4. Unstructured hybrids (MixIntegrate and Mix) failed to outperform the implemented methods, highlighting the importance of a well-designed hybridization strategy.

Thus, a simple combination of GA and PSO without a structured strategy does not provide significant advantages. High performance is achieved through specific modifications, such as the directed particle shift in PSO (SlipToBest) or structured alternation of GA and PSO operators (Alternating).

The theoretical significance of this research lies in the advancement and refinement of hybrid evolutionary algorithms for intelligent transportation systems. The practical value is in the potential real-world application of the proposed solutions for traffic signal control, which can reduce average travel time, lower fuel consumption and reduce harmful emissions.

Research limitations include the local scale of the simulated road network and certain simplifying assumptions regarding driver and pedestrian behavior. Future research will focus on expanding the modeling scope, considering seasonal and weather factors, and integrating adaptive parameter tuning strategies for evolutionary algorithms with deep learning and reinforcement learning techniques.

A promising research direction is the development of hybrid algorithms capable of adapting to dynamically changing traffic conditions. This will further enhance transportation infrastructure and provide long-term socio-economic benefits for urban traffic systems. ■

Acknowledgments

This research was supported by the Russian Science Foundation grant (Project No. 23-11-00080).

References

1. Zaripov E.A., Melnikov A.M., Akopov A.S. (2024) Simulation modeling and optimization of traffic flows in local sections of the street-road network using AnyLogic. *Information Technologies*, vol. 30, no. 4, pp. 183–189. https://doi.org/10.17587/it.30.183-189.

2. Akopov A.S., Zaripov E.A., Melnikov A.M. (2024) Adaptive control of transportation infrastructure in an urban environment using a real-coded genetic algorithm. *Business Informatics*, vol. 18, no. 2, pp. 48–66. https://doi.org/10.17323/2587-814X.2024.2.48.66

3. Eberhart R., Kennedy J. (1995) Particle swarm optimization. Proceedings of the *IEEE International Conference on Neural Networks, Perth, Australia*, pp. 1942–1948.

4. Holland J.H. (1992) *Adaptation in natural and artificial systems*, MIT Press.

5. Clerc M., Kennedy J. (2002) The particle swarm – explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73.

6. Shi Y., Eberhart R. (1998) A modified particle swarm optimizer. Proceedings of the *IEEE International Conference on Evolutionary Computation*, IEEE, pp. 69–73.

7. Bäck T., Fogel D.B., Michalewicz Z. (2000) *Evolutionary computation 1: Basic algorithms and operators*, CRC Press.

8. Wahde M. (2008) *Biologically inspired optimization methods: An introduction*, WIT Press.

9. Coello C.A.C., Lamont G.B., Van Veldhuizen D.A. (2007) *Evolutionary algorithms for solving multi-objective problems*, Springer.

10. Yoshida H., Kawata K., Fukuyama Y., Takayama S., Nakanishi Y. (2000) A particle swarm optimization for reactive power and voltage control considering voltage security assessment // *IEEE Transactions on Power Systems*, vol. 15, no. 4, pp. 1232–1239. https://doi.org/10.1109/59.898095

11. Zhang Y., Wang S., Ji G. (2015) A comprehensive survey on particle swarm optimization algorithm and its applications. *Mathematical Problems in Engineering*, pp. 1–38.

12. Ali M.Z., Pant M. (2009) Improving the performance of particle swarm optimization using differential evolution. Proceedings of the *IEEE Congress on Evolutionary Computation*, pp. 105–112.

13. Poli R., Kennedy J., Blackwell T. (2007) Particle swarm optimization. *Swarm Intelligence*, vol. 1, no. 1, pp. 33–57.

14. Engelbrecht A.P. (2007) *Computational intelligence: An introduction*. 2nd Edition, Wiley.

15. Goldberg D.E. (1989) *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley.

16. Michalewicz Z. (1996) *Genetic algorithms + Data structures = Evolution programs*, Springer.

17. Haupt R.L., Haupt S.E. (2004) *Practical genetic algorithms*, 2nd Edition, Wiley.

18. Tsai P.W., Chiu S.W., Pan J.S., Liao B.Y. (2010) Enhancing particle swarm optimization with multicriterion for engineering optimization. *Expert Systems with Applications*.

19. Ong Y.S., Lim M.H., Zhu N., Wong K.W. (2006) Classification of adaptive memetic algorithms: A comparative study. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*.

20. Zhou A., Qu B.Y., Li H., Zhao S., Suganthan P.N., Zhang Q. (2011) Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation*.

21. Bot D.M., Peeters J., Liesenborgs J., Aerts J. (2023) FLASC: A flare-sensitive clustering algorithm: extending HDBSCAN* for detecting branches in clusters. *arXiv:2311.15887*. https://doi.org/10.48550/arXiv.2311.15887

22. Anfyorov M.A. (2019) Genetic clustering algorithm. *Russian Technological Journal*, vol. 7, no. 6, pp. 134–150 (in Russian). https://doi.org/10.32362/2500-316X-2019-7-6-134-150

23. Beklaryan A.L., Beklaryan L.A., Akopov A.S. (2023) Simulation model of an intelligent transportation system for the 'smart city' with adaptive control of traffic lights based on fuzzy clustering. *Business Informatics*, vol. 17, no. 3, pp. 70–86. https://doi.org/10.17323/2587-814X.2023.3.70.86

About the authors

Evgeniy A. Zaripov

Junior Researcher, Laboratory of Dynamic Economic Models and Optimization, Central Economics and Mathematics Institute, Russian Academy of Sciences, 47, Nakhimovsky Ave., 117418 Moscow, Russia;

PhD Student, Assistant Lecturer, Department of Instrumental and Applied Software, MIREA – Russian Technological University, 78, Vernadsky Ave., 119454 Moscow, Russia;

E-mail: e.a.zaripov@ya.ru

ORCID: 0000-0003-1472-650X

Andranik S. Akopov

Doctor of Technical Sciences, Professor;

Professor of the Russian Academy of Sciences;

Chief Researcher, Laboratory of Dynamic Economic Models and Optimization, Central Economics and Mathematics Institute, Russian Academy of Sciences, 47, Nakhimovsky Ave., 117418 Moscow, Russia;

Professor, Department of Instrumental and Applied Software, MIREA – Russian Technological University, Russia, 78, Vernadsky Ave., 119454 Moscow, Russia;

E-mail: akopovas@umail.ru

ORCID: 0000-0003-0627-3037

DOI: 10.17323/2587-814X.2025.1.50.71

Development of a fuzzy optimization model for the formation of a portfolio of well-being program activities to increase employee productivity

Lev S. Mazelis [a]
E-mail: lev.mazelis@vvsu.ru

Gleb V. Grenkin [a]
E-mail: gleb.grenkin@vvsu.ru

Kirill I. Lavrenyuk [b]
E-mail: kir.lavrenyuk@mail.ru

Andrey A. Krasko [a]
E-mail: andrey.krasko@vvsu.ru

[a] Vladivostok State University, Vladivostok, Russia
[b] Yandex, Moscow, Russia

Abstract

This study was conducted within the framework of the urgent task of studying the processes of developing the human capital of an organization and increasing employee productivity. At the same time, the development process is viewed through the prism of creating and implementing various elements of the well-being program into the main corporate business processes of the organization. The purpose of this work is to develop a fuzzy method for forming an optimal portfolio of well-being program activities which will allow you to get as close as possible to the target values of key performance indicators (KPIs) of employees on a given planning horizon. To achieve this goal, a hypothesis is put forward about the possibility of building a tool that allows, based on the functional dependencies of influence channels, to form an optimal portfolio of well-being program activities that increases the efficiency of the organization. The method developed consists of a model representing a fuzzy programming problem and a method for finding its solution. A distinctive feature of the model is the consideration of two levels of uncertainty in the formation of an optimal portfolio of activities related to the reliability of estimates of numerical coefficients of functional dependencies of channels of influence and a set of parameters of constraints determined by experts. An integral indicator is used as the target function of the model, which characterizes the degree to which the target values of key employee performance indicators are achieved, taking into account the importance of each of them for the organization. The optimization variables in the model are binary variables that determine the inclusion of a certain event in the well-being program of an organization at a specific time within a given planning period. The limitations in the model are: the total amount of financial resources allocated for the implementation of the well-being program; the amount of investment in a specific area of the well-being program; an increase in the integral indicator of competence of each employee. From a practical point of view, the proposed method will make it possible to form a well-founded portfolio of well-being program activities, the implementation of which has the maximum possible positive impact on employee productivity.

Keywords: well-being program, employee burnout, competence development, portfolio optimization, fuzzy approach

Citation: Mazelis L.S., Grenkin G.V., Lavrenyuk K.I., Krasko A.A. (2025) Development of a fuzzy optimization model for the formation of a portfolio of well-being program activities to increase employee productivity. *Business Informatics*, vol. 19, no. 1, pp. 50–71. DOI: 10.17323/2587-814X.2025.1.50.71

Introduction

In today's world, where the business environment grows more dynamic and competitive, organizations constantly seek ways to improve employee performance and efficiency. An effective approach that has captured the attention of researchers and practitioners alike is implementing comprehensive measures to improve employee well-being in business processes.

Well-being programs are popular because they help create a healthy and productive work environment, improving the well-being of employees and increasing their productivity. Research suggests that well-being programs have a positive impact on various aspects of employees' lives, including physical health, emotional state, social connections and professional development. Companies implementing well-being programs can expect improved quality of their employees' work, lower stress levels and increased satisfaction. Organizations can achieve greater stability of their human resources and an improved emotional microclimate, boosting their competitiveness in the labor market.

Despite the growing popularity of well-being programs, the question stands as to which interventions and approaches are most effective. Studies propose different approaches to the design of well-being programs which differ in their focus, content and implementation methods. Furthermore, implementing similar well-being programs in different organizations often yields significantly different results. This testifies to the fact that the program's effects are also affected by the organization's internal and external factors, marked by high uncertainty and difficulty in forecasting. This creates the need to conduct more in-depth research focused on developing tools to create optimal well-being programs to raise employee performance.

Thus, this study is relevant due to the fact that there is a need to create tools for the formation of an optimal set of well-being program activities to maximize employee productivity. At the same time, when developing this tool, it is important to take into account the uncertainties and risks of the internal and external environment of the organization, as well as the resource constraints in it.

The subject under study is the development of employees' human capital and the reduction of their burnout through a well-being program, and the object is fuzzy tools for creating the optimal portfolio of well-being program measures to maximize employees' KPIs.

1. Literature review

We conducted a comprehensive analysis of studies that address the stages of employee development through the corporate well-being program described in the conceptual model proposed by Mazelis et al. [1]. These studies can be logically organized into several larger groups.

1.1. Research analyzing the influence of well-being programs on the development of organization employees

In [2] the authors proved that a culture rooted in respect, inclusivity, fairness and teamwork significantly increases employees' proactiveness and contributes to innovations and the organization's long-term development.

The authors [3] discovered a connection between employees' well-being and the accumulation of professional knowledge and skills in the organization.

The authors [4] showed the positive mutual influence of some personnel management practices (e.g., L&D, employee engagement management) and the development of employees' competencies.

In article [5] the authors described the effects of continuous training of managers in service enterprises and their impact on their professional competencies. The study [6] shows that the rapid spread of vocational education instead of academic education negatively affects the development of human capital.

Despite the scarcity of studies assessing the influence of corporate well-being programs on the development of employees' competencies, all of them confirm the positive effects of such programs. However, available research considers either the influence of individual well-being elements on employee competence or the program's influence on individual competencies. Thus, it is desirable to model the influence of a comprehensive well-being program on the overarching development of employee competence.

1.2. Research analyzing the influence of well-being programs on burnout

The authors of [7] proved that a supportive work culture contributes to improved well-being, mental health, motivation and satisfaction from work.

Drawing on an analysis of data from 240 interviews with medical workers conducted in SPSS, the authors [8] reported a positive relationship between employee well-being and perceived organizational support, a negative relationship between perceived organizational support and emotional burnout, and a negative or opposite relationship between employee well-being and emotional burnout.

In [9] conclusions were made that corporate social support programs (as part of well-being programs) promote employee satisfaction, engagement, loyalty, and productivity and better management team performance in the face of uncertainty.

In [10] the authors tested hypotheses about the impact of combined measures in the workplace on employee burnout levels. However, they noted that the data used in the study are limited due to high heterogeneity, potential bias and a small sample. This warrants further exploration of these issues.

In [11] the authors described key trends in implementing the concept of well-being by organizations to incentivize employees' work. In [12] the impact of the implementation of corporate health and well-being management programs on the general condition of employees and their productivity was proven.

The studies cited prove individual elements of corporate well-being programs and their synergistic influence affect employee burnout levels. The nature of this phenomenon is still not described exhaustively. In this study, we tested the hypothesis that the level of employee burnout as an integral indicator based on the assessment of loyalty, satisfaction and engagement affects the deviation of employees' expectations regarding specific well-being program measures from reality. In turn, employees' expectations depend on their value orientations.

1.3. Research analyzing the influence of developing employee competence and their burnout on the organization's performance

The author [13] proposed an approach to improving employee satisfaction based on building a work-life balance and demonstrated its contribution to organizations' sustainable development.

In [14] the authors substantiated the influence of workplace socio-psychological background on employees' productivity through changes in their burnout levels. In [15] they demonstrated that burnout among IT specialists affects the speed and quality of their performance.

The results of [16] show that the work environment, support from higher-ups, adaptability and intrinsic motivation have tremendous impact (direct and indirect) on employee productivity. In [17] the authors described the impact of some HR management practices on the work of the entire organization through their impact on employee competencies.

The results of [18] show that employee competence and job characteristics significantly affect their motivation for work and effectiveness. The authors [19] demonstrated that developing employees' competencies has a statistically significant influence on organizations' sustainability. The works [20, 21] described the impact of various aspects of labor (e.g., professional competence, work environment, satisfaction with work, remuneration, etc.) on employee productivity.

In [22] reported on the results of research into the influence of organizational culture and employee competence on perceived stress and productivity.

In [23] the components of burnout and their relationship with employee performance indicators were studied. Using a correlation analysis of the sample, it was estimated how various factors within the company affect employee burnout, as well as how certain aspects of burnout affect their productivity.

We can conclude that the effects of employee competence and burnout as distinct phenomena on worker productivity and organizations have been thoroughly investigated. However, the combined influence of these factors is described poorly. The analyzed studies assess the effects on the integral indicator of productivity or efficiency rather than its aspects, such as KPI.

1.4. Research assessing relationships between well-being program measures, employee competence, burnout, and the organization's performance indicators calculated mathematically

The authors of [24] analyzed data from a survey of 403 employees using Path Analysis' SmartPLS functionality and refuted the correlation between employees' corporate social responsibility and well-being.

In [25] the authors developed a quantitative model for assessing the impact of the quality of workplace equipment and engagement on employee well-being in the hospitality industry.

In [26] a model is proposed that allows us to identify the relationship between elements of motivation and employee performance. The model revealed a stable and statistically significant positive relationship between the parameters.

In [27] confirmed through regression analysis that practices improving the psychological climate in the team and developing leadership qualities have a positive influence on an organization's performance.

In [28], the results obtained using the Smart PLS method indicate that a favorable working atmosphere and engagement act as resources that help prevent burnout. Furthermore, senior management should be cautious about increasing working hours.

It was demonstrated in [29] that a low level of employee engagement does not always lead to burnout, and at the same time, the more employees work, the more they feel burnout. The study used digital doubles of employees and applied the LISREL package to analyze apparent and hidden variables and quantitatively assess relationships.

Note that existing studies have not previously solved the problem of finding the optimal portfolio of well-being program activities in order to maximize the effectiveness of the organization's activities. Most of the work uses standard data analysis tools, on the basis of which conclusions are drawn about the presence or absence of relationships between various system parameters.

At the same time, it is worth noting that the instrumental part of portfolio optimization is quite developed. For example, the fundamental principles and approaches in the field of mathematical programming were laid down in [30]. There are also other approaches that are often used to form optimal portfolios and sets of projects in industry and the financial sector, for example, quadratic programming [31] and genetic optimization algorithms [32]. Previously, the authors have already worked in the field of portfolio optimization, for example, in [33] models of fuzzy multi-period optimization were developed to support decision-making when choosing a portfolio of projects within the framework of the institution's strategic development program, allowing for step-by-step planning of a portfolio of projects taking into account the interests and risks of stakeholders.

Additionally, it is worth noting that due to the presence of uncertainties associated with data collection based on a subjective assessment of an employee or his supervisor, the lack of unified approaches to assessing burnout, competence and other entities, there is a need to use a fuzzy approach to model the existing uncertainties of the internal environment and minimize possible risks.

* * *

In conclusion, we can draw attention to the following drawbacks of existing research, which our study aimed to mitigate:

- the effects stemming from the multiperiodicity of the implementation of corporate well-being program measures and the possibility of rolling planning are not considered;
- the optimization aspects of developing the portfolio of measures under the corporate well-being program

given the effects of its implementation are not addressed;

- uncertainties and risks generated by the collection of subjective raw data, a vital component in building mathematical models, are unaccounted for;
- the effects of the comprehensive influence of the corporate well-being program on employee burnout and competence are understudied;
- the comprehensive influence of employee competence and burnout on performance and efficiency is not fully explored.

Thus, we can conclude on the lack of a proper toolkit, which would allow:

first, to describe the impact of the well-being program on the development of professional and personal competencies of employees, their burnout and performance, and, consequently, on the effectiveness of the organization as a whole;

second, with limited resources, high competition for personnel, risks and uncertainties, to form an optimal set of well-being program activities that will contribute to maximum approximation to the target values of performance indicators for both an individual employee and the organization as a whole.

2. Research goals and objectives

This study aimed to develop a fuzzy method for creating the optimal complex of measures under the corporate well-being program to maximize progress toward employees' target KPIs over a given planning horizon.

The hypothesis is that it is possible to develop a method to create the optimal portfolio of well-being program measures to improve an organization's performance based on competency development and reduction of burnout balanced across staff and time slots.

To achieve the goal, we addressed the following objectives:

1. To develop a fuzzy dynamic model for creating the optimal portfolio of well-being program measures accounting for the organization's resource limitations. The implementation of program measures depends on their direction and affects the development of workers' professional and personal competencies, which makes it possible to maximize progress toward the target KPIs of employees and the goals of the organization.

2. To develop a method for finding the solution of the fuzzy optimization model that would allow us to determine the set of program measures for each time interval on the planning horizon and each employee considering the organization's resource constraints.

3. Fuzzy optimization model for creating the optimal portfolio of measures under the corporate well-being program

Let us consider the work of organization employees on the planning horizon T. At each time point $t \in \{0, 1, ..., T\}$, the company invests in various measures within its well-being program, which affect the development of employees' personal and professional competencies and reduce their burnout at time point $(t + 1)$. Employees, considering their competence, influence the achievement of their and the organization's KPIs at time point $(t + 1)$. A critical aspect is that in commercial companies, the achievement of goals at time point t directly affects the volume of investment in the well-being program at time point $(t + 1)$.

The same employees with the same level of competence and preservation of business processes at different time intervals can achieve different KPIs in the organization. This relates to the fact that workers' productivity and efficiency are affected by their burnout level. Under employee burnout level, we understand their combined physical and emotional state, which affects how quickly they perform their work and how often they make mistakes in the business processes for which they are responsible. Mazelis et al. [34] demonstrated that burnout is a consequence of the deviation of employees' expectations about well-being program measures in the organization from what the company offers.

Mazelis et al. [1] proposed a conceptual model of employee competence development and burnout reduction by implementing the well-being program. Under conditions of uncertainties, risks, and limited resources, the management faces the task of optimally allocating financial resources between measures within the corporate well-being program to improve the performance of employees and the company through direct or indirect impact on employee competence development and the reduction of burnout (*Fig. 1*). The indicators of productivity used in this study are employees' KPIs.

Mazelis et al. [34–36] built the functional relationships of the influence channels provided in the conceptual model (*Fig. 1*).

Let us consider the work of organization employees on the planning horizon T. At each time point $t = 0, 1, ..., T$, the company invests resources in measures under the well-being program. Each measure can apply to different employees. Introducing these measures affects employee competence and changes in the deviation of employees' expectations regarding the well-being program elements from their implementation, which in dynamics leads to changes in their burnout level.

