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

Vol. 19 No. 2 – 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, build. 4, Shablovka 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.2025.2.7.24

Product matching in digital marketplaces: Multimodal model based on the transformer architecture

Artem Yu. Varnukhov
E-mail: varnuhov_ayu@usue.ru

Dmitry M. Nazarov
E-mail: slup2005@mail.ru

Ural State University of Economics, Yekaterinburg, Russia

Abstract

In this paper we analyze the problem of intelligent product matching in digital marketplaces for which one requires evaluation of similarity of various records that describe products but may differ in format, content or volume of multimodal data. The subject area of this scientific research represents an intersection of entity resolution (ER) problem solving methods: record matching and multimodal data analysis. It is of extreme relevance in a fast-growing platform economy with the e-commerce market expanding exponentially. The main purpose of this research is to develop and test an intelligent multimodal model based on transformer architecture to improve the accuracy and robustness of product matching in digital marketplaces. The authors developed a model integrating textual, visual and tabular attributes which enables us to identify similar products, find competitive offers, detect duplicates and perform product clustering and segmentation in a more effective manner. The proposed approach is based on the self-

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

attention mechanism which enables contextual-semantic relations modeling of various-nature data. In order to extract the vector representation of text descriptions, language models are applied, in particular the Sentence-BERT architecture; for the graphical component Vision Transformer is used; and tabular data are processed using specialized learning mechanisms based on TabTransformer structured data. The experiment we carried out demonstrated that the developed multimodal model efficiently solves the task of product matching in digital marketplaces in an environment of significant variability of product items and data heterogeneity. Additionally, the results suggest that the model can be adapted successfully for application in other product categories. The results obtained confirm the efficiency and expediency to apply the multimodal approach for digital marketplace product matching implementation. This allows the e-commerce market participants to significantly improve the quality of inventory management, increase pricing efficiency and strengthen their competitive advantages.

Keywords: digital marketplace, contextual-semantic identification, competitive offers search, product matching, machine learning, deep learning, transformer architecture, data mining

Citation: Varnukhov A.Yu., Nazarov D.M. (2025) Product matching in digital marketplaces: Multimodal model based on the transformer architecture. *Business Informatics*, vol. 19, no. 2, pp. 7–24.
DOI: 10.17323/2587-814X.2025.2.7.24

Introduction

The rapid development of the platform economy in recent decades has been driven by the emergence and widespread adoption of digital platforms serving as intermediaries between buyers and sellers. Marketplaces as digital platforms enable sellers to access a broad audience, while providing buyers with convenient services for searching and comparing product offers from various sellers. At the same time, managing product data and matching have become a complex scientific challenge the resolution of which inextricably affects both the direct and indirect economic performance indicators of all participants in e-commerce platforms. Direct indicators include factors influencing the choice of pricing model, while indirect indicators involve such aspects as the effectiveness of product promotion, accuracy of recommendations and consumer satisfaction. It is also evident that these direct and indirect indicators are causally linked. The relevance of this research is determined by the growing volume of e-commerce markets and expanding assortment of goods on digital platforms, which create a need for more advanced mathematical and instrumental methods in economics capable of automated and even intelligent and reliable product matching within a large flow of heterogeneous information in digital marketplaces.

From a scientific point of view, the issue of product matching (identification of product records) is closely related to the task of entity resolution (hereinafter referred to as ‘ER’), which is a process of aligning entities while accounting for potential duplicates. However, in the context of marketplaces, this problem is extended by the use of multimodal data in the product description. Indeed, product descriptions in marketplaces generally include not only textual information but also images and characteristics presented in a tabular form.

Below is a brief review of reference literature divided into two logical sections: personalization with regard to the specifics of e-commerce, and ER in the context of algorithmic approaches to matching.

A number of studies representing the scientific foundation of this research focus on analyzing user behavior, developing recommendation systems and evaluating the features of product categories in e-commerce (H. Angermann, M. Mao, F.T. Abdul Hussien, A. Fletcher, P. Ristoski, M. Cheung, et al.). Most researchers emphasize the importance of an integrated approach application to process the textual descriptions, images and metadata in order to improve recommendation accuracy and detect hidden relationships within large datasets (big data). Analysis of these studies in the context of matching highlights the need to apply multimodal data processing technologies to accurately match not only product titles and characteristics but also their visual attributes.

Another line of research within the subject area of this study involves classical tasks such as duplicate record detection, entity matching and the development of algorithms for accurate data integration (W.W. Cohen, S.S. Aanen, J. Devlin, A.K. Elmagarmid, N. Reimers, J. Wang, H. Köpcke, et al.). The works of these authors examine various similarity metrics, including Levenshtein distance and its analogs, language models and case-based learning methods enabling the construction of patterns for identification of potential matches. Additionally, the issue of scalability of solutions is often raised, from the development of batch-processing software to high-load systems processing dynamic data streams in real time. The rapid advancement of deep learning architectures, particularly those based on transformers caused a shift from automated feature search to intelligent methods detecting complex patterns. These approaches are especially relevant for digital marketplaces, where the promptness and reliability of processing large-scale data with various economic properties, including price dynamics represent crucial factors.