Each measure under the well-being program that requires financial investment relates to a particular direction (the program is divided into K directions) and focuses on a set of employees. The volume of investments attributable to the i-th employee in the

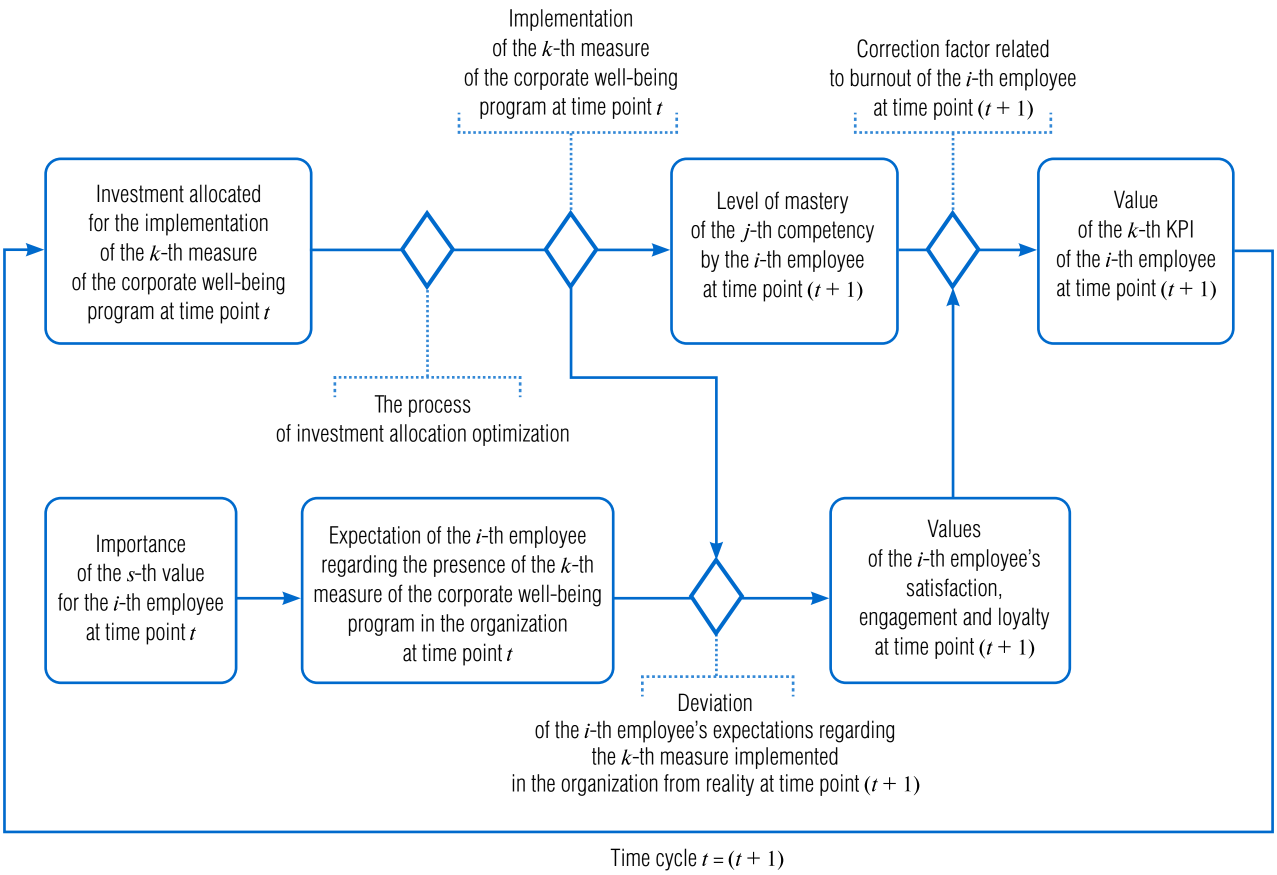


Fig. 1. Conceptual model for describing the process of developing the competence of organization employees.

implementation of all program measures belonging to the k-th direction of well-being shall be denoted by $z_{ik}(t)$.

The goal of optimization as part of our study is to maximize the integral indicator V, which describes the attainment of employees' target KPIs:

$$V = \frac{1}{I}\sum_{i=1}^{I}\left(\sum_{m=1}^{M}\beta_m \cdot \frac{y_{im}(T)}{\tilde{y}_{im}}\right) \to \max, \qquad (1)$$

where

$y_{im}(t)$ is the value of the m-th KPI of the i-th employee at time point t;

β_m is the importance coefficient of the m-th KPI;

$\tilde{y}_{im}$ is the target value of the m-th KPI for the i-th employee;

i is the number of the company employee, $i = 1, 2, ..., I$;

m is the number of the employee's KPI, $m = 1, 2, ..., M$.

Let us consider the constructed functional descriptions of the influence channels given in the conceptual model (*Fig. 1*).

1. Mazelis et al. [36] provided the econometric models of panel data, which functionally describe dependencies between the structure and volume of financial investments into measures under different directions of the well-being program and employee's levels of competencies:

$$x_{ij}(t+1) = \gamma_j x_{ij}(t) + \sum_{k=1}^{K}\alpha_{jk} z_{ik}(t), \qquad (2)$$

where

x_{ij} is the level of development of the j-th competency in the i-th employee at time point t;

$z_{ik}(t)$ is the volume of investments in the i-*th* employee of the company in the k-*th* direction of well-being at time point t.

To eliminate multicollinearity and reduce dimensionality while preserving information as much as possible, we used principal component analysis, in which a rotation-free method was employed to overcome difficulties in interpreting the loading matrix.

2. The channel of influence of investments in the implementation of the program measures on changes in the deviations of employees' expectations from the actual state of the corporate program is described as follows:

$$q_{ik}(t+1) = \max\left\{\min\{q_{ik}(t) + 2\frac{z_{ik}(t)-\mu_k}{v_k-\mu_k}, 1\}, -1\right\}, \qquad (3)$$

where

$q_{ik}(t)$ is the deviation of the i-th employee's expectations regarding the k-th direction of well-being from reality, $q_{ik}(t) \in [-1,1]$;

μ_k is the minimal investment in the k-th direction per employee per quarter beginning from which expectations start to change;

v_k is the maximum investment in the k-th direction per employee per quarter after which changes no longer occur.

3. Mazelis et al. [34] proposed a fuzzy model of the dependence of employee burnout on the discrepancy between their expectations about the company's corporate environment and the actual content of the well-being program, which includes several stages:

i) forming an integral indicator of expectations E_i^{INT}, equal to the weighted sum of $a_{ik}q_{ik}$, where a_{ik} (from 0 to 1) is the importance coefficient of the k-th direction of well-being measures for the i-th employee and q_{ik} is the degree of the employee's satisfaction with the implementation of the k-th measure (from –1 to 1);

ii) presenting the areas of the set of values of burnout indicators and the integral expectation indicator as a union of non-intersecting intervals of different lengths, each of which is considered as some category of the respective indicator; optimal weighting factors and fuzzy category boundaries are found by minimizing the partitioning quality functional

$$J = \sum_{i=1}^{I}\sum_{s=1}^{S}u_{is}d_s^2\left(E_i^{INT}\right), \qquad (4)$$

where

d_s is the distance from a point on the axis of the integral expectation index to the s-th interval;

u_{is} is the degree of the i-th point's membership in the s-th interval;

iii) the solution to the optimization problem is found iteratively: a) finding the optimal partitioning of points into classes by minimizing the functional at fixed weight coefficients w_k, b) finding weighting coefficients at the given division of points into classes by solving the problem of unconditional optimization of the functional

$$\frac{J}{\|w\|^2}, \text{ where } w = (w_1, w_2, ..., w_k);$$

iv) building a matrix of correspondence between fuzzy categories of the integral indicator of expectations and burnout levels.

After the defuzzification of the fuzzy piecewise constant regression based on the correspondence matrix, we obtain the dependence of the burnout index of the i-th employee on the integral expectation indicator:

$$b_i(t) = \psi\,(a_{i1}q_{i1}(t), \ldots, a_{iK}q_{iK}(t)). \qquad (5)$$

The graph of this dependence is presented in *Fig. 2*.

4. Mazelis et al. [35] provided a fuzzy model that allows assessing the influence of employee competencies and burnout levels on the achievement of KPIs. The model includes several stages:

i) forming an integral indicator of employee competence as a weighted sum of individual professional and personal competencies with optimal values of weighting coefficients;

ii) building fuzzy categories for the integral competence indicator and each KPI:

- representing the area of values of each indicator as a combination of non-overlapping intervals of different lengths, each of which is considered a category of employee competence (KPI);
- the categories are constructed based on the minimization of the cross-entropy function, and each element of the sample is described by a fuzzy number characterizing the degree of the element's membership in the constructed categories;

iii) adjusting KPI values for burnout by calculating the reduced KPI based on the solution of the de-

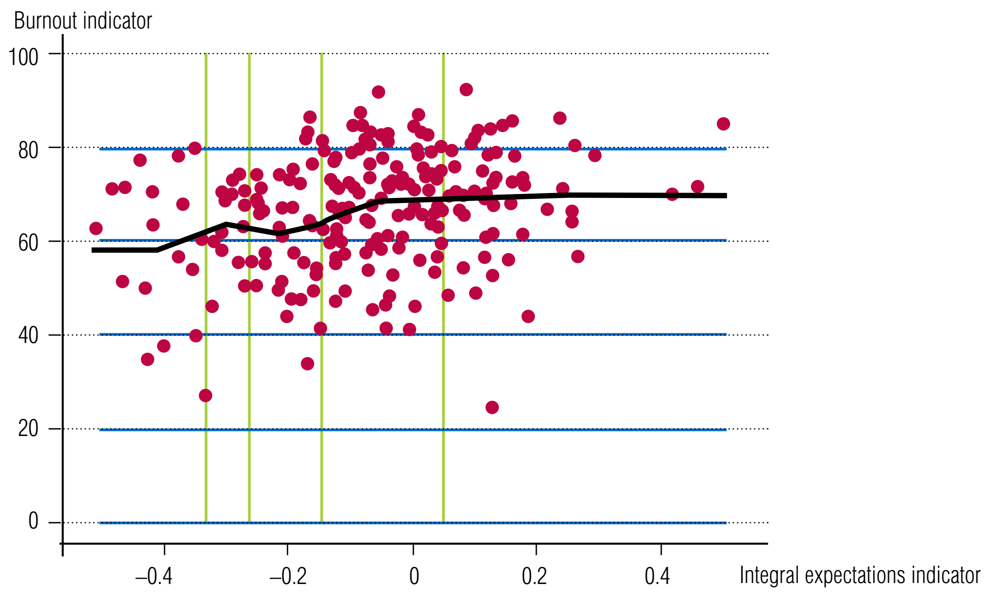


Fig. 2. Dependence of the burnout indicator of the *i*-th employee on the integral expectations indicator.

viation minimization problem using the weighted least squares method, considering the fuzzy membership of points to categories;

iv) building a sample distribution of KPIs for each category of the integral competence indicator based on the constructed correspondence matrix, which determines the distribution of points belonging to a competence category. This will allow us to estimate the probability of possible KPI values for specific values of competence and burnout.

Defuzzifying the dependence found; we obtain:

$$\frac{y_{im}(t)}{\tilde{y}_{im}} = \varphi_m\left(x_{i1}(t),\ldots,x_{iJ}(t)\right) + W_{m0} + W_{m1}b_i(t), \qquad (6)$$

where x_{ij} is the level of development of the j-th competency of the i-th employee at time point t;

$b_i(t)$ – the integral burnout indicator of the i-th employee.

The dependence of the i-th employee's KPI on the integral competence indicator is presented in *Fig. 3*.

The limits used in the optimization model are as follows.

1. The total amount of financial resources invested into the well-being program is limited by the company's budget and is defined by its C-level as part of human resources management strategy:

$$\sum_{k=1}^{K}\sum_{i=1}^{I} z_{ik}(t) \le Z(t). \qquad (7)$$

The budget is determined by top management as part of forming the human resources plan depending on certain cycles in the company.

2. Let us assume that investments in each well-being direction are formed as a sum of measures in which an employee participates. Each measure is characterized by its cost per person. Therefore, $z_{ik}(t)$ equals the total cost of the selected activities related to the k-th direction. There is a maximum allowable amount of investment that a company can allocate in a well-being direction:

$$\sum_{i=1}^{I} z_{ik}(t) \le Z_k(t), \qquad (8)$$

where $Z_k(t)$ is the limit on investments in the k-th direction of well-being, which is a trapezoidal fuzzy number expertly set by the company's managers.

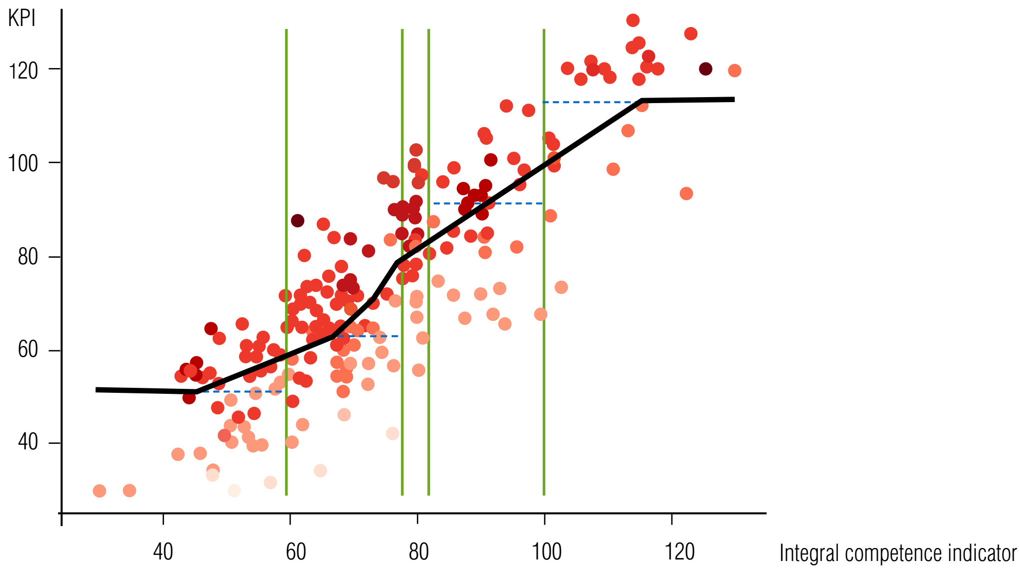


Fig. 3. Dependence of the KPI of the *i*-th employee on the integral competence indicator.

This limit is intended to ensure that those responsible for well-being do not direct all investments to the simplest and most costly activities, but try to address as many aspects of well-being as possible.

3. In one unit of time, the increase in the integral competence indicator due to investment in the corporate program does not exceed the fuzzy value c:

$$\sum_{k=1}^{K}\left(\sum_{j=1}^{J} w_j \alpha_{jk}\right) z_{ik}(t) \le c. \tag{9}$$

Thus, the optimization model of creating the corporate program influencing employees' KPIs through developing their competencies considering burnout takes the following form:

$$\begin{cases} f(T)=\dfrac{1}{I}\sum_{i=1}^{I}\left[\begin{array}{l}\varphi\left(x_{i1}(T),\dots,x_{iJ}(T)\right)+W_0+\\ +W_1\psi\left(a_{i1}q_{i1}(T),\dots,a_{iK}q_{iK}(T)\right)\end{array}\right]\to\max \\ q_{ik}(t+1)=\max\left\{\min\left\{q_{ik}(t)+2\dfrac{z_{ik}(t)-\mu_k}{v_k-\mu_k},1\right\},-1\right\}, \\ t=0,\dots,T-1 \\ x_{ij}(t+1)=\gamma_j x_{ij}(t)+\sum_{k=1}^{K}\alpha_{jk} z_{ik}(t), t=0,\dots,T-1 \\ \sum_{k=1}^{K}\sum_{i=1}^{I} z_{ik}(t)\le Z(t), t=0,\dots,T-1 \\ \sum_{i=1}^{I} z_{ik}(t)\le Z_k(t), t=0,\dots,T-1, k=1,\dots,K \\ \sum_{k=1}^{K}\left(\sum_{j=1}^{J} w_j\alpha_{jk}\right) z_{ik}(t)\le c,\ k=1,\dots,K,\ i=1,\dots,I \end{cases} \tag{10}$$

The limits on investment in the k-th direction of well-being and the competence increment of an individual employee are fuzzy inequalities.

The proposed model constitutes a fuzzy programming problem because some parameters are fuzzy numbers. To solve this problem of finding the optimal portfolio of program activities, we applied the approach described by Anshin et al. [37] and Wang and Hwang [38] to reduce it to a non-fuzzy problem at given confidence levels.

Let us specify the confidence levels λ_{Zk}, λ_c for limits on investment in the k-th direction of well-being and the growth of the integral employee competence indicator, respectively. In this case, the system of model limits will take the following form:

$$\begin{cases} N_{\sum_{i=1}^{I} z_{ik}(t)}\left(Z_k(t)\right)\ge\lambda_{Zk} \\ N_{\sum_{k=1}^{K}\left(\sum_{j=1}^{J} w_j\alpha_{jk}\right) z_{ik}(t)}(c)\ge\lambda_c, \end{cases} \tag{11}$$

where $N_A(B) > \lambda$ means that A satisfies limit B with a confidence level λ. This condition is equivalent to the inequality:

$$\min_x \max\left(1-\mu_A(x),\mu_B(x)\right)>\lambda, \tag{12}$$

where $\mu_Y(x)$ is the membership function of the fuzzy number Y.

Let the volume of investment in the k-th direction of well-being program measures at time point t be specified by the trapezoidal number $\sum_{i=1}^{I} z_{ik}(t)=(z_1,z_2,z_3,z_4)$ and the limit on the maximum amount of investment in the k-th direction of the well-being program – by the fuzzy trapezoidal number $Z_k(t) = (0, 0, r_3, r_4)$. Then the limit $N_{\sum_{i=1}^{I} z_{ik}(t)}\left(Z_k(t)\right)\ge\lambda_{Zk}$ is equivalent to:

$$\left[\left(1-\lambda_{Zk}\right)\cdot z_3+\lambda_{Zk}\cdot z_4\right]\le\lambda_{Zk}\cdot r_3+\left(1-\lambda_{Zk}\right)\cdot r_4. \tag{13}$$

Similarly, fuzzy inequalities are transformed into non-fuzzy ones, allowing one to obtain a non-fuzzy problem of Boolean programming, where the variables define whether the measure is included in the program.

To find the suboptimal solution, we performed the linearization of the dependencies:

1. To linearize the function φ in point x, we need to find the slope coefficient of the dependence. At the boundary between classes, the value of the slope coefficient is calculated as the angular coefficient of the corresponding segment of the piecewise linear function. At the remaining points x, the slope coefficient is linearly interpolated.

2. In linearizing the dependence of $q_{ik}(t+1)$ on $z_{ik}(t)$, we should note that the number of possible values of $z_{ik}(t)$ is small (averaging at about 25 based on real data). Therefore, if we introduce binary variables corresponding to each i, each k, and each of the possible values of $z_{ik}(t)$ and calculate how much the degree of KPI achievement changes with this investment, we obtain an integer linear programming problem. In this case, it will be necessary to limit the variables attributable to a specific i and a specific k: their sum must equal 1. The set of possible values of the variables $z_{ik}(t)$ is found through dynamic programming.

In solving the integer linear programming problem, we used the *scipy.optimize.milp* package in Python. Software implementation available in the repository https://github.com/lapkin25/fuzzy-data-clustering in the folder "fuzzy_optimization."

4. Results of approbation of the method based on the example of subdivisions of a commercial company

Let us consider creating the optimal portfolio of measures under the well-being program to achieve maximum progress toward employees' target KPIs over one year divided into four periods (quarters).

We examined the IT and HR structural subdivisions of a commercial organization with a total headcount of 95 employees. The raw data collection for this example is described in previous papers by Mazelis et al. [1, 34, 36]. At the starting point in time $t = 0$ the average values of staff competence in terms of personal qualities is the fuzzy number {61; 68; 84; 91}, in terms of professional qualities in the context of HR functions – {49; 55; 67; 73}, and in the context of IT functions – {38; 43; 53; 58}. The values of indicators characterizing burnout level are as follows: loyalty $b_1(0)$ = {46; 52; 64; 70}; engagement $b_2(0)$ = {57; 64; 78; 85}; satisfaction $b_3(0)$ = {57; 64; 78; 85}.

KPIs and their target values within the loss function are chosen in accordance with the company's strategic goals. The initial value of the integral indicator of employee KPI is 71.4%.

Each measure has a set of characteristics used to decide whether it should be included in the well-being program: minimum and maximum costs; minimum and maximum number of employees that can participate in it simultaneously; and the intended duration of the measure. An example of measures is given in *Table 1.*

The model allows us to form the optimal portfolio of measures of the corporate well-being program and the corresponding structure of financial resources distribution across program investment directions, employees and time points after finding the numerical solution. The model is updated for each upcoming quarter based on the results achieved in the current one.

Let us consider the results of the model solution under the condition of limiting the volume of investments in the program to 12.5 million rubles/quarter. Importantly, the results depend on the confidence level (λ), which is assigned by the company's management and represents some measure of rigidity when specifying parameters in a fuzzy form, i.e., the closer the value is to 1, the more rigid the parameter (0.5 corresponds to the default value). *Table 2* shows examples of confidence levels depending on the scenario.

Scenario 1 describes a situation where the organization strives to allocate a more rigidly defined amount of financial resources to develop its employee competence to provide for their career growth in the organization (this structure characterizes an HR strategy focused on nurturing internal human capital). This scenario encompasses such program directions as "Environment," "Career development" and "Development of skills." Investments in other blocks of directions have a gap and are mainly driven by the features of the team working in the company.

Scenario 2 arises if the organization deems it important to create a positive and transparent corporate environment (i.e., the HR strategy focuses on creating a positive employer brand and attracting experts from the external market). This scenario covers such blocks of program directions as "Financial well-being," "Healthy lifestyle" and "Corporate infrastructure."

Table 1.

Examples of well-being program measures

Direction of well-being	Measure	Mode of costs, thousand rubles	Number of employees, people	Time period
Implementing corporate benefits	Voluntary medical insurance	80	1	12
	Corporate discounts from partners	6	1	1
	Contributions to a non-profit pension fund	10	1	1
	Additional vacation days	25	1	12
	Corporate bonus program	10	1	1
Integrating employees into industry processes	Participation of speakers in industry conferences	60	1	1
	Publication of original articles	40	1	1
	Organization of industry communities	1300	100	3
	Organization of team days	7000	500	6
…	…	…	…	…

Scenario 3 creates a situation where the organization prioritizes neither of the directions. This scenario is appropriate if the company wishes to create a mixed HR strategy and identify the most effective directions depending on the team's features.

Next, let us consider the structure of financial resource allocation for the year according to the three scenarios (*Fig. 4*).

Analyzing *Fig. 4*, we should highlight the following:

1) in scenario 1, the priority directions of the well-being program receiving about 43% of the investments are "External development of SOFT competencies" (10.7%), "Implementing a talent management system" (10.4%), "Integrating employees into industry processes" (10%), "Creating adaptive work processes" (7.2%) and "Organizing healthy eating" (5%). The largest amount of investment is allocated to the areas that are part of the priority blocks of the program and provide the greatest impact on employee competence development;

2) in scenario 2, the priority directions of the well-being program receiving around 42% of the investments are "Creating adaptive work processes" (9.4%), "Integrating employees into industry processes" (8.7%), "External development of SOFT competencies" (8.3%), "Implementing a talent management sys-

Table 2.

Confidence levels

Parameter	Scenario 1	Scenario 2	Scenario 3
Integral competence indicator	0.6	0.6	0.6
Limitation on the volume of investments into the “Financial well-being” well-being block	0.6	0.9	0.7
Limitation on the volume of investments into the “Environment” well-being block	0.9	0.6	0.7
Limitation on the volume of investments into the “Career development” well-being block	0.9	0.6	0.7
Limitation on the volume of investments into the “Healthy lifestyle” well-being block	0.6	0.9	0.7
Limitation on the volume of investments into the “Development of skills” well-being block	0.9	0.6	0.7
Limitation on the volume of investments into the “Corporate infrastructure” well-being block	0.6	0.9	0.7

tem” (8.1%) and “Organizing healthy eating” (7.4%). This set of priority directions is similar to scenario 1, yet the investment structure is significantly different. Specifically, more financial resources are invested in program directions that fall within the priority clusters of the strategy;

3) scenario 3 shows a similar trend with respect to priority directions, but the structure of financial resource distribution is again different.

Across all scenarios, two directions (“External development of HARD competencies” and “Meaning management”) receive virtually no funding. This can be explained by the low effect of the measures within these program directions on competencies or burnout reduction.

λ has a significant influence on creating the portfolio of well-being program measures and the structure of funding for program directions.

Our research shows that the dynamics of change in the volume of financial resources depending on λ is monotonous (increasing or decreasing) for most program directions. To be specific, the amount of allocated financial resources increases with λ for the directions of “Implementing corporate benefits,” “Integrating employees into the company’s corporate life,” “Implementing a talent management system,” “Internal development of HARD competencies,” etc. An opposite dynamic is observed for “Integrating employees into industry processes” and “Developing technological and team leadership.”

Analyzing the results by staff, we observed that the largest amount of funding is allocated to employees who fit the following portrait: “Employee between the ages of 31–40 with 3–6 years of total experience. Holds a position above the level of a senior specialist. Works in IT. Shows low levels of several personal competencies and high levels of professional competencies. Burnout is below the company average.”

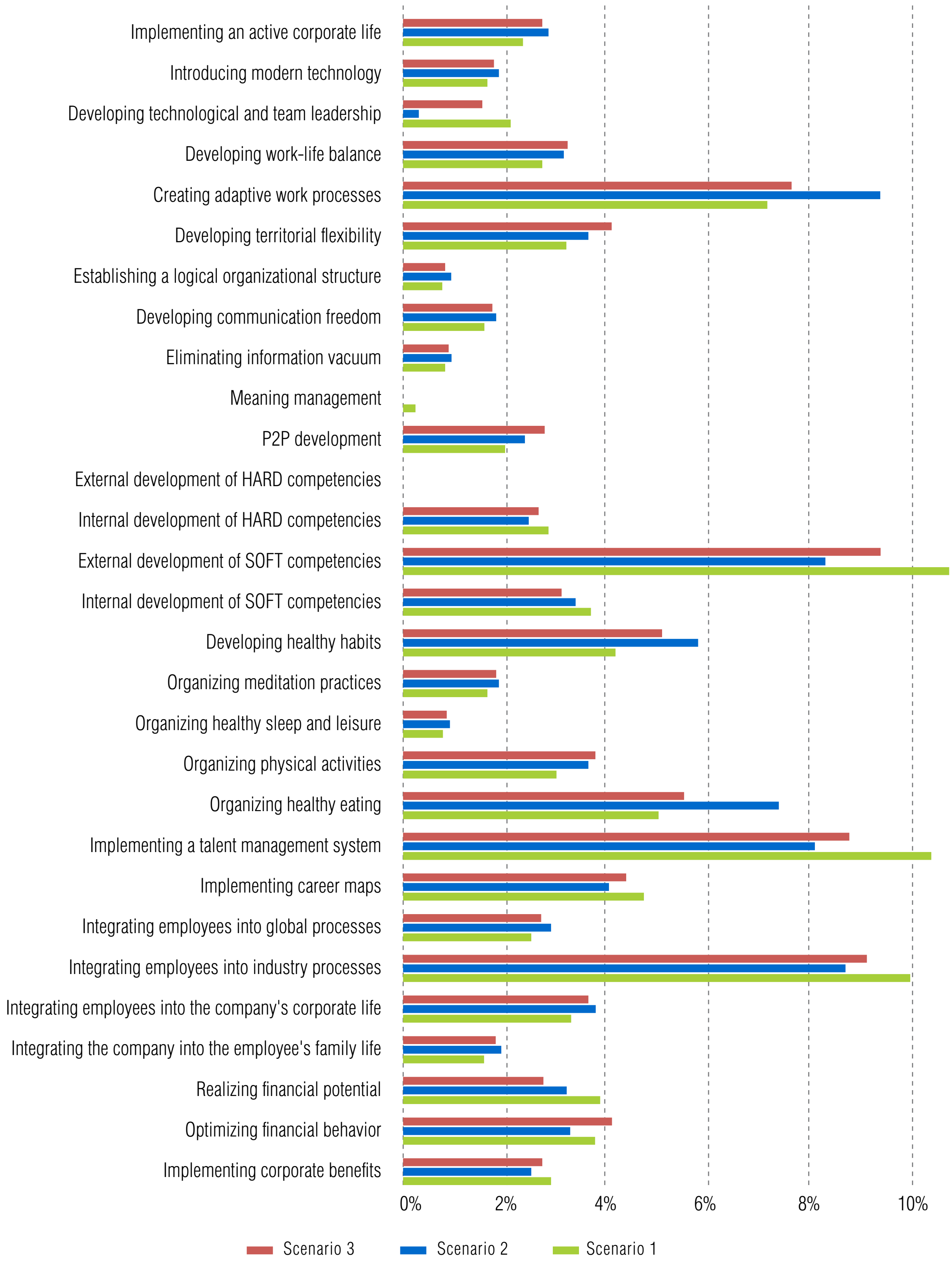


Fig. 4. The structure of financial resources allocation to the directions of the well-being program under three scenarios.

The final increase in the integral indicator of KPIs over 1 year amounted to: under scenario 1 – 11.85 units; under scenario 2 – 11.8 units; under scenario 3 – 12 units. As you can see, provided that the total investment volumes in the well-being program areas are identical, the key criterion for making a decision is the final value of the integral KPI. As part of this example, it is recommended to choose scenario number 3, i.e. a scenario in which the company's management artificially does not prioritize specific areas of the program. Thus, the tool proposed by the authors independently allows you to determine the optimal portfolio of program activities, taking into account the initial input parameters.

The final portfolio of measures is implemented in the corporate well-being program. We recommend conducting a flexible quarterly portfolio adjustment because the results of specific activities more accurately indicate their impact on the competence of a group of employees and their burnout reduction. Thus, this will introduce another indicator for the measures in terms of their contribution to the integral KPI. Future studies are planned to collect retrospective data on the implemented measures to modify the proposed model and improve its accuracy.

5. Discussion

Most studies investigating the effects of corporate well-being programs on an organization's functioning confirm the influence of such programs on human capital and employee productivity. However, these studies focus not on the overall process of human capital development using the program (from investment into personnel well-being to the productivity of their work) but rather its fragments. Gonsiorowska and Zieba [3], Salman et al. [4], Mathibe and Chinyamurindi [7], and Nursel [8] describe the influence of the program on various aspects of employee competence and burnout. Naczenski et al. [14], Kurniawan et al. [18], and Kim and Jung [22] explore the influence of different elements of employee competence and burnout on their productivity and efficiency. Several studies propose economic and mathematical models describing the interrelationships between elements of human capital development through the well-being program [26, 27].

Existing works do not consider the differentiated impact of well-being program directions on aspects of employees' work, on the one hand, and the synergetic effects arising from the implementation of several directions at once, on the other. The analyzed studies do not address the effects of multiperiodicity in implementing these programs and the optimization aspects of their development.

In the presented model, we accounted for these shortcomings and the demands of real commercial organizations.

A distinctive feature of our model is the consideration of two uncertainty levels in forming the optimal portfolio of well-being program measures. The first level has to do with the reliability of estimates of the numerical coefficients of influence channels' functional dependencies found by defuzzifying fuzzy piecewise constant regressions based on the matrices of correspondence between the fuzzy categories of the integral indicators of expectations, competencies, burnout level and KPIs. The second level relates to the need to set several limit parameters determined by experts (e.g., the finances invested in a specific direction of the program) in fuzzy numbers.

The model is marked by multiperiodicity and the opportunity to adjust the portfolio of well-being program measures for the entire planning horizon during quarterly monitoring of the achieved results, assessing deviations from the plan and reshaping the portfolio of activities based on new data. Thus, the model steadily improves the accuracy of executive decisions and employee efficiency.

Thus, our study rests on the theoretical foundation of prior research while also solving a different problem using mathematical apparatus. The need for new methods owes to the existing trend of improving financial resource management, on the one hand, and ensuring that the organization continuously develops and creates competitive advantages to secure a market share sufficient for success in the face of growing competition, on the other.

The approbation of the model on real business units of a commercial company enabled us to form a portfolio of well-being program measures depending on the priority assigned to the directions by company management. The analysis of prioritization scenarios highlighted the most significant directions that have the greatest influence on improving employees' competencies and reducing their burnout and the organization's overall efficiency. The reliability of the results is confirmed by the experts in personnel performance management who participated in our survey and research. We can conclude that the proposed method is a functional tool that allows one to create the optimal portfolio of measures under the corporate well-being program based on an organization's starting parameters. Therefore, the hypothesis posed in this study was confirmed.

Conclusion

In this study, we proposed a fuzzy method for creating the optimal portfolio of measures in a corporate well-being program to achieve maximum progress toward employees' target KPIs by developing their competencies and reducing burnout.

This method accounts for the drawbacks shown in the literature analysis and puts forward possible solutions: the effects stemming from the multiperiodicity of the implementation of a well-being program are considered; the optimization aspects of developing the portfolio of measures under the well-being program are examined; the uncertainties and risks associated with the subjectivity of input parameters are accounted for; the comprehensive influence of employee competence and burnout on productivity and efficiency is explored.

We built a fuzzy dynamic model to create the optimal portfolio of well-being program measures. The loss function used in the model is an integral indicator characterizing the attainment of target employee KPIs while accounting for the importance of each KPI for the organization.

The optimization variables in the model are binary variables that determine the inclusion of a measure into the organization's well-being program at a specific time within the given planning period.

The model proceeds from three types of limits:

1) the total amount of investment allocated to the well-being program is limited by the organization's available budget;

2) the permissible amount of financial resources to be invested in a direction of the well-being program has a top limit represented by a specified fuzzy number;

3) an increase in each employee's integral competence indicator due to investment in the well-being program does not exceed a given fuzzy value.

The features of the program are as follows.

First, the model relies on the functional dependencies of influence channels (between the categories of integral indicators of expectations, competencies, burnout level and KPIs) found by defuzzifying fuzzy piecewise constant regressions based on the matrices of correspondence between the fuzzy categories of integral indicators;

Second, the set of input parameters of model limits is specified in fuzzy numbers, which allows for either tightening or loosening the limits by setting their confidence level.

The fuzzy approach allows us to account for the uncertainties and risks associated with subjective raw data, an important component of the model.

The proposed method demonstrates theoretical significance because it is a tool for creating the optimal portfolio of measures within the corporate well-being program in terms of specific employees and points in time. This allows for structuring the allocation of financial resources to different program directions to improve the performance and efficiency of staff and the organization.

Our findings confirm the study hypothesis.

The study's practical value lies in the fact that it presents persons tasked with managing the efficiency and overall state of organization employees with an instrumental apparatus for determining and forming a well-grounded portfolio of measures for the corporate well-being program to influence employee productivity.

In the future, we plan to conduct an in-depth study to quantitatively assess the risks of deviation of actual progress toward the achievement of employees' target KPIs over the planning horizon from the ones forecasted based on the creation of the optimal well-being program. ■

Acknowledgments

The research was financed by the Russian Science Foundation grant № 23-28-01333[1], at Vladivostok State University.

References

1. Mazelis L., Lavrenyuk K., Grenkin G., Krasko A. (2023) Conceptual model for the development of employee competencies through the well-being implementation. *International Journal of Sustainable Development & Planning*, vol. 18, no. 11, pp. 3557–3566. https://doi.org/10.18280/ijsdp.181120

2. Zeyi L., Jiayin H. (2024) Key factors for sustainability – corporate culture and employee well-being. *Advances in Economics, Management and Political Sciences*, vol. 78, pp. 263–267. https://doi.org/10.54254/2754-1169/78/20241734

3. Gonsiorowska M., Zieba M. (2023) Employee well-being and its potential link with human knowledge risks. *24th European Conference on Knowledge Management*, pp. 430–436. https://doi.org/10.34190/eckm.24.1.1578

4. Salman M., Saleem I., Ganie S.A. (2023) Human resource management practices as antecedents of employee competencies: Empirical evidence from the banking industry. *Management and Labour Studies*, vol. 3, no. 48, pp. 381–398. https://doi.org/10.1177/0258042X221138362

5. Rudyk O., Maksymova J., Zaletska I. (2023) The importance of developing the professional competences of the staff of ser-vice sector enterprises. *Market Economy Modern Management Theory and Practice*, vol. 21, no. 3, pp. 461–475. https://doi.org/10.18524/2413-9998.2022.3(52).275826

[1] https://rscf.ru/project/23-28-01333/

6. Loyalka P., Huang X., Zhang L., et al. (2015) The impact of vocational schooling on human capital development in developing countries: Evidence from China. *The World Bank Economic Review*, vol. 30, no. 1, pp. 143–170. https://doi.org/10.1093/wber/lhv050

7. Mathibe M.S., Chinyamurindi W.T. (2021) Determinants of employee mental health in the South African public service: The role of organizational citizenship behaviors and workplace social support. *Advances in Mental Health*, vol. 19, no. 3, pp. 306–316. https://doi.org/10.1080/18387357.2021.1938153

8. Nursel A.M. (2020) The relationship between employee well-being, burnout and perceived organizational support in healthcare professionals. *Journal of International Health Sciences and Management*, vol. 6, no. 12, pp. 34–39. https://doi.org/10.48121/jihsam.788565

9. Konieczny G., Kolisnichenko P., Górska M., Górski T. (2023) The role of well-being in sustainable corporate development of companies. *Economics, Finance and Management Review*, vol. 15, no. 3, pp. 59–67. https://doi.org/10.36690/2674-5208-2023-3-59-67

10. Bes I., Shoman Y., Al-Gobari M., et al. (2023) Organizational interventions and occupational burnout: a meta-analysis with focus on exhaustion. *International Archives of Occupational and Environmental Health*, vol. 9, no. 96, pp. 1211–1223. https://doi.org/10.1007/s00420-023-02009-z

11. Genkin E.V., Ozerov S.L., Satina A.V. (2022) Implementation of the WELL-BEING program to maintain the emotional and psychological state of the company's employees. *Normirovaniye i Oplata Truda v Promyshlennosti (Labor Norming and Remuneration in Industrial Sector)*, no. 2 (in Russian). https://doi.org/10.33920/pro-3-2202-03

12. Konovalova V.G. (2018) Experience, development trends and problems of implementing corporate employee health and well-being programs. *Management of the Personnel and Intellectual Resources in Russia*, vol. 7, no. 5, pp. 19–29 (in Russian). https://doi.org/10.12737/article_5bd1c6797f9068.21933673

13. Rong C. (2024) The impact of employees' health and well-being on job performance. *Journal of Education, Humanities and Social Sciences*, vol. 29, pp. 372–378. https://doi.org/10.54097/9ft7db35

14. Naczenski L.M., de Vries J.D., van Hooff M.L.M., Kompier M.A.J. (2017) Systematic review of the association between physical activity and burnout. *Journal of Occupational Health*, vol. 59, no. 6, pp. 477–494. https://doi.org/10.1539/joh.17-0050-RA

15. Shih S.-P., Jiang J.J., Klein G., Wang E. (2013) Job burnout of the information technology worker: Work exhaustion, depersonalization, and personal accomplishment. *Information & Management*, vol. 50, no. 7, pp. 582–589. https://doi.org/10.1016/j.im.2013.08.003

16. Diamantidis A.D., Chatzoglou P. (2019) Factors affecting employee performance: An empirical approach. *International Journal of Productivity and Performance Management*, vol. 68, no. 1, pp. 171–193. https://doi.org/10.1108/IJPPM-01-2018-0012

17. Otoo F.N.K., Mishra M. (2018) Measuring the impact of human resource development (HRD) practices on employee performance in small and medium scale enterprises. *European Journal of Training and Development*, vol. 42, nos. 7/8, pp. 517–534. https://doi.org/10.1108/EJTD-07-2017-0061

18. Kurniawan A., Sanosra A., Qomariah N. (2023) Efforts to increase motivation and performance based on employee competency and job characteristics. *Journal of Economics, Finance and Management Studies*, vol. 6, no. 7, pp. 3153–3162. https://doi.org/10.47191/jefms/v6-i7-17

19. Ateke B.W., Nwulu C.S. (2018) Employee competency development and organizational resilience. *International Journal of Social Sciences and Management Research*, vol. 4, no. 3, pp. 31–44.

20. Dharmanegara I.B.A., Sitiari N.W., Wirayudha I.D.G.N. (2016) Job competency and work environment: the effect on job satisfaction and job performance among SMEs worker. *IOSR Journal of Business and Management*, vol. 18, no. 1, pp. 19–26. http://doi.org/10.9790/487X-18121926

21. Sitopu Y.B., Sitinjak K.A., Marpaung F.K. (2021) The influence of motivation, work discipline, and compensation on employee performance. *Golden Ratio of Human Resource Management*, vol. 2, no. 1, pp. 72–83. https://doi.org/10.52970/grhrm.v1i2.79

22. Kim J., Jung H.-S. (2022) The effect of employee competency and organizational culture on employees' perceived stress for better workplace. *International Journal of Environmental Research and Public Health*, vol. 19, no. 8. https://doi.org/10.3390/ijerph19084428

23. Matani M., Bidmeshki G.A. (2020) The role of burnout on reducing employees' performance. *Journal of Management and Accounting Studies*, vol. 6, no. 2, pp. 39–46. https://doi.org/10.24200/jmas.vol6iss02pp39-46

24. Trees Bolt E.E., Homer S.T. (2023) Employee corporate social responsibility and well-being: The role of work, family and culture spillover. *Employee Relations: The International Journal*, vol. 46, no. 7. https://doi.org/10.1108/ER-02-2023-0097

25. Kristiana Y., Sijabat R., Sudibjo N., Bernarto I. (2024) Hotel employee well-being: The role of job crafting and work engagement. *International Journal of Religion*, vol. 5, no. 7, pp. 200–211. https://doi.org/10.61707/az74f509

26. Saleh R. (2023) The effect of motivation factors on employee performance: Performance appraisal using data mining. Journa*l of Business Theory and Practice*, vol. 11, no. 1, pp. 43–58. https://doi.org/10.22158/jbtp.v11n1p43

27. Ozcelik H., Langton N., Aldrich H. (2008) Doing well and doing good: The relationship between leadership practices that facilitate a positive emotional climate and organizational performance. *Journal of Managerial Psychology*, vol. 23, no. 2, pp. 186–203. https://doi.org/10.1108/02683940810850817

28. Fastje F., Mesmer-Magnus J., Guidice R., Andrews M.C. (2023) Employee burnout: The dark side of performance-driven work climates. *Journal of Organizational Effectiveness: People and Performance*, vol. 10, no. 1, pp. 1–21. https://doi.org/10.1108/JOEPP-10-2021-0274

29. Rony Z., Pardosi H.D. (2021) Burnout digital monitoring on employee engagement at the company. *International Journal of Research in Business and Social Science*, vol. 10, no. 7, pp. 156–162. https://doi.org/10.20525/ijrbs.v10i7.1412

30. Williams H.P. (2013) *Model building in mathematical programming, 5th Edition*. Willey.

31. Ammar E., Khalifa H.A. (2003) Fuzzy portfolio optimization a quadratic programming approach. *Chaos, Solitons & Fractals*, vol. 18, no. 5, pp. 1045–1054. https://doi.org/10.1016/S0960-0779(03)00071-7

32. Akopov A.S. (2014) Parallel genetic algorithm with fading selection. *International Journal of Computer Applications in Technology*, vol. 49, nos. 3/4, pp. 325–331. https://doi.org/10.1504/IJCAT.2014.062368

33. Mazelis L., Solodukhin K., Tarantaev A. (2018) Fuzzy optimization models for project portfolio rolling planning taking into account risk and stakeholder interests. *The Journal of Social Sciences Research*, no. 3, pp. 201–210. https://doi.org/10.32861/jssr.spi3.201.210

34. Mazelis L.S., Lavreniuk K.I., Grenkin G.V. (2023) Analysis of the Relation Between Expectation of Employees from Corporate Environment and their Burnout. *Journal of Applied Economic Research*, vol. 22, no. 4, pp. 1034–1055 (in Russian). https://doi.org/10.15826/vestnik.2023.22.4.040

35. Mazelis L.S., Grenkin G.V., Lavreniuk K.I., Krasko A.A. (2024) Fuzzy method for assessing the impact of employee competencies and burnout on their achievement of key performance indicators. *Gibridnye i sinergeticheskie intellektualnye sistemy (Hybrid and synergistic intelligent systems). Collection of articles from the materials of the VII All-Russian Pospel Conference. Kaliningrad, St. Petersburg*, pp. 411–422 (in Russian).

36. Mazelis L.S., Lavreniuk K.I., Krasko A.A. (2023) Modeling the development of employee competence by investing in well-being program activities. *Market Economy Problems*, no. 3, pp. 53–67 (in Russian). http://doi.org/10.33051/2500-2325-2023-3-53-67

37. Anshin V.M., Demkin I.V., Nikonov I.M., Tsarkov I.N. (2008) *Models of project portfolio management under uncertainty*. Moscow: Moscow State Aviation Technological University (in Russian).

38. Wang J., Hwang W.-L. (2007) A fuzzy set approach for R&D portfolio selection using a real options valuation model. *Omega*, vol. 35, no. 3, pp. 247–257. https://doi.org/10.1016/j.omega.2005.06.002

About the authors

Lev S. Mazelis

Doctor of Sciences (Economics);

Professor, Department of Mathematics and Modeling, Vladivostok State University, 41, Gogol St., Vladivostok 690014, Russia;

E-mail: lev.mazelis@vvsu.ru

ORCID: 0000-0001-7346-3960

Gleb V. Grenkin

Candidate of Sciences (Physics and Mathematics);

Associate Professor, Department of Mathematics and Modeling, Vladivostok State University, 41, Gogol St., Vladivostok 690014, Russia;

E-mail: gleb.grenkin@vvsu.ru

ORCID: 0000-0002-1307-3757

Kirill I. Lavrenyuk

Project Manager, Yandex, 16, Lva Tolstogo St., Moscow 119021, Russia;

E-mail: kirlavrenyuk@yandex-team.ru

ORCID: 0000-0002-9092-3196

Andrey A. Krasko

Candidate of Sciences (Economics);

Associate Professor, Department of Mathematics and Modeling, Vladivostok State University, 41, Gogol St., Vladivostok 690014, Russia;

E-mail: andrey.krasko@vvsu.ru

ORCID: 0000-0002-6136-6893

DOI: 10.17323/2587-814X.2025.1.72.92

Assessment of risks of failure to achieve target values of indicators for an organization's intellectual capital based on a fuzzy model

Konstantin S. Solodukhin *
E-mail: k.solodukhin@mail.ru

Georgij S. Zavalin
E-mail: georgiy.zavalin08@vvsu.ru

Daria V. Makarova
E-mail: malnova.daria@vvsu.ru

Vladivostok State University, Vladivostok, Russia

* Corresponding Author

Abstract

The processes of formation and development of intellectual capital in the digital economy are ill structured processes occurring in conditions of a significant increase in the speed and unpredictability of changes in the external environment. This makes it extremely difficult to use previous experience and probabilistic forecasts when assessing the risks of failure to achieve strategic goals for the development of the intellectual capital of an organization. At the same time, undesirable deviations in achieving these goals can lead to significant negative consequences. In this regard, there is a need to develop appropriate fuzzy methods and models, all of which determines the relevance of this work. The purpose of this study was to develop a fuzzy method for assessing the risks of failure to achieve the strategic goals of an organization in the field of intellectual capital development. The method is based on a fuzzy model developed by the authors which allows us to take into account the uncertainty tolerance of the decision maker. Testing the method on the example of a specific organization showed the possibility of its practical applicability. We provide quantitative assessments and qualitative interpretations of the risk levels of failure to achieve target indicators for the development of the intellectual capital of an organization (a large regional university).

Keywords: intellectual capital, fuzzy model, risks of failure to achieve goals, project portfolio risk management, tolerance to uncertainty

Citation: Solodukhin K.S., Zavalin G.S., Makarova D.V. (2025) Assessment of risks of failure to achieve target values of indicators for an organization's intellectual capital based on a fuzzy model. *Business Informatics*, vol. 19, no. 1, pp. 72–92. DOI: 10.17323/2587-814X.2025.1.72.92

Introduction

The formation and implementation of strategic goals are the core of the strategic process in an organization [1, 2]. Other key features of the strategic process are strategic reflection (search for the best ways to achieve goals) and strategic actions (implementation of programs, projects, events) [3].

The strategic process (like any other activity of an economic entity) is carried out in conditions of uncertainty, the presence of which leads to the risks of failure to achieve goals.

The system of strategic goals of an organization may include goals that require unconditional fulfillment. Failure to achieve these goals may lead to the termination of its existence. In this regard, the accuracy of assessing the risks of failure to achieve such goals is of critical importance [4]. The presence of risks of failure to achieve the goal requires a reserve of "dead" ("insurance") resources, the use of which is assumed in the event of threats on the way to the goal [5]. Increasing the accuracy of risk assessment will avoid unnecessary "freezing" of resources.

Inaccuracy in assessing the risks of failure to achieve goals may also lead to unnecessary expenditure of resources allocated at the stage of planning strategic actions (formation of development plans and programs, selection of projects). In turn, this may lead to overfulfillment of some goals with corresponding negative consequences [6].

From the standpoint of the operational theory of risk management, the level of risk of deviation from the organization's goals can be considered a characteristic of the quality of management. Within the framework of the theory, economic risk is considered as a category reflecting the measure of the reality of an undesirable deviation from the goals and the amount of losses caused by this deviation. In this case, the assessment of the measure of deviation can be carried out, for example, in the form of additional costs or a decrease in the economic effect caused by the implementation of a management decision [7].

In the theory of economic risk management, the analysis and comparison of information about the economic situation can occur in two different situations (time slices): at the stage of decision preparation and at the stage of decision implementation [7].

For the second stage, the methods for assessing the risks of failure to achieve goals are devoted to identifying and assessing possible obstacles – events in the external environment that prevent the achievement of goals. Depending on the method of assessing obstacles, models can be probabilistic [8–13] or fuzzy [14–18]. There are also models in which the fuzzy approach is supplemented by probabilistic analysis [19].

At the first stage, the generation of decision options (management actions) takes place, the sufficiency (for decision-making) of the available information is assessed, the initial information is supplemented and clarified, the results of applying the selected decisions and their compliance with the set goals are predicted [7].

A management decision (a decision option) is a decision to select from a variety of possible strategic measures (each of which has the necessary resources and certain consequences) a specific set of measures aimed at achieving goals [20]. In most cases, strategic measures can be considered as projects (reconstruction and development), and accordingly, the choice of a set of measures represents the formation of a project portfolio.

Project Portfolio Risk Management (PPRM) is an established field of research that focuses on the processes of identifying and balancing project portfolio risks while striving to maximize the value received by the organization, as reflected in the achieved impact on strategic objectives [21]. Thus, PPRM focuses on the management capabilities of the organization to reduce the negative impact of risks and expand opportunities, taking into account the interdependencies between projects and risks [22–25].

In PPRM, both established and promising research areas can be identified [21]. One of these areas is the development of models and methods for assessing project portfolio risks (PPR).

Existing methods for assessing portfolio risk are based on various approaches to modeling uncertainty. In stochastic models, uncertainty is described by a probability distribution. In fuzzy models, uncertainty is described by a membership function [26]. In non-stochastic game models, an unstructured set of potentially possible values of an elementary event is specified [7].

Risk management theory shifts the emphasis towards the study of ill structured phenomena [7]. At present, these phenomena occur in conditions of a significant increase in the speed and unpredictability of changes in the external environment. This makes it extremely difficult to use previous experience (historical data) [27] and probabilistic forecasts.

In this regard, in recent years, a wide variety of fuzzy models have been actively developed to assess the risks of a project portfolio and take into account the resulting risk assessments when selecting a portfolio. It is worth noting the work of Mohagheghi and co-authors [28], who proposed an original approach to selecting the optimal portfolio of projects to achieve sustainable development goals using interval-valued fuzzy sets to take into account uncertainty and various criteria (indices) reflecting risks. In the work of Shatalova [29], fuzzy models for assessing the comparative effectiveness of innovative projects in the formation of a program for the technological development of an industrial enterprise under conditions of non-stochastic uncertainty, taking into account risks, are proposed. In the work of Zaidouni et al. [30], fuzzy factor analysis is used to predict the risks of an IT project portfolio. The works of Mehlawat et al. [31], Khanjani Shiraz et al. [32], Deng and Yuan [33], Rahiminezhad Galankashi et al. [34], Mohseny-Tonekabony et al. [35], Wang et al. [36], Dehghani [37], Nguyen et al. [38], Abtahi [39], Yang et al. [40] should also be noted.

At the same time, the potential of the fuzzy approach in analyzing project portfolio risks has not been fully realized [21].

Achieving strategic goals leads to a change in strategic potential – "the totality of 'strategic' resources at the disposal of an enterprise that are of decisive importance for the capabilities and boundaries of the enterprise's functioning in certain conditions" [41, p. 352]. In the context of the development of the digital economy, the most important strategic resource of an organization is its intellectual capital.

The intellectual capital development strategy is part of the overall development strategy of the organization. The set of strategic goals that contribute to the development of intellectual capital is a subset of the set of all strategic goals of the organization. Undesirable deviations in achieving these goals lead to the negative consequences described above. At the same time, assessing the risks of such deviations is significantly more difficult than for most other strategic goals due to the specifics of intellectual capital: a large number of implicit and "qualitative" development factors, a strong dynamic impact of organization's intellectual capital components, etc. [42, 43].

In this regard, it is precisely fuzzy models for assessing the organization's intellectual capital and its components that come to the fore. Considerable attention is paid to such models in the review by Cosa et al. [44]. Of the relevant publications published no earlier than 2020, the following should be noted: Bustamante et al. [45], Kozlovskyi et al. [46], Çevik and Arslan [47], Pokrovskaia et al. [48], Lucchese et al. [49], Gross-Gołacka et al. [50].

At the same time, such an important issue (for the above reasons) as assessing the risks of not achieving the organization's intellectual capital targets remained untouched by the developers of the models.

Thus, the purpose of this study is to develop a fuzzy method for assessing the risks of failure to achieve strategic goals to support management decision-making in the area of developing the organization's intellectual capital.

The work contains four main sections. The first section provides a basic scheme of the method and describes the content of its main stages. The proposed method is based on a fuzzy model, which is described in detail in the second section of the article. The third section presents the results of testing the method using the example of a specific organization (a large regional university). The "Discussion" section formulates the advantages of the method; its contribution to the development of various areas of research; the possibilities and conditions for its use for organizations of various types and industries.

1. The method for assessing the risks of failure to achieve the target indicators of the organization's intellectual capital

The basic scheme of the proposed method for assessing the risks of failure to achieve strategic goals for the development of an organization's intellectual capital (hereinafter referred to as organization's intellectual capital) is shown in *Fig. 1.*

First of all, it is necessary to form a causal field of OIC development indicators according to the previously proposed scheme [51]. The selection of explicit and implicit factors located at the lowest level of the hierarchy and the assessment of their influence are

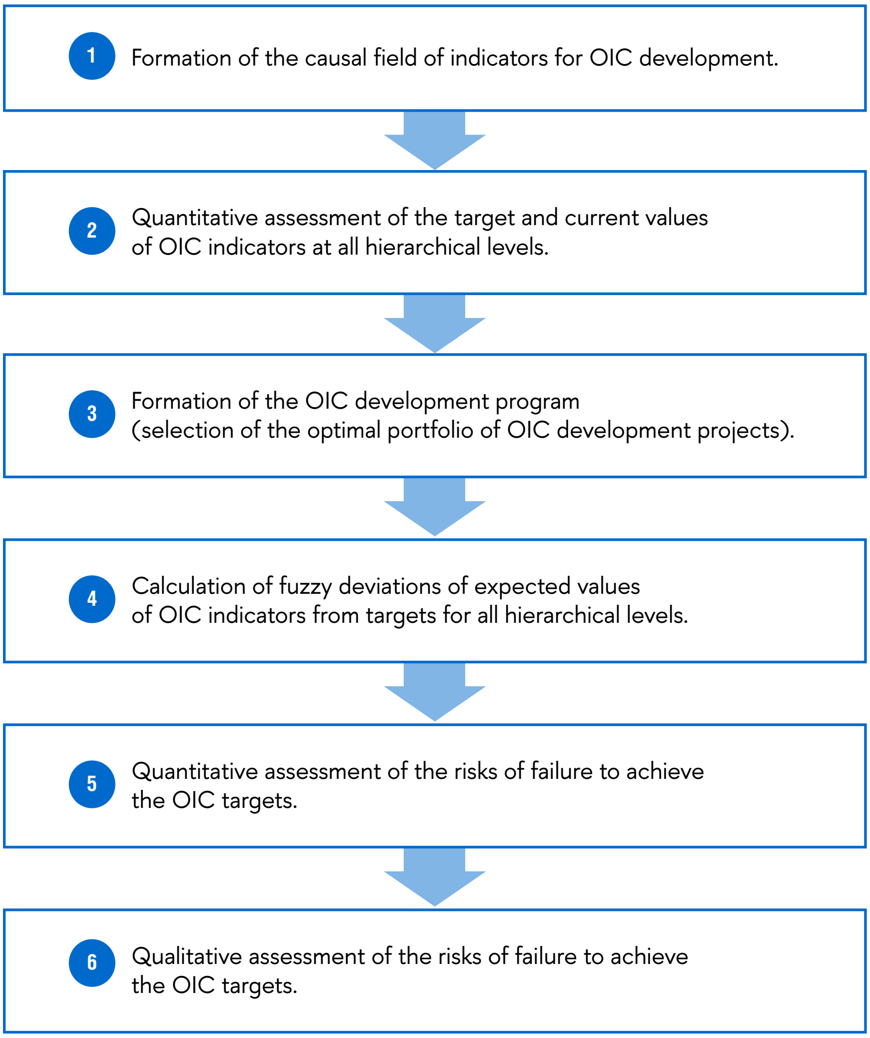


Fig. 1. Basic scheme of the method for assessing the risks of failure to achieve the organization's goals for the development of intellectual capital.

based on the author's fuzzy model. An example of a hierarchical structure of OIC development indicators for a specific organization (university) is given in section 3 (*Fig. 2*) [52, 53].

Some of the OIC indicators of the lowest level of the hierarchy are naturally assessed in quantitative scales. The other part (as well as almost all indicators of the overlying levels) is assessed in qualitative scales. Accordingly, significant difficulties arise in the process of moving up the hierarchy from bottom to top when evaluating the OIC integrated indicators up to the integral indicator of the intellectual capital of the organization as a whole. Another difficulty in assessing the OIC development indicators is related to the emergence of cycles in the hierarchy. In this regard, a fuzzy model was proposed that allows for a quantitative assessment of the OIC development indicators at all hierarchical levels [52].

This model was developed to assess the current level of OIC. At the same time, it can be successfully used in setting target values for OIC development indicators. Strategic goals in the field of OIC development arise within the framework of the general strategy of the organization, formalized on the basis of a modified Balanced Scorecard [51]. Accordingly, as a rule, crisp target values of the "quantitative" resulting indicators of such goals are known. The values of the "qualitative" resulting indicators are often not explicitly included in the strategy. And in the event that they are somehow established, it is unclear how to determine the target values of various integral OIC development indicators (corresponding to the main components of the OIC, the organization's abilities for certain cognitive activities and various aspects within the framework of certain types of cognitive activities). This is what this model can be used for.

After the fuzzy target and current values of all indicators in the hierarchy have been quantitatively determined, a program for the OIC strategic development can be formed. This program is a set of strategic events (projects), the implementation of which results in the achievement (to one degree or another) of the strategic goals for the OIC development. Thus, the values of the lagging indicators of these goals (indicators of the OIC development) reach the target values or approach them.

In [53], fuzzy optimization models for forming a project portfolio for the OIC development taking into account risks are proposed. Fuzzy optimization problems are reduced to crisp problems of Boolean quadratic programming using the methods proposed in [54–56]. In this case, it is necessary to set the satisfaction degrees for the objective function and each constraint. The satisfaction degrees determine the rigidity of the constraints and can influence the composition of the portfolio. It can be considered that the set satisfaction degrees determine the decision maker's tolerance to uncertainty [57].

The proposed models allow us to form an optimal portfolio of projects for the OIC development (depending on the selected target function and the specified constraints). At the same time, we obtain the expected fuzzy values of all OIC indicators in the hierarchy (when implementing this portfolio). Accordingly, fuzzy relative deviations of expected values from target values (degrees of failure to achieve target indicators) can be calculated.

These deviations reflect the risks of failure to achieve the target indicators of the OIC development. In turn, additional quantitative information can be obtained for convenience and increased accuracy of the qualitative interpretation of the risk level. First of all, each fuzzy deviation can be defuzzified by the selected method. In addition, a deviation fuzziness index can be calculated, which reflects the degree of blurriness of the fuzzy deviation [58].

In [42], a method for interpreting the risk level based on correspondence with the maximum possible value is proposed. Within the framework of an alternative approach, the correspondence coefficients of each fuzzy deviation to the intervals into which the universal set is divided are calculated [18]. The distribution of the correspondence coefficients determines the qualitative risk assessment.

The correspondence coefficients are the relative areas of figures bounded by the membership function curve from above and a given alpha level from below. It is important to note that the choice of the alpha level (as well as the choice of satisfaction degrees) determines the decision maker's tolerance to uncertainty: the lower the alpha, the greater the tolerance. A change in the decision maker's tolerance to uncertainty can lead to a change in the qualitative risk assessments. Moreover, as will be shown below, these changes can be two-sided: for some indicators, with an increase in the alpha level, the risk assessments will increase, while for others, on the contrary, they will decrease. Taking into account the decision maker's tolerance to uncertainty is a significant advantage of the proposed method for assessing the risks of failure to achieve the goals for the OIC development.

2. Fuzzy model

Let a causal field of OIC development indicators be formed. Fuzzy values of "qualitative" indicators of the lowest level are defined. Fuzzification of crisp values of "quantitative" indicators is carried out. Fuzzy inference systems (or simplified procedures [59]) are specified, allowing us to obtain fuzzy values of OIC indicators of all hierarchy levels. Target values are defined and current values of all indicators are calculated [52].

Let scenarios of possible changes in the internal and external environment be given, their fuzzy probabilities be defined. A set of possible projects for the OIC development are formed, and their budgets are fuzzy. Changes in the current values of the indicators of the lowest level (and, through them, all indicators in the hierarchy) are determined as a result of the implementation of each project for each scenario [53].

Note that the values of all the OIC indicators (except for the lowest level of the hierarchy), scenario probabilities, project budgets and changes in the indicators of the lowest level are Gaussian fuzzy numbers. The values of the indicators of the lowest level are trapezoidal fuzzy numbers.

We are formulating a program for the OIC development based on one of the proposed fuzzy optimization models (which are fuzzy Boolean quadratic programming problems) [53]. Let us recall that the models differ in their objective functions (maximization of specific utility or minimization of portfolio risk) and constraints (on the total budget of the development program, on the value of the program risk, or on the value of the expected specific utility). Let us also recall that the composition of the optimal project portfolio depends on the specified satisfaction degrees (on the objective function and constraints). For the selected optimal portfolio, we have expected fuzzy values of all OIC indicators.

We calculate the degree of failure to achieve the OIC target indicators as relative deviations of expected values from the target ones as follows:

$$\Delta_i = \frac{IC_i^T - IC_i^E}{IC_i^T}, \qquad (1)$$

where

Δ_i is degree of non-achievement of the target value of the i-th OIC indicator;

IC_i^T is target value of the i-th OIC indicator;

IC_i^E is expected value of the i-th OIC indicator.

Note that in the general case, due to the fuzziness of the variables,

$$\Delta_i \neq 1 - \frac{IC_i^E}{IC_i^T}.$$

It is important to note that if $IC_i^E > IC_i^T$ (for the chosen fuzzy set comparison method), there may still be risks of not achieving the target value of the i-th indicator. In this case, we will calculate the fuzzy relative deviations of expected values from target values using the formula:

$$\Delta_i = \frac{IC_i^E - IC_i^T}{IC_i^E}. \qquad (2)$$

Let $IC_i^T > IC_i^E$. Then $(\xi_i; \eta_i) \subseteq (a_i, b_i)$, where $(\xi_i; \eta_i)$ is the support of the fuzzy set Δ_i, a_i is the infimum of the support of the fuzzy set

$$\frac{IC_i^T - IC_i^T}{IC_i^T},$$

b_i is the supremum of the support of the fuzzy set

$$\frac{IC_i^T}{IC_i^T}.$$

If $IC_i^E > IC_i^T$, then a_i should be taken as the infimum of the support of the fuzzy set

$$\frac{IC_i^E - IC_i^E}{IC_i^E},$$

and b_i should be taken as the supremum of the support of the fuzzy set

$$\frac{IC_i^E}{IC_i^E}.$$

Let's consider bijections $f_i : (a_i; b_i) \to (0;1)$,

$$f_i(x) = \frac{x - a_i}{b_i - a_i}. \quad (3)$$

These mappings allow us to move from fuzzy deviations Δ_i to fuzzy sets Δ_i' whose supports are subsets of the interval (0;1).

We calculate the fuzziness indices of the sets using a given method [60]. We denote them θ_i. We perform defuzzification of fuzzy sets Δ_i' using one of the selected methods. The resulting crisp values are denoted by ω_i.

We divide the interval (0; 1) into h intervals (0; ξ_1), (ξ_1; ξ_2), ..., (ξ_{h-1};1) (these intervals may be equal, but in general this is not necessary).

We calculate the coefficients of correspondence of fuzzy sets Δ_i' to these intervals ($0 \leqslant \alpha < 1$). We denote them $\delta_1^i, \delta_2^i, \ldots, \delta_h^i$.

For $\alpha = 1$, δ_j^i is the measure of intersection of the core of the fuzzy set Δ_i' and the interval $(\xi_{j-1};\xi_j)$. As a defuzzification method for a single alpha level, it is reasonable to use the mean maximum method.

A qualitative assessment of the risk of failure to achieve the target value of the i-th OIC indicator occurs in a given linguistic scale based on the distribution of coefficients $\delta_1^i, \delta_2^i, \ldots, \delta_h^i$. If $IC_i^T > IC_i^E$, then the higher the correspondence coefficients with smaller lower indices and the smaller the coefficients with larger lower indices, the lower the risk level, and vice versa. If $IC_i^E > IC_i^T$, then the risk level is lower, the smaller the correspondence coefficients with smaller subscripts and the larger the coefficients with larger subscripts.

Also taken into account are ω_i and θ_i. Note that the smaller θ_i, the more reliable the qualitative risk assessments obtained, and vice versa.

In the future, a formalized procedure for determining a certain integral coefficient (as a function of variables $\delta_1^i, \delta_2^i, \ldots, \delta_h^i, \omega_i, \theta_i$) may be proposed, on the basis of which the interpretation of the risk level will occur. However, the selection of such a function is complicated by the non-monotony in the general case of the sequence $\delta_1^i, \delta_2^i, \ldots, \delta_h^i$.

3. Approbation of the method

The method was tested on the example of a large regional university (Vladivostok State University, VVSU). Using the example of this university, some fuzzy models were previously tested which formed the basis of the method we developed [51–53].

The hierarchical system (causal field) of OIC development indicators is shown in *Fig. 2*. "Qualitative" OIC indicators at the lowest-level of OIC hierarchy are highlighted in yellow, while "quantitative" OIC indicators are marked in green.

The target values of the "quantitative" indicators were taken from the university's strategy, and the current values were taken from the management accounting system. The crisp current and target values have been fuzzified using the method proposed in [52]. Previously, the current values of the "quality" indicators were also obtained in the form of Gaussian fuzzy numbers [52]. Their target values were determined in a similar way (*Table 1*).

Next, fuzzy target and current values of the OIC indicators of all higher levels of the hierarchy were calculated. For this purpose, among other things, the bases of fuzzy production rules for the lowest-level of the hierarchy were formed. The target and current values of the university's intellectual capital indicators at all levels of the hierarchy, defuzzified using the center of gravity method, are presented in *Table 2.*

At the next step, a set of strategic measures (projects) was formed that contribute to the development of the intellectual capital of the university. The project budgets were expertly assessed in a given linguistic scale. Weighted average expert estimates of project budgets in the form of fuzzy Gaussian numbers are shown in *Table 3.*

Most of the projects (projects 1, 2, 3, 5, 8) are clearly aimed at increasing the human capital of the university.

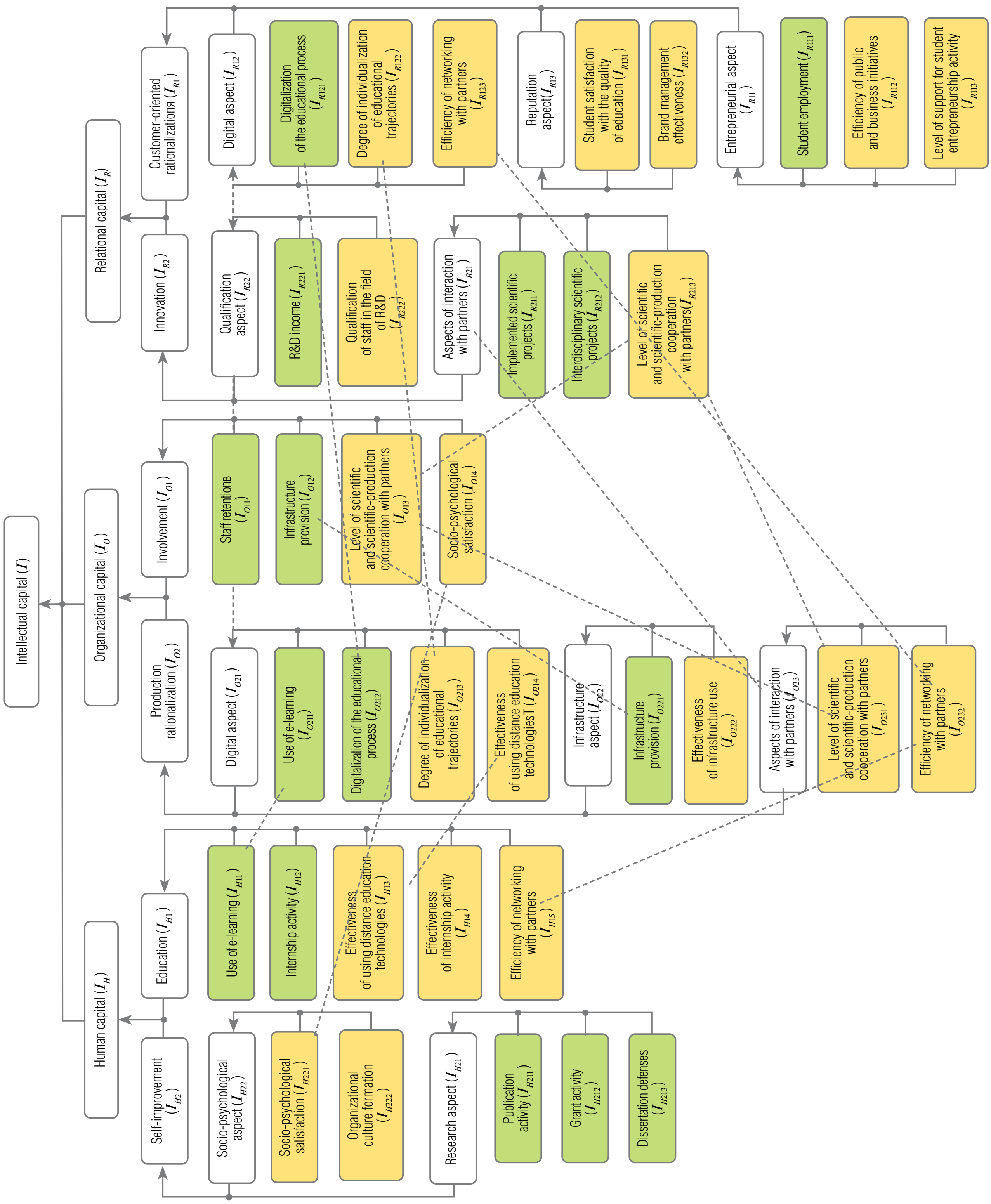


Fig. 2. The causal field of indicators of the university's intellectual capital development.

Table 1.

Current and target values of "qualitative" indicators of the university's intellectual capital[1]

Indicator	Structural component of the OIC	Cognitive activity	Parameters of the approximating Gaussian			
			Current value		Target value	
			μ	σ	μ	σ
Effectiveness of using distance education technologies (I_{H13}, I_{O214})	Human capital	Education	3.1492	0.4778	5.5782	0.6227
Effectiveness of internship activity (I_{H14})	Human capital	Education	1.3452	0.2555	2.7812	0.3451
Socio-psychological satisfaction (I_{H221}, I_{O14})	Human capital, Organizational capital	Self-improvement, Involvement	7.4240	0.8361	8.1243	0.7362
Organizational culture formation (I_{H222})	Human capital	Self-improvement	5.3308	0.5708	6.4687	0.6391
Level of scientific and scientific-production cooperation with partners (I_{O13}, I_{O231}, I_{R213})	Organizational capital, Relational capital	Involvement, Customeroriented rationalization, Innovation	3.1332	0.4081	5.2634	0.6132
Effectiveness of infrastructure use (I_{O222})	Organizational capital	Involvement, Production rationalization	9.0226	0.7395	9.3461	0.6564
Degree of individualization of educational trajectories (I_{O213}, I_{R122})	Organizational capital, Relational capital	Production rationalization, Customeroriented rationalization	1.3452	0.2555	2.6843	0.3154
Efficiency of networking with partners (I_{H15}, I_{O232}, I_{R123})	Organizational capital, Relational capital	Production rationalization, Customeroriented rationalization	3.1492	0.4778	5.5617	0.5996
Student satisfaction with the quality of education (I_{R131})	Relational capital	Customeroriented rationalization	5.3308	0.5708	6.9453	0.6124
Brand management effectiveness (I_{R132})	Relational capital	Customeroriented rationalization	7.3888	0.6720	8.2164	0.7985
Efficiency of public and business initiatives (I_{R112})	Relational capital	Customeroriented rationalization	3.1492	0.4778	5.8041	0.6124
Level of support for student entrepreneurship activity (I_{R113})	Relational capital	Customeroriented rationalization	3.1654	0.5360	5.7926	0.6309
Qualification of staff in the field of R&D (I_{R222})	Relational capital	Innovation	5.3308	0.5708	7.8098	0.7068

[1] A fragment of this table was presented in one of the authors' previous articles [52]. In this article, the last two columns have been added, which provide additionally calculated Gaussian parameters for the target values of the indicators.

Table 2.

Defuzzified target, current and expected values of target indicators of the university's intellectual capital

Fuzzy variable	Current value	Target value	Expected value
I	5.131	6.942	6.408
I_H	5.226	6.933	6.441
I_O	5.012	6.985	5.547
I_R	4.612	6.502	5.862
I_{H1}	2.829	5.218	4.426
I_{H2}	5.515	7.145	7.085
I_{O1}	5.459	6.878	7.464
I_{O2}	5.012	7.035	5.547
I_{R1}	4.612	6.502	5.862
I_{R2}	4.852	7.150	7.938
I_{H21}	4.099	6.517	5.635
I_{H22}	6.660	7.659	8.772
I_{O21}	2.829	5.205	4.454
I_{O22}	7.189	8.214	7.189
I_{O23}	3.939	6.775	4.452
I_{R11}	5.000	7.047	6.463
I_{R12}	2.829	5.440	3.452
I_{R13}	5.341	6.591	6.773
I_{R21}	2.885	5.623	3.672
I_{R22}	5.415	8.393	9.141

At the same time, each of them also affects organizational capital, and almost all of them (except the fifth) affect relational capital (albeit to a lesser extent). Projects 4 and 7 are primarily focused on increasing relational capital, while project 4 also partially affects human and organizational capital (due to the internal stakeholders of the university). Project 6 is primarily designed to ensure the growth of organizational capital, while it also significantly affects the relational capital and, to some extent, the human capital of the university.

Next, three scenarios were considered, the fuzzy probabilities of which are approximated by Gaussians: $\mu = 0.2955$, $\sigma = 0.0318$ (for the pessimistic scenario); $\mu = 0.5238$, $\sigma = 0.0497$ (for the realistic scenario); $\mu = 0.1974$, $\sigma = 0.0226$ (for the optimistic scenario).

Table 3.

Projects for the development of the university's intellectual capital [53]

Project number	Project name	Project budget (parameters of the approximating Gaussian)	
		μ (million rubles)	σ
1	Conducting training for educators in digital educational technologies, including MOOC creation technologies	12.24	1.71
2	Organization of educators' internships at enterprises	5.18	0.82
3	Enhancement of the system of material and non-material rewards and incentives for personnel	20.93	2.32
4	Identification of requests from stakeholders (applicants, parents, students, employers, teaching community) to the university	3.87	0.45
5	Organization of events (business, creative, sports, professional) aimed at team building	4.21	0.74
6	Development of the university's infrastructure component	18.36	2.17
7	Conducting socially oriented and socially significant activities based on the university	6.53	0.98
8	Comprehensive support for the development of scientific activities at the university	20.34	3.19

Within the framework of the scenarios, changes in the lowest-level OIC indicators are determined as a result of the implementation of each project. Weighted average expert responses in the form of Gaussian fuzzy numbers are partially shown in *Table 4*. The first project is missing from the table, since its implementation leads to changes in the values of only those indicators that other projects do not significantly affect.

The formation of the OIC development program was carried out according to the criterion of the minimum risk of the program with restrictions on the program budget (70 million rubles) and the value of the expected specific utility (0.6) with satisfaction degree of 0.9. The program includes projects with numbers: 1, 2, 3, 4, 8. The expected specific utility of the portfolio is 0.64; the expected budget of the portfolio is 62.56 million rubles.

The calculated expected fuzzy changes in all indicators of OIC development as a result of the implementation of the optimal portfolio allowed us to obtain the expected fuzzy values of OIC development indicators at all levels of the hierarchy (their values defuzzified by the centroid method are shown in *Table 2*). For four indicators (I_{O1}, I_{R2}, I_{H22}, I_{R13}), the centers of gravity of fuzzy expected values exceeded the centers of gravity of fuzzy target values, which, as noted above, does not exclude the risks of failure to achieve the target values of the indicators.

Then, using formula (1), fuzzy values of relative deviations of expected values from the target values (fuzzy degrees of non-achievement) were calculated.

Table 4.

Fuzzy changes in the lowest-level indicators of the university's intellectual capital development as a result of project implementation within scenarios (fragment)

OIC indicators / Project numbers	I_{H211}	I_{H221} (I_{O14})	I_{O11}	I_{R111}	I_{R113}	I_{R131}	I_{R132}	I_{R221}	I_{R222}
2				(0.241; 0.020) (0.139; 0.044) (0.078; 0.022)		(0.238; 0.022) (0.132; 0.019) (0.086; 0.011)		(0.152; 0.011) (0.106; 0.046) (0.048; 0.023)	(0.122; 0.011) (0.92; 0.046) (0.037; 0.023)
3	(0.150; 0.023) (0.090; 0.019) (0.035; 0.008)	(0.218; 0.013) (0.152; 0.030) (0.097; 0.011)	(0.225; 0.043) (0.220; 0.012) (0.153; 0.042)					(0.221; 0.048) (0.139; 0.014) (0.053; 0.024)	
4		(0.156; 0.016) (0.124; 0.013) (0.089; 0.009)		(0.231; 0.047) (0.210; 0.043) (0.121; 0.049)	(0.102; 0.046) (0.025; 0.043) (0.014; 0.021)	(0.153; 0.040) (0.075; 0.016) (0.028; 0.043)	(0.134; 0.040) (0.068; 0.044) (0.032; 0.012)	(0.148; 0.011) (0.102; 0.046) (0.051; 0.023)	
5		(0.241; 0.018) (0.141; 0.019) (0.081; 0.014)	(0.042; 0.013) (0.023; 0.006) (0.021; 0.005)						
6		(0.148; 0.015) (0.118; 0.012) (0.076; 0.008)				(0.147; 0.038) (0.091; 0.015) (0.059; 0.007)			
7				(0.136; 0.018) (0.087; 0.009) (0.048; 0.007)	(0.049; 0.037) (0.025; 0.023) (0.016; 0.006)		(0.227; 0.015) (0.210; 0.023) (0.149; 0.011)		
8	(0.205; 0.027) (0.131; 0.031) (0.073; 0.012)		(0.189; 0.016) (0.154; 0.030) (0.097; 0.029)					(0.226; 0.049) (0.204; 0.037) (0.117; 0.012)	(0.233; 0.049) (0.224; 0.037) (0.146; 0.012)

Next, the functions were selected, allowing for the transition from fuzzy deviations to fuzzy normalized deviations (formula (2)), for which the Yager indices of fuzziness with linear Hamming metric [61] (IOF), centers of gravity (COG) and means of maxima (MOM) were calculated.

The interval (0; 1) was divided into five equal intervals (0; 0.2), (0.2; 0.4), ..., (0.8; 1), for which the coefficients of deviation correspondence were calculated (at alpha levels of 0; 0.5 and 1).

Let us give an example of the corresponding calculations for the I_{O2} indicator, which characterizes the cognitive activity "Production rationalization".

The fuzzy normalized deviation of the expected value of the I_{O2} indicator from its target value is shown in *Fig. 3*. The fuzziness index of this fuzzy set is 0.269; center of gravity is 0.493; means of maxima is 0.647.

Table 5 shows the values of the coefficients of correspondence of the fuzzy set to the specified five intervals for different alpha levels.

We present in *Table 6* the calculated coefficients for all indicators and the qualitative interpretations of the risk levels of failure to achieve the target indicators for the development of the university's intellectual capital obtained on their basis (on the linguistic scale {Very Low (VL); Low (L); Medium (M); High (H); Very High (VH)}).

It is easy to see that the resulting qualitative interpretations of risk levels may depend significantly on the chosen alpha level. Thus, for the integral indicator of human capital I_H, the assessment of the risk of failure to achieve its target value changes from "Low" (at $\alpha = 0$) to "Medium" (at $\alpha = 0.5$) and, further, to "High" (at $\alpha = 1$). For the integral indicator of organizational capital I_O, the assessment of the risk of failure to achieve its target value, on the contrary, decreases from "Medium" to "Low" with an increase in alpha. For the I_{H2} and I_{O2} indicators corresponding to the cognitive activities "Self-improvement" and "Production rationalization", with an increase in alpha, the risk assessment increases from "Medium" to "High" (for the I_{O2} indicator, this is even more clearly demonstrated by the results presented in *Table 5*). Here, it is worth noting the low fuzziness index of the Δ'_{H2}, which indicates the reliability of the risk estimates for I_{H2}. At the same time, the risk assessment for the I_{O21} indicator, which characterizes the "Digital aspect" of "Production rationalization", changes from "Low" to "Medium".

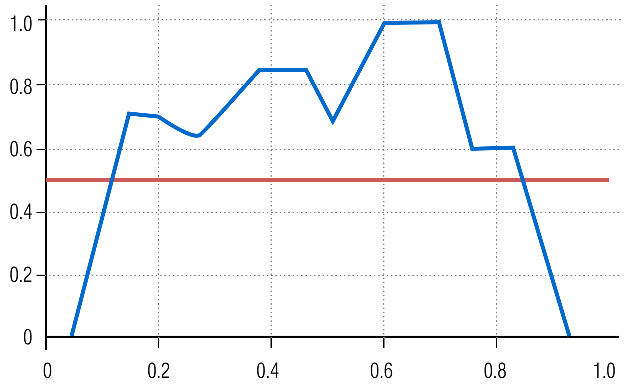


Fig. 3. Fuzzy normalized deviation of the I_{O2} expected value from target value.

Table 5.

Risk assessments of failure to achieve the target value of the I_{O2} indicator

Alpha level	Interval matching coefficients					Interpretation of the risk level
	0–0.2	0.2–0.4	0.4–0.6	0.6–0.8	0.8–1	
0	0.35	0.72	0.81	0.85	0.24	Medium
0.1	0.31	0.69	0.79	0.84	0.20	Medium
0.2	0.27	0.65	0.76	0.72	0.16	Medium
0.3	0.22	0.61	0.73	0.79	0.12	Medium
0.4	0.18	0.54	0.69	0.76	0.08	Medium
0.5	0.13	0.45	0.63	0.71	0.04	Medium
0.6	0.07	0.32	0.54	0.64	0	Medium
0.7	0	0.15	0.40	0.57	0	High
0.8	0	0.04	0.21	0.56	0	High
0.9	0	0	0.06	0.51	0	High
1	0	0	0	0.47	0	High

Table 6.

Risk assessments of failure to achieve the target indicators of the university's intellectual capital

Fuzzy variable	Normalized degree of non-achievement			Interval matching coefficients (α = 0/ α = 0.5/ α = 1)					Interpretation of the risk level (α = 0/ α = 0.5/ α = 1)
	IOF	COG	MOM	0–0.2	0.2–0.4	0.4–0.6	0.6–0.8	0.8–1	
I	0.331	0.464	0.456	0.19 / 0 / 0	0.81 / 0.65 / 0.06	0.95 / 0.92 / 0.64	0.62 / 0.34 / 0	0.08 / 0 / 0	M / M / M
I_H	0.275	0.462	0.598	0.39 / 0 / 0	0.71 / 0.54 / 0	0.81 / 0.67 / 0.21	0.57 / 0.42 / 0.21	0.07 / 0 / 0	L / M / H
I_O	0.362	0.489	0.332	0.13 / 0 / 0	0.85 / 0.71 / 0.70	0.92 / 0.89 / 0.20	0.76 / 0.55 / 0	0.10 / 0 / 0	M / M / L
I_R	0.263	0.464	0.450	0.18 / 0 / 0	0.64 / 0.36 / 0	0.88 / 0.78 / 0.36	0.46 / 0.11 / 0	0.07 / 0 / 0	M / M / M
I_{H1}	0.311	0.428	0.424	0.15 / 0 / 0	0.79 / 0.64 / 0.29	0.85 / 0.77 / 0.52	0.32 / 0.03 / 0	0.00 / 0 / 0	M / M / M
I_{H2}	0.148	0.491	0.613	0.23 / 0 / 0	0.53 / 0.09 / 0	0.56 / 0.12 / 0	0.59 / 0.31 / 0.12	0.05 / 0 / 0	M / H / H
I_{O1}	0.313	0.639	0.634	0 / 0 / 0	0.39 / 0.02 / 0	0.88 / 0.85 / 0.73	0.97 / 0.97 / 1	0.74 / 0.33 / 0.12	L / L / L
I_{O2}	0.269	0.493	0.647	0.35 / 0.13 / 0	0.72 / 0.45/ 0	0.81 / 0.63 / 0	0.85 / 0.71 / 0.47	0.24 / 0.04 / 0	M / M / H
I_{R1}	0.263	0.464	0.450	0.18 / 0 / 0	0.64 / 0.36 / 0	0.88 / 0.78 / 0.36	0.46 / 0.11 / 0	0.07 / 0 / 0	M / M / M
I_{R2}	0.206	0.611	0.604	0.03 / 0 / 0	0.37 / 0 / 0	0.65 / 0.37 / 0.36	0.72 / 0.31 / 0.46	0.57 / 0 / 0	L / L / L
I_{H21}	0.360	0.457	0.450	0.28 / 0.07 / 0	0.77 / 0.59 / 0.29	0.93 / 0.91 / 0.79	0.61 / 0.28 / 0	0.05 / 0 / 0	M / M / M
I_{H22}	0.263	0.530	0.638	0.02 / 0 / 0	0.25 / 0 / 0	0.36 / 0.08 / 0	0.64 / 0.44 / 0.23	0.53 / 0 / 0	L / L / L
I_{O21}	0.312	0.423	0.471	0.21 / 0 / 0	0.83 / 0.70 / 0	0.85 / 0.74 / 0.61	0.27 / 0.02 / 0	0 / 0 / 0	L / L / M
I_{O22}	0.306	0.379	0.375	0.14 / 0 / 0	0.78 / 0.67 / 0.67	0.62 / 0.38 / 0.38	0.07 / 0 / 0	0 / 0 / 0	L / L / L
I_{O23}	0.233	0.518	0.500	0.05 / 0 / 0	0.56 / 0.18 / 0	0.89 / 0.85 / 0.56	0.60 / 0.22 / 0	0.13 / 0 / 0	M / M / M
I_{R11}	0.321	0.403	0.397	0.19 / 0 / 0	0.86 / 0.76 / 0.51	0.80 / 0.71 / 0.45	0.23 / 0 / 0	0 / 0 / 0	L / L / L
I_{R12}	0.312	0.497	0.457	0.04 / 0 / 0	0.59 / 0.34 / 0.12	0.92 / 0.92 / 0.68	0.51 / 0.18 / 0	0.06 / 0 / 0	M / M / M
I_{R13}	0.374	0.564	0.562	0.03 / 0 / 0	0.08 / 0 / 0.4	0.71 / 0.52 / 0.37	0.34 / 0.09 / 0	0.05 / 0 / 0	L / M / M
I_{R21}	0.373	0.511	0.494	0.02 / 0 / 0	0.70 / 0.52 / 0.31	0.94 / 0.94 / 0.94	0.77 / 0.61 / 0.31	0.09 / 0 / 0	M / M / M
I_{R22}	0.277	0.333	0.342	0.34 / 0.01 / 0	0.89 / 0.82 / 0.57	0.54 / 0.25 / 0	0.08 / 0 / 0	0 / 0 / 0	L / L / L

Let us pay special attention to the indicators for which the center of gravity of the expected value exceeds the centroid of the target value (I_{O1}, I_{R2}, I_{H22}, I_{R13}). For all these indicators, at zero alpha, the risk of not achieving the target value is "Low," as well as for the other alphas (excluding only I_{R13}, for which, with an increase in alpha, the risk assessment reaches the value "Medium"). In no case did the risk assessment reach the value of "Very Low," which is apparently due to the fact that the excess of expected values over target values (by centers of gravity) was not very significant. Moreover, for the I_{R13} indicator, this excess turned out to be the smallest, which is why the risk assessment of not achieving its target value reached the "Medium" value.

4. Discussion

Analysis of the results obtained shows the following.

1. The developed method for assessing the risks of failure to achieve target indicators in the sphere of development of the intellectual capital contributes to the instrumental components of the methodology for managing the formation and development of intellectual capital, the theory of economic risk management, as well as PPRM (project portfolio risk management) as an independent area of research. Testing the method on the example of a specific organization (a large regional university) demonstrates the possibility of its practical application.

2. The proposed method is fuzzy. Its use does not require the availability of historical data and probabilistic forecasts and is therefore promising for the study of ill structured phenomena and processes (namely, the processes of formation and development of intellectual capital in the digital economy) in the context of a significant increase in the speed and unpredictability of changes in the external environment.

3. The fuzziness of the method makes it possible to take into account the decision maker's tolerance to uncertainty. This possibility is realized: at the stage of formation of the causal field of OIC indicators (when choosing fuzzy "cut-off boundaries" of explicit and implicit factors and methods of defasification); at the stage of evaluating current and determining target values of indicators (when choosing fuzzy inference systems (bases of fuzzy production rules and fuzzy inference algorithms)); at the stage of forming a OIC development program (when setting limits on risk and choosing satisfaction degrees); at the stage of quantitative assessment and qualitative interpretation of the risk level of failure to achieve the targets for the OIC development (when calculating the coefficients of correspondence of normalized deviations of expected values of indicators from the targets (namely, choosing the alpha level), as well as the possible setting of the formula for calculating the integral risk level). Differences in decision makers' tolerance for uncertainty lead to differences in assessments of the risks of failure to achieve target indicators.

4. The use of trapezoidal fuzzy numbers as membership functions of linguistic scale values ensures simplicity and transparency when conducting expert surveys. In turn, a further transition to Gaussian fuzzy numbers (and, further, arbitrary fuzzy numbers) allows us to level out the shortcomings that arise when using trapezoidal fuzzy numbers (in particular, when reducing fuzzy optimization problems to crisp Boolean quadratic programming problems).

5. The developed method is universal in the sense of possible applicability to various types of organizations of different industry affiliations (all stages of the basic scheme of the method will be standard). At the same time, the specifics of a particular organization will be manifested: in the causal field of OIC development indicators (a set of indicators of the two lower levels of the hierarchy); ways to assess current and determine target values of indicators; approaches to the formation of the OIC development program.

6. The success of the practical use of the method for a particular organization depends not so much on its type and industry affiliation as on the fulfillment of the following requirements. First of all, it is the presence of a formalized organization's strategy, in which strategic goals related to the OIC development are embedded in the hierarchical system of all strategic goals of the organization. In this case, the resulting indicators of strategic goals must be established. The organiza-

tion's management accounting system must include the ability to obtain quantitative values of the necessary indicators. Calculating the values of fuzzy variables requires appropriate software adapted to the given organization.

7. The method is also universal in the sense of its possible applicability to assessing the risks of not achieving the target values of various hierarchical indicator systems with a large number of implicit factors and "qualitative" indicators.

Conclusion

A method for assessing the risks of failure to achieve the OIC target indicators has been developed. The proposed method is based on the following fuzzy models: a fuzzy model for forming a causal field of OIC development indicators in conjunction with the organization's strategy and types of cognitive activity; a fuzzy model for assessing ICO indicators of all hierarchical levels; fuzzy models for optimizing the portfolio of ICO development projects taking into account risks.

The proposed method has the following distinctive features. The OIC indicator system is a multi-level hierarchical system with possible cycles. One of the levels of this system is formed by indicators reflecting the organization's ability to perform various types of cognitive activity linked to the main components of the OIC. At the lowest level of the hierarchy there are explicit and implicit factors of OIC development, the process of identifying which is quite flexible. When determining the current and target values of the OIC indicators, an original method of assessing the "qualitative" indicators and integral indicators of the OIC of all hierarchical levels is used, which allows us, among other things, to obtain numerical estimates of the spread (blurriness) of the calculated values of the indicators. When forming a OIC development program, it is possible to choose various fuzzy optimization models in which risk is taken into account either in the objective function or in fuzzy constraints. The scenario approach to modeling changes in internal and external conditions, which underlies fuzzy models for selecting the optimal portfolio of projects for the OIC development, allows us to obtain expected values of the OIC indicators at all levels of the hierarchy. The proposed method for obtaining additional information when calculating fuzzy deviations of expected values of indicators from target ones allows us to increase the accuracy of the qualitative interpretation of the level of risks of failure to achieve the organization's goals for the development of intellectual capital.

Another significant advantage of the proposed method is the ability to take into account the decision maker's tolerance to uncertainty. This opportunity is realized in the formation of a causal field of OIC indicators, in optimizing the portfolio of OIC development projects and, most importantly, in interpreting the risk levels of failure to achieve OIC development goals.■

Acknowledgments

The study was supported by the Russian Science Foundation (RSF) as part of research project No. 23-28-01091, https://rscf.ru/project/23-28-01091/.

References

1. Barry D. (1987) *The relationship of strategic goals and planning processes to organizational performance.* Unpublished Ph. D. Dissertation, University of Maryland.

2. Gurkov I.B. (2008) *Factors of formation and mechanisms of realization of strategic goals of Russian companies*. Report at the Economics Section of the Department of Social Sciences of the Russian Academy of Sciences. March 13, 2008 (in Russian).

3. Gurkov I.B. (2007) Integrated metrics of strategy process – an attempt of theoretical synthesis and empirical validation. *Russian Management Journal*, vol. 5, no. 2, pp. 3–28 (in Russian).

4. Mityakov S.N., Mityakov E.S. (2023) Developing the theory of economic security risks and thresholds. *The Bulletin of the Institute of Economics of the Russian Academy of Sciences*, no. 5, pp. 83–113 (in Russian). https://doi.org/10.52180/2073-6487_2023_5_83_113

5. Vasilkov Yu.V., Gushchina L.S. (2017) Risk analysis of not achievement of the objectives in case of control of the organization. *Proceedings of Voronezh State University. Series: Economics and Management*, no. 1, pp. 5–12 (in Russian).

6. Morozov V.O. (2013) Dependence formalization between level of achievement of the strategic objective and values of its indicators on the basis of sign-variable function of usefulness. *Modern Problems of Science and Education*, no. 6, pp. 457 (in Russian).

7. Kachalov R.M. (2012) *Economic risk management. Theoretical foundations and applications*. Moscow; St. Petersburg: Nestor-History (in Russian).

8. Gushchina L.S., Vasilkov Yu.V. (2017) Technique of the account of risks at planning enterprise developments. *Modern High Technologies. Regional Application*, no. 2(50), pp. 105–122 (in Russian).

9. Lapochkina V.V., Dolgova V.N., Orshanskaya Yu.O., Shkilev I.N. (2020) Assessing the risk of benchmark and additional indicators of the Science national project. *National Interests: Priorities and Security*, vol. 16, no.12, pp. 2338–2362 (in Russian). https://doi.org/10.24891/ni.16.12.2338

10. Chereshnev V.A., Vasil'eva A.V., Korobitsyn B.A. (2017) Assessing the economic efficiency of socially oriented government programs by simulation modeling methods. *Economic Analysis: Theory and Practice*, vol. 16, no.1, pp. 174–187 (in Russian). https://doi.org/10.24891/ea.16.1.174

11. Pishchalkina I., Tereshko E., Suloeva S. (2023) Application of self-organizing maps for risk assessment of mining and metallurgical enterprises. *Sustainable Development and Engineering Economics*, no. 1(7), pp. 28–44. https://doi.org/10.48554/SDEE.2023.1.2

12. Novoselova I.Yu., Novoselov A.L. (2023) Planning and implementation of Federal projects in the Arctic regions, taking into account risk factors. *Economics, Taxes & Law*, vol. 16, no. 6, pp. 49–59 (in Russian). https://doi.org/10.26794/1999-849X-2023-16-6-49-59

13. Novoselova I.Yu., Novoselov A.L. (2023) Methods of risk assessment for the implementation of economic development projects in the Arctic regions. *Economics, Taxes & Law*, vol. 16, no. 3, pp. 109–119 (in Russian). https://doi.org/10.26794/1999-849X-2023-16-3-109-119

14. Kachalov R.M., Sleptsova Y.A. (2014) Modelling of procedure regulations of economic risk with application of fuzzy logic theory. Proceedings of the *III International Youth Scientific and Practical Conference*. Saratov: Saratov State University, pp. 78–83 (in Russian).

15. Sleptsova Y.A. (2016) *Risk management in the activities of a manufacturing enterprise based on the tools of systemic economic theory and fuzzy logic*. Moscow: CEMI (in Russian).

16. Tselykh A.N., Tselykh L.A., Prichina O.S. (2014) Fuzzy logic methods in the management of production processes. *Izvestiya SFedU. Engineering Sciences*, no. 1(150), pp. 111–119 (in Russian).

17. Strebkova L.N. (2014) Risk assessment of enterprise based on application of fuzzy neural network. *Vestnik NSUEM*, no. 3, pp. 147–154 (in Russian).

18. Mazelis L.S., Solodukhin K.S., Lavrenyuk K.I. (2017) Fuzzy model of socio-economic system development risks analysis on the stakeholder approach basis. *Tyumen State University Herald. Social, Economic, and Law Research*, vol. 3, no. 3, pp. 242–260 (in Russian). https://doi.org/10.21684/2411-7897-2017-3-3-242-260

19. Dudin M.N., Lyasnikov N.V., Protsenko O.D., Tsvetkov V.A. (2017) Quantification and risk assessment of hydrocarbon resources development projects in the Arctic region. *Tyumen State University Herald. Social, Economic, and Law Research*, vol. 12, no. 4, pp. 168–195 (in Russian). https://doi.org/10.18288/1994-5124-2017-4-07

20. Solodukhin K.S. (2019) Fuzzy strategic decision-making models based on formalized strategy maps. *AEBMR-Advances in Economics, Business and Management Research*, vol. 47, Proceedings of the International Scientific Conference "Far East Con" (ISCFEC 2018), pp. 543–547. https://doi.org/10.2991/iscfec-18.2019.136

21. Micán C., Fernandes G., Araújo M. (2021) Project portfolio risk management: a structured literature review with future directions for research. *International Journal of Information Systems and Project Management*, vol. 8, no. 3, pp. 67–84. https://doi.org/10.12821/ijispm080304

22. Hofman M., Spalek S., Grela G. (2017) Shedding new light on project portfolio risk management. *Sustainability*, vol. 9, no. 10, pp. 1798-1816. https://doi.org/10.3390/su9101798

23. Ghasemi F., Sari M., Yousefi V., Falsafi R., Tamošaitienė J. (2018) Project portfolio risk identification and analysis, considering project risk interactions and using Bayesian networks. *Sustainability*, vol. 10, no. 5, pp. 1609–1631. https://doi.org/10.3390/su10051609

24. Guan D., Guo P., Hipel K., Fang L. (2017) Risk reduction in a project portfolio. *Journal of Systems Science and Systems Engineering*, vol. 26, no. 1, pp. 3–22. https://doi.org/10.1007/s11518-016-5296-2

25. Benaija K., Kjiri L. (2015) Hybrid approach for project portfolio selection taking account of resources management and interactions between projects. *Journal of Digital Information Management*, vol. 13, no. 6, pp. 451–461.

26. Kleiner G.B., Smolyak S.A. (2000) *Econometric dependencies: methods and principles of construction*. Moscow: Nauka (in Russian).

27. Kettunen J., Salo A. (2017) Estimation of downside risks in project portfolio selection. *Production and Operations Management*, vol. 26, no. 10, pp. 1839–1853. https://doi.org/10.1111/poms.12727

28. Mohagheghi V., Mousavi S.M., Vahdani B. (2016) A new multi-objective optimization approach for sustainable project portfolio selection: A real-world application under interval-valued fuzzy environment. *Iranian Journal of Fuzzy Systems*, vol. 13, no. 6, pp. 41–68. https://doi.org/10.22111/ijfs.2016.2821

29. Shatalova O.M. (2023) *Efficiency of innovative processes: fuzzy multiple modeling and evaluation in conditions of non-stochastic uncertainty*. Izhevsk: Publishing House of IzhSTU (in Russian).

30. Zaidouni A., Idrissi M.A.J., Bellabdaoui A. (2023) A Sugeno ANFIS model based on fuzzy factor analysis for IS/IT project portfolio risk prediction. *Journal of Information and Communication Technology*, vol. 23, no. 2, pp. 139–176. https://doi.org/10.32890/jict2024.23.2.1

31. Mehlawat M.K., Gupta P., Khan A.Z. (2023) An integrated fuzzy-grey relational analysis approach to portfolio optimization. *Applied Intelligence*, vol. 53, pp. 3804–3835. https://doi.org/10.1007/s10489-022-03499-z

32. Khanjani Shiraz R., Tavana M., Fukuyama H. (2020) A random-fuzzy portfolio selection DEA model using value-at-risk and conditional value-at-risk. *Soft Computing*, vol. 24, pp. 17167–17186. https://doi.org/10.1007/s00500-020-05010-7

33. Deng X., Yuan Y. (2021) A novel fuzzy dominant goal programming for portfolio selection with systematic risk and non-systematic risk. *Soft Computing*, vol. 25, pp. 14809–14828. https://doi.org/10.1007/s00500-021-06226-x

34. Rahiminezhad Galankashi M., Mokhatab Rafiei F., Ghezelbash M. (2020) Portfolio selection: a fuzzy-ANP approach. *Financial Innovation*, vol. 6, no. 1, pp. 1–34. https://doi.org/10.1186/s40854-020-00175-4

35. Mohseny-Tonekabony N., Sadjadi S.J., Mohammadi E., Tamiz M., Jones D.F. (2024) Robust, extended goal programming with uncertainty sets: An application to a multi-objective portfolio selection problem leveraging DEA. *Annals of Operations Research*. https://doi.org/10.1007/s10479-023-05811-7

36. Wang B., Li Y., Wang S., Watada J. (2018) A multi-objective portfolio selection model with fuzzy value-at-risk ratio. *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3673–3687. https://doi.org/10.1109/TFUZZ.2018.2842752

37. Dehghani F. (2023) Simulation of annealing for portfolio selection in mean-pseudo-variance fuzzy model. *SSRN Electronic Journal*, pp. 1–12. https://doi.org/10.2139/ssrn.4660934

38. Nguyen V.D., Duyen N.K., Hai N.M., Duy B.K. (2023) Multicriteria portfolio selection with intuitionistic fuzzy goals as a pseudoconvex vector optimization. *Lecture Notes on Data Engineering and Communications Technologies*, vol. 187, pp. 68–79. https://doi.org/10.1007/978-3-031-46573-4_7

39. Abtahi S.H. (2023) Uncertain random portfolio optimization based on skew chance distribution. *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 23, no. 1, pp. 44–55. https://doi.org/10.5391/ijfis.2023.23.1.44

40. Yang X., Liu W., Chen S., Zhang Y. (2021) A multi-period fuzzy mean-minimax risk portfolio model with investor's risk attitude. *Soft Computing*, vol. 25, pp. 2949–2963. https://doi.org/10.1007/s00500-020-05351-3

41. Kleiner G.B. (2008) *Enterprise strategy*. M.: Publishing house "Delo" ANKH (in Russian).

42. Nazarov D.M. (2016) The evaluation model of implicit factors on the basis of fuzzy-set descriptions. *Far Eastern Federal University News*, no. 4(80), pp. 3–17 (in Russian). https://doi.org/10.5281/zenodo.220793

43. Makarova D.V., Nedoluzhko O.V., Solodukhin K.S. (2024) The role of economic digitization in the development of the organization intellectual capital theory. Proceedings of the *X international Scientific and Practical Conference "Intelligent engineering economics and Industry 5.0 (IEEI_5.0_INPROM)", St. Petersburg, April 25–28* (eds. D.G. Rodionov, A.V. Babkin), pp. 215–219 (in Russian). https://doi.org/10.18720/IEP/2024.2/50

44. Cosa M., Pedro E., Urban B. (2024) How to assess the intellectual capital of firms in uncertain times: a systematic literature review and a proposed model for practical adoption. *Journal of Intellectual Capital*, vol. 25, no. 7, pp. 1–22. https://doi.org/10.1108/JIC-05-2023-0096

45. Bustamante A., Liberona D., Ferro R. (2024) Approach to measuring organizational performance from the perspective of intellectual capital. *Communications in computer and information science*, vol. 2152, pp. 73–85. https://doi.org/10.1007/978-3-031-63269-3_6

46. Kozlovskyi S., Syniehub P., Kozlovskyi A., Lavrov R. (2022) Intellectual capital management of the business community based on the neuro-fuzzy hybrid system. *Neuro-Fuzzy Modeling Techniques in Economics*, vol. 11, no. 11, pp. 25–47. https://doi.org/10.33111/nfmte.2022.025

47. Çevik G., Arslan Ö. Analytic evaluation of intellectual capital for ship management companies under a fuzzy environment (2022) *Journal of ETA Maritime Science*, vol. 10, no. 3, pp. 185–194. https://doi.org/10.4274/jems.2022.41033

48. Pokrovskaia N., Margulyan Ya., Lvin Yu., Bulatetskaia A. (2020) Neuro-technologies and fuzzy logic for intellectual capital evaluation in education and business. *IOP Conference Series: Materials Science and Engineering*, vol. 940, article 012090. https://doi.org/10.1088/1757-899X/940/1/012090

49. Lucchese M., Aversano N., Di Carlo F., Polcini P.T. (2020) Assessing the intellectual capital and related performance in the teaching process using FES models: First evidence in Italian universities. *WSEAS transactions on business and economics*, vol. 17, pp. 325–344. https://doi.org/10.37394/23207.2020.17.34

50. Gross-Gołacka E., Kusterka-Jefmańska M., Jefmański B. (2020) Can elements of intellectual capital improve business sustainability? – The perspective of managers of SMEs in Poland. *Sustainability*, vol. 12, no. 4, article 1545. https://doi.org/10.3390/su12041545

51. Zavalin G.S., Nedoluzhko O.V., Solodukhin K.S. (2023) Formation of the causal field of indicators for an organization's intellectual capital development: A concept and a fuzzy economic and mathematical model. *Business Informatics*, vol. 17, no. 3, pp. 53–69. https://doi.org/10.17323/2587-814X.2023.3.53.69

52. Nedoluzhko O.V., Solodukhin K.S. (2024) Quantitative assessment of university's intellectual capital based on fuzzy model. *University Management: Practice and Analysis*, vol. 28, no. 1, pp. 34–49 (in Russian). https://doi.org/10.15826/umpa.2024.01.003

53. Makarova D.V., Nedoluzhko O.V., Solodukhin K.S., Zavalin G.S. (2024) Fuzzy optimization models for intellectual capital enhancing project portfolio selection under risk. *Journal of System and Management Sciences*, vol. 14, no. 7, pp. 1–19. https://doi.org/10.33168/JSMS.2024.0701

54. Anshin V.M., Dyomkin I.V., Tsarkov I.V., Nikonov I.M. (2008) On application of fuzzy set theory to the problem of project portfolio selection. *Issues of Risk Analysis,* vol. 5, no. 3, pp. 8–21 (in Russian).

55. Dubois D., Prade H. (1988) *Possibility Theory*. New York: PlenumPress.

56. Wang J., Hwang W.-L. (2007) A fuzzy set approach for R&D portfolio selection using a real option valuation model. *Omega*, vol. 35, no. 3, pp. 247–257.

57. Furnham A., Marks J. (2013) Tolerance of ambiguity: A review of the recent literature. *Psychology*, vol. 4, no. 9, pp. 717–728. https://doi.org/10.4236/psych.2013.49102

58. Minaev Yu.N., Filimonova O.Yu., Minaeva J.I. (2012) Index of fuzziness of fuzzy sets in context of concepts "Data Mining". *Problems of Informatization and Management*, vol. 3, no. 39, pp. 95–101 (in Russian).

59. Nazarov D.M. (2016) *Methodology of fuzzy set evaluation of implicit factors in organizational activities*. Ekaterinburg: Ural State Economic University Press (in Russian).

60. De Luca A., Termini S. (1972) A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and control*, vol. 20, no. 4, pp. 301–312. https://doi.org/10.1016/S0019-9958(72)90199-4

61. Yager R.R. (1979) On the measure of fuzziness and negation. Part I: Membership in the unit interval. *International Journal of General Systems*, vol. 5, no. 4, pp. 221–229. https://doi.org/10.1080/03081077908547452

About the authors

Konstantin S. Solodukhin

Doctor of Sciences (Economics), Professor;

Professor, Mathematics and Modeling Department, Head of Laboratory, Strategic Planning Laboratory, Vladivostok State University, 41, Gogolya Str., Vladivostok 690014, Russia;

E-mail: k.solodukhin@mail.ru

ORCID: 0000-0003-3619-1219

Georgij S. Zavalin

Head of Department, Data Mining Department, Intern Researcher, Strategic Planning Laboratory, Vladivostok State University, 41, Gogolya Str., Vladivostok 690014, Russia;

E-mail: georgiy.zavalin08@vvsu.ru

ORCID: 0000-0003-4519-0242

Daria V. Makarova

Leading Specialist, Analytical Department, Intern Researcher, Strategic Planning Laboratory, Vladivostok State University, 41, Gogolya Str., Vladivostok 690014, Russia;

E-mail: malnova.daria@vvsu.ru

ORCID: 0009-0002-8207-3010

DOI: 10.17323/2587-814X.2025.1.93.107

The method for the land plot value appraisal as part of the single real estate object, based on game theory approach*

Michael B. Laskin
E-mail: laskin.m@iias.spb.su

St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, Russia

Abstract

In mass real estate valuation, in cadastral valuation, there is a problem of splitting the value of a single real estate object into the value of land plot and buildings (improvements) located on it. One of the key information sources for real estate valuation is market data. Such data may contain information on offer prices, as well as actual transaction prices (for example, in mortgage transactions) for the whole object. At the same time, in the accounting policy of enterprises different rates of land and property tax often require separate accounting of the value of land plots and the buildings located on them. The problem of such splitting of a single object's value is the subject of permanent discussions in the valuation community. There are no established methods. This article proposes a method of splitting the value of a single property object based on the approach borrowed from co-operative game theory. A simple game formulation of the problem and its fair solution based on the Shepley value are considered. Simple and well-interpretable computational formulas are obtained, which allow us to split the market value of single objects on large data sets in minimum time. The proposed method is new in the theory and practice of valuation.

* The article is published with the support of the HSE University Partnership Programme

Keywords: single real estate object, Shepley value, multiple linear regression, log-normal price distribution, property value splitting

Citation: Laskin M.B. (2025) The method for the land plot value appraisal as part of the single real estate object, based on game theory approach. *Business Informatics*, vol. 19, no. 1, pp. 93–107. DOI: 10.17323/2587-814X.2025.1.93.107

Introduction

One of the current problems of real estate valuation, especially in cadastral valuation, is the problem of splitting the estimated value of a single real estate object1 (SREO) into shares of the value of the land plot (LP) and improvements on it (buildings, structures, etc.). A detailed description of appraisal practice situations that require such a division is given in [1].

In [1], as one of the possible methods for such splitting, a method based on Shapley value is proposed. It should be noted that the application of the Shepley value in machine learning has received quite a lot of attention by researchers in various applied fields (see, for example, [2–7]). Including for estimating the factors' degree of influence in a linear regression model [8–9]. Article [10] is devoted to the application of a multiple linear regression model and Shepley value in the study of the relationship between land and buildings. Article [10] based on market dataset from the city of Montreal.

The present paper considers an approach to solving the problem of splitting the estimated value of SREO arising from the ideology of applying the Shapley value. In this case, the target variable is the estimated value of the SREO, and the pricing factors[2] are the area of land and the area of improvements on it (buildings and structures).

1. Problem statement

Let's introduce the following notations:

V is the offer (or transaction) price for the SREO;

SB is the area of improvements within the SREO;

SP is the area of the land plot within the SREO.

Suppose, that there are *n* observations in the form of three-dimensional vectors (V_i, SP_i, SB_i), $i = \overline{1,n}$ and it was possible to construct some estimated functional dependence $Y = f(sp, sb)$, in some sense best reflecting the relationship between the target variable and factors (by *Y* we will understand the estimation of the SREO value, by *sp*, *sb* – fixed values of the area of land and improvements, respectively). We will consider the estimation by the formula $Y = f(sp, sb)$ as a result of the cumulative influence of factors SP, SB at the fix values $SP = sp$, $SB = sb$, and functions $Y_1 = f_1(sp)$ and $Y_2 = f_2(sb)$ – as a result of estimates for each of the factors separately, on the same initial data as $Y = f(sp, sb)$.

It is required to create a model of the dependence of the total value of the object and to distribute the value between the land plot and improvements, applying some fairness criterion.

2. Solution method (construction of the Shepley value)

In our case, only two factors are considered. The area of the LP and the area of improvements. In terms of game theory, it is the simplest cooperative game with

[1] A single real estate object is an object which includes a land plot and buildings located on it.

[2] A linear regression model may involve more factors. This article considers two of these factors, since the problem is to split the value of SREO only between these two factors.

two participants (for game theory see, for example, [11, 12]). The Shepley value in this case represents some optimal value distribution, where the contribution of LPs and improvements to the total value is equal to the average contribution of all possible "coalitions." There are only two possible coalitions (*sp, sb*) and (*sb, sp*). Let us construct a table of the values of the "win function," the calculation of the Shepley value and the shares to be assigned to the value of LP and improvements (*Table 1*).

Table 1.

Calculation of the share of the land component and the share of buildings in the SREO valuation using the Shepley value

Coalitions / factors	*sp*	*sb*
(*sp*, *sb*)	Y_1	$Y - Y_1$
(*sb*, *sp*)	$Y - Y_2$	Y_2
Mean by coalitions (Shepley value)	$\frac{Y + Y_1 - Y_2}{2}$	$\frac{Y - Y_1 + Y_2}{2}$
LP shares/building shares	$\frac{Y + Y_1 - Y_2}{2 \cdot Y}$	$\frac{Y - Y_1 + Y_2}{2 \cdot Y}$

In order to transfer the game approach to the problem of splitting the estimated value of SREO, the following conditions must be fulfilled:

1. Efficiency.

2. Additivity across coalitions.

3. Symmetry.

4. Factors that do not affect the outcome do not participate in the model (in game terminology, the "dummy axiom").

We consider the SREO valuation, i.e., the facility has a non-zero building area and a non-zero land area. Hence, $Y_1 > 0$, $Y_2 > 0$. The inequalities $Y > Y_1$, $Y > Y_2$ ensure the fulfilment of the efficiency and additivity conditions (on coalitions). The fulfilment of the third and fourth conditions is ensured by the problem conditions: there are no identical factors in the training model; there are no factors that do not influence the result.

The fulfilment of conditions $Y_1 > 0$, $Y_2 > 0$, $Y > Y_1$, $Y > Y_2$ ensures the existence of the kernel, hence the existence of the sharing and the existence of the Shepley value as the only fair sharing.

Failure to fulfil the condition $Y > Y_1$ ($Y \leqslant Y_1$) means that the evaluation of SREO only on the first factor "area of LP" gives a value not less than the evaluation on two factors. In this case, the improvements in the SREO have negative or zero value.

Failure to fulfil the condition $Y > Y_2$ ($Y \leqslant Y_2$) (means that the assessment of SREO only by the second factor "area of improvements" gives a value not less than the assessment by two factors. In this case, the LP within the SREO[3] has a negative or zero value.

The calculation of the share of the value of land and improvements within the SREO shown in *Table 1* is independent of the model used to estimate the value of the SREO. However, the choice of model may depend on how the market data are organized. In game theory terms, this means that a separate part of the study is the selection of the characteristic function, i.e. the function that forms the winning rule.

3. Construction of the characteristic function

It should be noted that the choice of the characteristic function significantly affects the result of calculations. In some cases, it is possible to obtain a model of the form

[3] It should be noted that free (or conditionally free) land and land within the SREO are different types of real estate. For free (conditionally free) land it is assumed that there is a turnover in the real estate market. There is no turnover of land plots within the SREO as separate plots.

$$V = a + b \cdot sp + c \cdot sb \quad (1)$$

with satisfactory quality indicators.

In this case, we look for Y_1 and Y_2 in the form $Y_1 = a_1 + b_1 \cdot sp$ and $Y_2 = a_2 + c_2 \cdot sb$.

The share on LP is equal to

$$\frac{Y + Y_1 - Y_2}{2 \cdot Y} = \frac{1}{2} + \frac{a_1 - a_2 + b_1 \cdot sp - c_2 \cdot sb}{2 \cdot \left(a + b \cdot sp + c \cdot sb\right)},$$

the share on improvements is

$$\frac{Y - Y_1 + Y_2}{2 \cdot Y} = \frac{1}{2} + \frac{a_2 - a_1 - b_1 \cdot sp + c_2 \cdot sb}{2 \cdot \left(a + b \cdot sp + c \cdot sb\right)}.$$

A formalized representation in the form of a game. Suppose that two people play the following game: the first player chooses the value of *sp* and gets the payoff Y_1, the second chooses the value of sb and gets the payoff Y_2; if they join in a coalition, they get the payoff $Y = V$ for two. Conditions for the existence of 'sharing' $Y_1 > 0$, $Y_2 > 0$, $Y > Y_1$, $Y > Y_2$, depend on the signs of the coefficients of the linear regression models. How to distribute the gain, if it exists, between the players (in our subject area between the LP and the improvements)?

In real estate valuation problems, a model of the form (1) is often inapplicable due to asymmetric distributions of the values *V, sp, sb*. This means asymmetric errors distribution of the linear regression model. It was shown in [13], that the price distributions formed by successive comparisons converge to a log-normal distribution. The same fact was pointed out by the authors [14, 15]. The areas of improvements often follow the same distribution (see, for example, [16]). The same tendency is also observed for the area of land plots. It should be noted that for land plots the distribution of areas in separate market sectors may not be confirmed by statistical tests due to a large number of comparison objects of the same area (e.g., suburban typical residential settlements formed in the Soviet period). Assuming that the values *V, sp, sb* are jointly distributed log-normally[4], the linear model can be considered in the form:

$$\ln(V) = a + b \cdot \ln(sp) + c \cdot \ln(sb), \quad (2)$$

and the general model of the problem in the form:

$$V = e^a \cdot sp^b \cdot sb^c. \quad (3)$$

In this case, we look for Y_1 and Y_2 in the form

$$\ln(Y_1) = a_1 + b_1 \cdot \ln(sp) \text{ and } \ln(Y_2) = a_2 + c_2 \cdot \ln(sb) \quad (4)$$

and we obtain

$$Y = e^a \cdot sp^b \cdot sb^c;\ Y_1 = e^{a_1} \cdot sp^{b_1};\ Y_2 = e^{a_2} \cdot sb^{c_2}.$$

The share on the LP is equal to

$$\frac{Y + Y_1 - Y_2}{2 \cdot Y} = \frac{1}{2} + \frac{e^{a_1} \cdot sp^{b_1} - e^{a_2} \cdot sb^{c_2}}{2 \cdot e^a \cdot sp^b \cdot sb^c},$$

the share on improvements is equal to

$$\frac{Y - Y_1 + Y_2}{2 \cdot Y} = \frac{1}{2} - \frac{e^{a_1} \cdot sp^{b_1} - e^{a_2} \cdot sb^{c_2}}{2 \cdot e^a \cdot sp^b \cdot sb^c}. \quad (5)$$

The conditions for the existence of "sharing" $Y_1 > 0$, $Y_2 > 0$ are fulfilled, for the conditions $Y > Y_1$, $Y > Y_2$ to be fulfilled, the inequalities depending on models (3) and (4) must be fulfilled

$$e^a \cdot sp^b \cdot sb^c > e^{a_1} \cdot sp^{b_1} \text{ and } e^a \cdot sp^b \cdot sb^c > e^{a_2} \cdot sb^{c_2},$$

which on the plane (*sp, sb*) give the region bounded by the conditions

$$sb > e^{\frac{a_1 - a}{c}} \cdot sp^{\frac{b_1 - b}{c}} \text{ and } sp > e^{\frac{a_2 - a}{b}} \cdot sb^{\frac{c_2 - c}{b}}. \quad (6)$$

As will be shown below, there can also be combined models. The data can be arranged in such a way that different types of models are constructed for the values Y, Y_1, Y_2. In any case, the principle of constructing the Shepley value (*Table 1*) will not change.

[4] In any case, it is often found that asymmetrically distributed prices allow one to proceed to construct a linear model in logarithms for which the conditions of the Gauss-Markov theorem are satisfied.

When estimating the market value of shares of land and improvements within the SREO, it may well turn out that either vacant land or a house on it may be valued more highly than the SREO. Such situations are well interpreted: in the first case the building deteriorates the land compared to the free (conditionally free) land; in the second case the improvements are valued so expensively that the land (as part of the SREO) is worthless or has a negative value compared to them. In cadastral valuation, such interpretation is impossible – the cadastral appraiser is, in any case, obliged to assign some positive cadastral value to both the land and the improvements. At the same time, cadastral valuation performed by mass valuation methods should be carried out as market valuation [17].

For the purposes of cadastral valuation, we modify the values , as follows

$$Y_1 = \begin{cases} f_1(sp), \text{ if } f_1(sp) < Y \\ Y = f(sp, sb), \text{if } f_1(sp) \geq Y, \end{cases}$$

$$Y_2 = \begin{cases} f_2(sb), \text{ if } f_2(sb) < Y \\ Y = f(sp, sb), \text{if } f_2(sb) \geq Y. \end{cases} \quad (7)$$

This choice of Y_1, Y_2, will result in positive cadastral values for both land and improvements.

4. Concept of information support of cadastral offices in calculating cadastral values of land and buildings included in single real estate objects

The approach proposed above is easily implemented in such environments as Python, statistical package R. Their advantage is openness and accessibility to any user. Any appraiser (researcher) interested in applying modern methods of working with large volumes of data can independently master the necessary set of skills to obtain high-quality analytical results. Cadastral offices have their own databases with which the results of apprising can be easily interfaced, as all of them can be downloaded from specialized packages in required formats (as a rule, files with .csv extension). The practical implementation of the method proposed above can be represented by the flowchart shown in *Fig. 1.*

Fig. 1. Block diagram of the programmed code for calculating the shares of values of land and buildings included in the SREO.

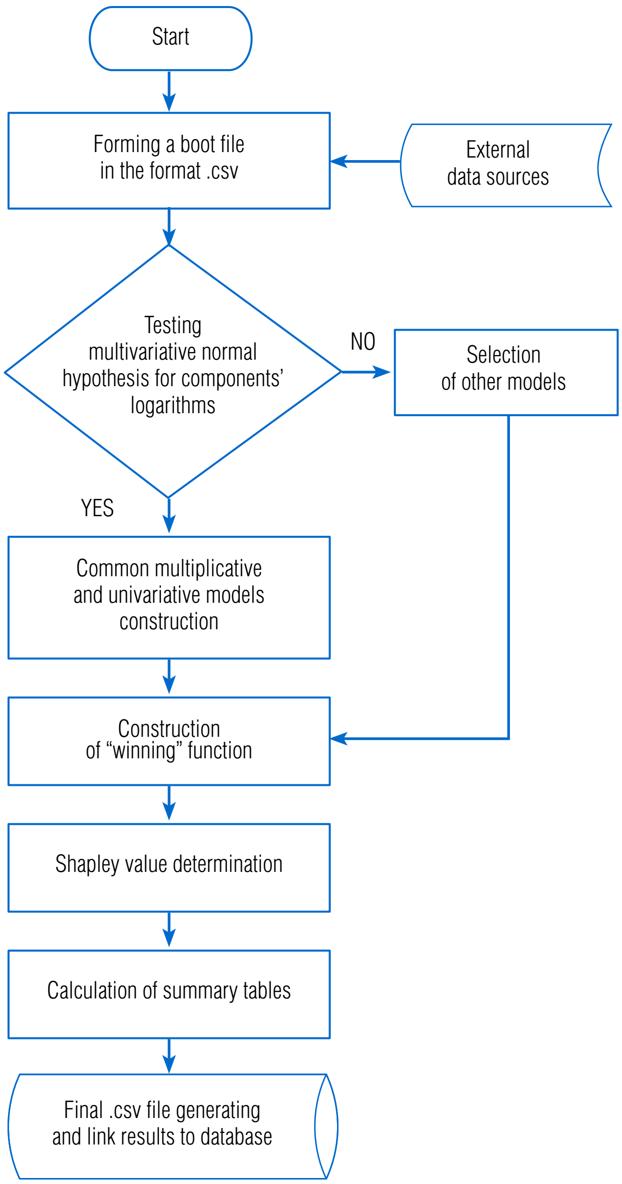


Fig. 1. Block diagram of the programmed code for calculating the shares of values of land and buildings included in the SREO.

A simple program code in an environment such as the R statistical package allows not only to calculate the cost shares for a specific object, but also to create tables (even reference tables, subject to careful selection of data by time and location) and link the results to cadastral databases.

Let us consider an example using data from the article [1] and provided by its authors (sources of primary data sites krasnodar.cian.ru, avito.ru). The total number of observations (objects of comparison) in the article [1] is 49, of which 39 objects are SREO, the remaining 10 are vacant land plots. The region is Krasnodar, and the type of permitted use is individual residential construction. The data are presented in *Table 2*. The calculations were performed in the statistical package R (readers, who are not familiar with R, can be recommended books [18, 19]).

Table 2.

Data on areas of land, buildings and prices of comparison objects

№	Land area, sq. m	Building area, sq. m	Price, million rubles	№	Land area, sq. m	Building area, sq. m	Price, million rubles
1	650.0	80.0	18.6	26	600.0	300.0	47.0
2	500.0	242.0	30.0	27	215.0	0.0	4.0
3	1120.0	720.0	78.0	28	700.0	145.0	16.0
4	580.0	300.0	65.0	29	280.0	97.1	12.5
5	120.0	62.8	6.5	30	600.0	36.0	10.0
6	580.0	49.0	3.1	31	520.0	0.0	12.0
7	800.0	370.0	125.0	32	420.0	50.0	8.8
8	500.0	0.0	15.0	33	450.0	84.0	6.3
9	215.0	123.6	7.0	34	220.0	150.0	12.0
10	640.0	260.0	16.0	35	500.0	84.0	3.9
11	383.0	43.0	9.9	36	450.0	100.0	18.5
12	600.0	50.0	7.8	37	900.0	0.0	35.0
13	616.0	149.0	18.0	38	200.0	160.0	12.0
14	707.0	0.0	8.0	39	314.0	66.3	11.0
15	400.0	90.0	10.0	40	454.0	90.0	9.5
16	300.0	90.0	6.4	41	850.0	0.0	25.0
17	600.0	160.0	19.3	42	300.0	88.0	7.8
18	2330.0	0.0	90.0	43	100.0	50.0	2.5
19	360.0	270.0	25.4	44	490.0	91.0	15.0
20	450.0	85.7	10.5	45	400.0	0.0	14.0
21	350.0	150.0	6.3	46	500.0	0.0	10.0
22	613.0	0.0	23.0	47	400.0	108.0	4.6
23	200.0	56.3	7.5	48	460.0	120.0	17.0
24	700.0	350.0	47.0	49	150.0	44.0	3.0
25	860.0	106.8	6.5				

First of all, it should be noted that vacant (conditionally vacant) land and land within the SREO are different types of real estate. For the first case, there is a market where free land is traded. For the second case, there is no market, the land within the SREO is not sold separately from the buildings and structures located on them; they can be sold only in conjunction with improvements. To build a general model $Y = f\ (sp,\ sb)$ we use data on 39 objects of SREO. To build the model $Y_1 = f_1(sp)$ we use data on vacant land; to build the model $Y_2 = f_2\ (sb)$ we use data on SREO as they have improvements.

The linear regression model on the two factors does not give a satisfactory model (*Table 3*). There is a clear asymmetry of errors and two factors of the model have unsatisfactory *t*-criterion values.

In contrast, the linear regression model built for logarithms of variables gives acceptable results (*Table 4*).

Table 3.

Results obtained using the library function *lm*() of the statistical package R for the linear regression model of the form (1)

Model: $V = a + b \cdot sp + c \cdot sb$					
Residuals:	min	1 Q	median	3 Q	max
	–26.59	–4.96	–1.15	3.91	71.34
Coefficients	Estimate	Standard error	*t*-test statistic	*p*-value *t*-test	
a	–7.191	5.677	–1.267	0.213	<0.05
b	0.015	0.014	1.075	0.289	<0.05
c	0.132	0.023	5.609	0	<0.05
RSE	14.67				
R^2	0.646	Adjusted R^2	0.626		
F-test statistic	32.81	*p*-value *F*-test	0		

Table 4.

Results obtained using the library function *lm*() of the statistical package R for the model in logarithms of the form (2)

Model: $\ln(V) = a + b \cdot \ln(sp) + c \cdot \ln(sb)$					
Residuals:	min	1 Q	median	3 Q	max
	–0.969	–0.237	0.007	0.351	0.351
Coefficients	Estimate	Standard error	*t*-test statistic	*p*-value *t*-test	
a	–4.181	0.921	–4.538	0	<0.05
b	0.425	0.170	2.500	0.017	<0.05
c	0.874	0.132	6.641	0	<0.05
RSE	0.5				
R^2	0.7	Adjusted R^2	0.682		
F-test statistic	41.83	*p*-value *F*-test	0		

For the studied real estate sector, we obtained a model (characteristic function) of the form:

$$V = e^a \cdot sp^b \cdot sb^c = e^{-4.1812} \cdot sp^{0.4246} \cdot sb^{0.8739}. \qquad (8)$$

Substituting the fixed values of *sp* and *sb* into formula (8) we obtain an estimate of the cost of the SREO.

To build the model by area of the SREO only, we select 10 free SREOs from the original data set. For this set we manage to build a satisfactory linear regression model (*Table 5*).

We obtain

$$a_1 = -7.5822,\ b_1 = -0.0413,$$
$$Y_1 = -7.5822 + 0.0413 \cdot sp.$$

In order to build a model by improvement area only, we use only those properties that have buildings (there are 39 of them). For this set we can build a satisfactory regression model only for logarithms (*Table 6*).

Table 5.

Results obtained using the library function *lm*() of the statistical package R for a univariate linear regression model of the form (4)

Model: $V = a_1 + b_1 \cdot sp$					
Residuals:	min	1 Q	median	3 Q	max
	–13.642	–2.393	1.343	4.49	5.38
Coefficients	Estimate	Standard error	*t*-test statistic	*p*-value *t*-test	
a_1	–7.582	3.243	–2.338	0.047	<0.05
b_1	0.041	0.003	11.867	0	<0.05
RSE	6.117				
R^2	0.947	Adjusted R^2	0.941		
F-test statistic	143.2	*p*-value *F*-test	0		

Table 6.

Results obtained using the library function *lm*() of the statistical package R for the univariate model in logarithms of the form (4)

Model: $\ln(V) = a_2 + c_2 \cdot \ln(sb)$					
Residuals:	min	1 Q	median	3 Q	max
	–0.957	–0.398	0.029	0.362	1.104
Coefficients	Estimate	Standard error	*t*-test statistic	*p*-value *t*-test	
a_2	–2.344	0.594	–3.945	0	<0.05
c_2	1.026	0.125	8.233	0	<0.05
RSE	0.532				
R^2	0.647	Adjusted R^2	0.637		
F-test statistic	67.78	*p*-value *F*-test	0		

We obtain

$$a_2 = -2.3443,\ c_2 = 1.0262,\ Y_2 = e^{-2.3443} \cdot sb^{1.0262}.$$

The share on land is $\frac{Y + Y_1 - Y_2}{2}$, the share on building is $\frac{Y - Y_1 + Y_2}{2}$.

Let $sb = 300$, $sp = 600$, then we get:

$$Y = V = e^{-4.1812} \cdot 600^{0.4246} \cdot 300^{0.8739} = \\ = 35.833 \text{ (million rubles)},$$

$$Y_1 = -7.5822 + 0.0413 \cdot 600 = 17.198 \text{ (million rubles)},$$

$$Y_2 = e^{-2.3443} \cdot 300^{1.0262} = 33.412 \text{ (million rubles)}.$$

The share for land is equal to

$$\frac{Y + Y_1 - Y_2}{2 \cdot Y} = \frac{35.833 + 17.198 - 33.412}{2 \cdot 35.833} \approx 27\%,$$

the share for the building is equal to

$$\frac{Y - Y_1 + Y_2}{2 \cdot Y} \approx 73\%$$

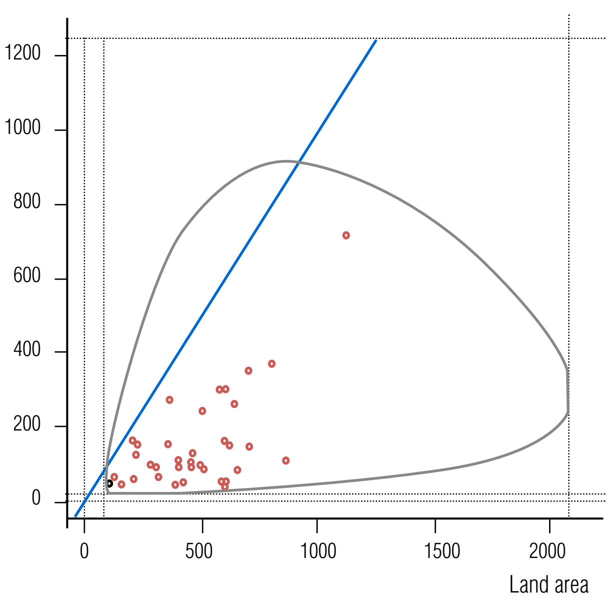


Fig. 2. Scattering diagram of the observed pairs (SP_i, SB_i), $i = \overline{1,n}$.

or in rubles 35.833 · 27% ≈ 9.675 (million rubles) for the land and 35.833 · 73% ≈ 26.158 (million rubles) for the improvements.

Thus, the total appraised value of the SREO with a land plot of 600 square meters and a house on it with an area of 300 square meters is RUR 35.833 million, of which RUR 9.675 million should be attributed to the value of the land plot and RUR 26.158 million to the value of the house. The shares to be allocated to the land and improvements for any values of *sb*, *sp* can be calculated in a similar way. *Table 2* shows the calculated shares of the value of the land within the SREO at different *sb*, *sp*. values. The improvement shares are the difference between 100% and the values in *Table 7*.

In *Table 7*, the blank fields correspond to two cases:

- the value of improvements dominates the value of LP to such an extent that the LP within the SREO has an estimated negative value (bottom left corner of the table);
- the value of the LP dominates the value of the improvements to such an extent that the improvements reduce the value of the LP within the SREO compared to the value of the free LP (upper right corner of the table).

5. Additional model justification and area of factor variation

The model of joint influence of factors on the value of the SREO of the form $V = e^a \cdot sp^b \cdot sb^c$ was constructed as a multiple linear regression model in logarithms. To construct it, it was sufficient to logarithm the factors and the target variable and fit the regression equation. Overall, the model has satisfactory statistical performance (*Table 4*), but has a standard deviation of $RSE = 0.4979$, indicating a noticeable scatter in the observations (V_i, SP_i, SB_i), $i = \overline{1,n}$. Checking the logarithms of the observations for joint normality can provide further argument in favor of this model (on joint normality see, for example, [20]). *Table 8* shows the result of the Mardia test (*MVN* library of the R statistical package).

Table 7.

Shares of the value of land within the SREO

Building area, sq. m	Land area, sq. m							
	200	400	600	800	1000	1200	1400	1600
20	**15%**							
60		68%						
100		**41%**	75%					
140		**28%**	**56%**	76%	93%			
180		**20%**	**44%**	**62%**	76%	88%	99%	
220		**15%**	**36%**	**52%**	**65%**	76%	85%	94%
260		**10%**	**30%**	**45%**	**57%**	**66%**	75%	83%
300		7%	**26%**	**40%**	**50%**	**59%**	**67%**	**74%**
340		4%	**22%**	**35%**	**45%**	**54%**	**61%**	**68%**
380		2%	**19%**	**32%**	**41%**	**49%**	**56%**	**63%**
420			**17%**	**29%**	**38%**	**46%**	**52%**	**58%**
460			**15%**	**26%**	**35%**	**43%**	**49%**	**54%**
500			**13%**	**24%**	**33%**	**40%**	**46%**	**51%**
540			**11%**	**22%**	**30%**	**37%**	**43%**	**48%**
580			**10%**	**20%**	**29%**	**35%**	**41%**	**46%**
620			8%	**19%**	**27%**	**33%**	**39%**	**44%**
660			7%	**17%**	**25%**	**32%**	**37%**	**42%**
700			6%	**16%**	**24%**	**30%**	**35%**	**40%**
740			5%	**15%**	**23%**	**29%**	**34%**	**39%**
780			4%	**14%**	**21%**	**27%**	**33%**	**37%**

The *MVN* library of the R statistical package (see [21, 22] for a detailed description), in addition to the Mardia test, contains other tests of joint normality, such as the Royston, Henze-Zirkler, Dornick-Hansen and other tests, which for these data also give positive results. Thus, we have acceptable justifications for the choice of the model, albeit with noticeable standard errors (spread) of the data. Similar validations can be obtained for paired observations (SP_i, SB_i), $i = \overline{1,n}$..

Regression models are widely used in the estimation literature (see, e.g., [23–25]). The author of [26] points out the reasons why it is not recommended to extend regression models beyond the domain of observations. *Figure 2* shows the observed 39 values of the pairs: the area of LP is the area of improvements.

The curved line in *Fig. 2* is the 90% level line of the model log-normal distribution of areas of land and areas of improvements. The straight line is the bisec-

Table 8.

Mardia test result on joint normality of logarithms of initial data

Multivariative normality test Mardia

Test:	Statistic	*p*-value	Result (YES – «+», NO – «–»)
Mardia skewness	16.31	0.091	YES
Mardia kurtosis	–0.458	0.647	YES
Mardia MVN			YES

Univariate normality	Components	Statistic	*p*-value	Result
Lilliefors (Kolmogorov-Smirnov)	Component 1	0.121	0.161	YES
Lilliefors (Kolmogorov-Smirnov)	Component 2	0.131	0.087	YES
Lilliefors (Kolmogorov-Smirnov)	Component 3	0.11	0.27	YES

Sample description

Component numbe	Sample size	Mean	Standard error	Median	Min	Max	1 Q	3 Q	Skewness	Kurtosis
1	39	2.497	0.884	2.351	0.912	4.828	1.909	2.904	0.621	0.139
2	39	6.018	0.537	6.109	4.605	5.726	5.727	6.397	–0.752	0.129
3	39	4.718	0.692	4.575	3.584	4.288	4.288	5.043	0.569	–0.259

tor of the first coordinate angle. Above the bisector are cases where the area of improvements is larger than the area of the LP. As a rule, it is assumed that for private houses the area of improvements does not exceed the area of land, but such cases are possible and, as can be seen from *Fig. 2*, the constructed model allows it. However, in our two-dimensional case, the application of the model is better restricted to the area inside the closed curve in *Fig. 2.* A comparison of *Table 7* and *Fig. 2* shows that the shares of the value of LPs are calculated for this area. At the same time, *Table 7* leaves blank fields that have an interpretation (see above) in terms of determining the shares of land and improvements in the market value of the SREO. In cadastral valuation, for any combination of area of land (*sp*) and improvements (*sb*), the cadastral value should be reported as a positive value and, in general, the cadastral value should be estimated as market value or close to it. It is this understanding of cadastral value that minimizes possible claims against cadastral valuation. What to do? In this case, we should set Y_1 and Y_2 so that it does not exceed Y for any combination of areas of land (*sp*) and improvements (*sb*). For example,

$$Y = V = e^{a} \cdot sp^{b} \cdot sb^{c} = e^{-4.1812} \cdot sp^{0.4246} \cdot sb^{0.8739},$$

Table 9.

Share of the value of land within the EON for cadastral purposes

		Land area, sq. m							
		200	400	600	800	1000	1200	1400	1600
Building area, sq. m	20	**17%**	61%	67%	71%	74%	76%	77%	78%
	60	7%	54%	61%	66%	69%	71%	73%	74%
	100	4%	**41%**	58%	63%	66%	69%	71%	72%
	140	3%	**31%**	**56%**	61%	65%	67%	69%	71%
	180	3%	**25%**	**44%**	**59%**	63%	66%	68%	70%
	220	2%	**21%**	**36%**	**52%**	**62%**	65%	67%	69%
	260	2%	**18%**	**30%**	**45%**	**57%**	**64%**	66%	68%
	300	2%	16%	**26%**	**40%**	**50%**	**59%**	**65%**	**67%**
	340	1%	14%	**23%**	**35%**	**45%**	**54%**	**61%**	**67%**
	380	1%	13%	**21%**	**32%**	**41%**	**49%**	**56%**	**63%**
	420	1%	12%	**19%**	**29%**	**38%**	**46%**	**52%**	**58%**
	460	1%	11%	**18%**	**26%**	**35%**	**43%**	**49%**	**54%**
	500	1%	10%	**16%**	**24%**	**33%**	**40%**	**46%**	**51%**
	540	1%	9%	**15%**	**22%**	**30%**	**37%**	**43%**	**48%**
	580	1%	9%	**14%**	**20%**	**29%**	**35%**	**41%**	**46%**
	620	1%	8%	14%	**19%**	**27%**	**33%**	**39%**	**44%**
	660	1%	8%	13%	**17%**	**25%**	**32%**	**37%**	**42%**
	700	1%	7%	12%	**16%**	**24%**	**30%**	**35%**	**40%**
	740	1%	7%	12%	**15%**	**23%**	**29%**	**34%**	**39%**
	780	1%	7%	11%	**14%**	**21%**	**27%**	**33%**	**37%**

$$Y_1 = \begin{cases} -7.5822 + 0.0413 \cdot sp, \\ \quad \text{if } -7.5822 + 0.0413 \cdot sp < Y \\ Y, \text{if } -7.5822 + 0.0413 \cdot sp \geq Y, \end{cases}$$

$$Y_2 = \begin{cases} e^{-2.3443} \cdot sb^{1.0262}, \text{if } e^{-2.3443} \cdot sb^{1.0262} < Y \\ Y, \text{if } e^{-2.3443} \cdot sb^{1.0262} \geq Y. \end{cases}$$

Such conditions are easily realized in the script of the statistical package R and for them the shares of land and improvements in the SREO and their monetary expressions can be calculated. The table of shares for the cadastral value of land within the SREO is given in *Table 9*.

The fractions of land (similarly, the fractions of improvements as an addition up to 100%) in *Table 9* meet the objectives of cadastral valuation — the values of all land and buildings located on them will be positive, in total coinciding with the estimated model value of the

SREO. The fractions of land in *Table 9* approximately correspond to the fractions of land in *Table 7* in the main, most important part of the tables corresponding to the area of observations in *Fig. 2* (in bold), i.e., the estimation of the market value share in these fields approximately corresponds to the cadastral one. The discrepancies become apparent as one shifts to the lower left corner and to the upper right corner. These are the areas where the cadastral value is obliged to assign a positive cadastral value in any case, even if the assigned cadastral value differs from the market value share estimate. The bottom left decreases the share of the value of the land within the SREO, the top right decreases the share of the value of the improvements within the SREO.

Conclusion

Land plots within the SREO and vacant (or conditionally vacant) land plots belong to different types of real estate. Land plots within the SREO are not traded on the market without improvements located on them. Their market value can be obtained only as a result of the SREO value sweep. Free land plots are traded on the real estate market; for them objects of comparison can be selected and their market value can be estimated both by the comparative and income approaches.

The Shepley value allows to determine a fair distribution of shares of the value of land and improvements within the SREO.

When estimating the market value of the shares of land and improvements within the SREO, negative values of land and improvements may be obtained. In the first case, the value of the improvements is greater than the value of the SREO (the land within the SREO has a negative value), in the second case, the value of the land is greater than the value of the SREO (the buildings deteriorate the land plot within the SREO compared to the free land plot).

When determining the cadastral value, the land and improvements on it as part of the SREO should have a positive value. The proposed methodology, based on the Shepley value and the appropriate selection of the characteristic function, allows us to split the value of SREO into the values of land and improvements, basically corresponding to the market value. Discrepancies with the market value appear only in rare cases away from the area of observation and only due to special requirements imposed on the cadastral value (base for taxation and accounting). This is a case that illustrate the differences between cadastral and market values.

The proposed method allows us to obtain data from cadastral databases, external sources, performing calculations in a specialized environment and uploading the results of calculations in formats easily linked to cadastral databases. ■

References

1. Kuznetsov D.D., Myagkov V.N. (2023) Land – separately, building – separately. How to solve the problem of unfolding the market value of the pledged object. *Bank Lending*, no. 6(112), pp. 60–76 (in Russian).

2. Bure V.M., Staroverova K.Yu. (2019) Applying cooperative games with coalition structure for data clustering. *Automation and Remote Control*, vol. 80, no. 8, pp. 1541–1551.

3. Rozemberczki B., Watson L., Bayer P., et al. (2022) The Shapley value in machine learning. *arXiv:2202.05594v2*. https://doi.org/10.48550/arXiv.2202.05594

4. Roth J., Bajorath J. (2024) Machine learning models with distinct Shapley value explanations decouple feature attribution and interpretation for chemical compound predictions. *Cell Reports Physical Science*, vol. 5, no. 8, pp. 102–110. https://doi.org/10.1016/j.xcrp.2024.102110

5. Li M., Sun H., Huang Y., Chen H. (2024) Shapley value: from cooperative game to explainable artificial intelligence. *Autonomous Intelligent Systems*, vol. 4, article 2. https://doi.org/10.1007/s43684-023-00060-8

6. Feretzakis G., Sakagianni A., Anastasiou A., et al. (2024) Integrating Shapley values into machine learning techniques for enhanced predictions of hospital admissions. *Applied Sciences*, vol. 14(13), article 5925. https://doi.org/10.3390/app14135925

7. Kjersti A., Jullum M., Loland A. (2020) Explaining individual predictions when features are depended: More accurate approximations to Shapley values. *arXiv:1903.10464v3*. https://doi.org/10.48550/arXiv.1903.10464

8. Mikheenko A.M., Savich D.S. (2020) Application of Shepley value in regression analysis. *Vestnik of I. Kant Baltic Federal University. Ser.: Physic-mathematical and technical sciences*, no. 2, pp. 84–94 (in Russian).

9. Lundberg S.M., Allen P.G. (2017) A unified approach to interpreting model predictions. *ArXiv:1705.07874v2*. https://doi.org/10.48550/arXiv.1705.07874

10. Özdilek Ü. (2020) Land and building separation based on Shapley values. *Palgrave Communications*, vol. 6, article 68. https://doi.org/10.1057/s41599-020-0444-1

11. Petrosyan L.A., Zenkevich N.A., Semina E.A. (1998) *Game Theory*. Moscow: Higher School (in Russian).

12. Mazalov V.V. (2010) *Mathematical theory of games and applications*. St. Petersburg–Moscow–Krasnodar: Lan (in Russian).

13. Rusakov O., Laskin M., Jaksumbaeva O. (2016) Pricing in the real estate market as a stochastic limit. Log Normal approximation. *International Journal of the Mathematical models and methods in applied sciences*, vol. 10, pp. 229–236.

14. Aitchinson J., Brown J.A.C. (1963) *The Lognormal distribution with special references to its uses in economics*. Cambridge: University Press.

15. Ohnishi T., Mizuno T., Shimizu C., Watanabe T. (2011) On the evolution of the house price distribution. *Understanding Inflation Dynamics of the Japanese Economy*, Working Paper Series no. 56.

16. Laskin M.B. (2017) Adjustment of the market value by the pricing factor 'area of the object'. *Property relations in the Russian Federation*, no. 8(191), pp. 86–99 (in Russian).

17. Rosreestr Order No. N/0336 dated 08/04/2021 (as amended on 09/11/2024) "*On Approval of Methodological Guidelines on State Cadastral Valuation*" (Registered with the Ministry of Justice of Russia on 12/17/2021 N 66421). Available at: https://www.consultant.ru/document/cons_doc_LAW_403900/ (accessed 30 January 2025) (in Russian).

18. Kabakov P. (2014) *R in action. Data analysis and visualization in the R language*. Moscow: DMK-Press (in Russian).

19. Zaryadov I.S. (2010) *Introduction to statistical package R: types of variables, data structures, reading and writing information, graphics*. Moscow: PUDN Publishing House (in Russian).

20. Feller W. (1971) *An introduction to probability theory and its applications, vol. 2.* John Wiley & Sons; 2nd edition.

21. Korkmaz S., Goksuluk D., Zararsiz G. (2014) MVN: An R package for assessing multivariate normality. *The R Journal*, vol. 6(2), pp. 151–162. https://doi.org/10.32614/RJ-2014-031

22. Korkmaz S., Goksuluk D., Zararsiz G. (2019) *MVN: An R package for assessing multivariate normality*. Trakya University, Faculty of Medicine, Department of Biostatistics, Edirne, TURKEY, MVN version 5.8 (Last revision 2019-09-27). Available at: https://cran.r-project.org/web/packages/MVN/vignettes/MVN.pdf (accessed 30 January 2025).

23. Puzynya N.Y. (2020) *Modern market trends and market value assessment: collective monograph*. St. Petersburg: Izd-vo Sankt-Peterburgskogo universiteta (in Russian).

24. Barinov N.P. (2022) Application of regression analysis methods in the tasks of individual and mass valuation of real estate objects. *Valuation Issues*, no. 1(106), pp. 34–46 (in Russian).

25. Anisimova I.N., Barinov N.P., Gribovsky S.V. (2004) Accounting for different types of price-forming factors in multidimensional models of real estate valuation. *Valuation Issues*, no. 2, pp. 2–15 (in Russian).

26. Barinov N.P. (2019) Calculation of the uncertainty interval of value estimation by methods of the comparative approach. *Valuation Issues*, no. 4(98), pp. 2–10 (in Russian).

About the author

Michael B. Laskin

Doctor of Sciences (Economics), Candidate of Sciences (Physics and Mathematics), Associate Professor;

Chief Scientist, St. Petersburg Federal Research Center of the Russian Academy of Sciences, 39, 14th Line V.O., St. Petersburg 199178, Russia;

Professor, Department of Information Systems in Economics, St. Petersburg State University, 7–9 Universitetskaya Embankment, St Petersburg 199034, Russia;

E-mail: laskinmb@yahoo.com

ORCID: 0000-0002-0143-4164