Thus, the existing research confirms the importance of applying advanced economic and mathematical methods in the development of product matching systems in marketplaces based on multimodal data processing. However, several questions remain open in the reviewed literature: how to optimally combine features, which approaches are most effective for analyzing multimodal data, and how to adapt the developed models to the new products and data sources. This study proposes a multimodal model based on a transformer architecture that can process data from various sources and enables the sequential and efficient integration of different modalities through an attention mechanism and multi-level representations for more accurate product matching. The approach proposed contributes to the mathematical and instrumental methods in economics aimed at building robust systems for big data analysis, while also supporting scalability and accommodating the high dynamism of e-commerce platforms. The scientific contribution of this paper includes the development and validation of an approach that allows for the simultaneous integration of textual, visual and tabular data within a unified modular system based on a transformer architecture.

The aim of this study is to develop and validate such a model that ensures high accuracy and robustness in product matching across large datasets of product listings, which are typically found in modern digital marketplaces.

The article is divided into four sections. The first section is devoted to the potential and applicability of product matching on digital marketplaces. The second section analyzes the existing approaches and formulates the product matching task for digital marketplaces. The third section presents the development of a multimodal product matching model for digital marketplaces (MMMP). The fourth section discusses the application of the model for evaluating product similarity in the Wildberries digital marketplace.

1. Potential and applicability of product matching in digital marketplaces

Product matching refers to the process of aligning (linking) products to determine which items are identical or essentially similar, ensuring that the same product is correctly recognized even if it appears under different titles, descriptions, or identifiers. For example, when equivalent models of a smartphone from the same brand are offered by different sellers in marketplaces, their specifications, images and even the naming format may vary significantly.

Product matching is a fundamental component widely used in various modern digital ecosystems, including marketplaces, classified ad platforms, e-commerce services and other online platforms. From the buyer's perspective, product matching significantly enhances the shopping experience by enabling quick and easy comparison of similar product offerings from different sellers [1]. For example, by implementing a product matching system that consolidates similar listings into a unified and well-structured collection of comparable offers, buyers can be spared the need to look through numerous variations within the same product category. Such a system would protect buyers from encountering duplicate or misleading offers that often make product selection a confusing and time-consuming task. Instead, buyers would be able to quickly compare prices, reviews, and seller ratings, which ultimately improves search relevance, enhances overall shopping satisfaction and supports more informed and rational decision-making. Moreover, product matching can also be used to generate personalized recommendations based on the analysis of consumer behavior [2], preferences and past purchases, helping to discover the most suitable products, increasing commitment and enhancing the overall value of the platform.

From the seller's perspective, product matching helps analyze competitive offers [3], enabling them to adjust and develop their pricing strategies based on real-time market trends [4]. Digital marketplaces accumulate vast amounts of product data; however, without proper processing and structural organization, much of this information remains fragmented, inconsistent and difficult to analyze effectively. For instance, many sellers list the same product on platforms with slight variations in the name, images, description, or characteristics. Without product matching, sellers are forced to manually track and compare thousands of listings, which is a highly time-consuming and costly process prone to errors. Automated identification and linking of similar products make it possible to consolidate scattered information into a unified coherent data repository which becomes a valuable source for further analysis and decision-making based on the current market environment. Possessing such a data source enables the application of algorithmic pricing models that rely on market assessments, as they allow for real-time monitoring of competitor prices and market demand fluctuations. This opens the door to dynamic pricing strategies. As a result, sellers can adapt their offerings in real time according to consumer behavior and the existing market environment, ensuring that product prices remain competitive and attractive to buyers without affecting margins. Such flexibility enables the identification of emerging trends and allows for proactive responses to seasonal fluctuations, promotional activities and economic shifts, ultimately contributing to long-term and sustainable development in the highly competitive digital platform market. Product matching also plays a crucial role in advertising and marketing campaigns. One of its main advantages in this context is the ability to optimize advertising costs. By correctly identifying similar products offered by other sellers, unnecessary competition can be avoided, helping to eliminate wasted expenditures and redirect the budget toward more profitable niches. In addition to advertising, product matching also enhances cross-selling opportunities. By intelligently linking complementary, frequently purchased, or otherwise related items, sellers can create more attractive offers for consumers.

Another important application of product matching algorithms lies in their role in fighting fraud and detecting counterfeit goods, thereby assisting platform operators in maintaining trust, safety and the integrity of ecosystems [5]. With the rapid development and growing influence of e-commerce platforms on public life, incidents involving fraudulent listings, counterfeit items, and deliberately misleading products are becoming increasingly common, posing a serious challenge to the industry. Such illicit activities not only harm end consumers but also undermine overall trust in digital platforms and the business of law-abiding sellers [6]. The need for counterfeit detection is especially crucial in many product categories, such as electronics, pharmaceuticals and cosmetics. In these segments, counterfeit goods can pose not only financial risks but also direct threats to consumer health and safety. In this context, product matching algorithms serve as a vital tool for preventing the appearance of listings and offers from unauthorized resellers or the sale of low-quality counterfeit products disguised as reputable brands.

Based on the above, it is reasonable to conclude that the potential and capabilities of product matching go well beyond the scope of a simple tool for solving a single task. It can be stated that product matching is a fundamental technology that enables the optimization of various business processes, supports fraud prevention and enhances the efficiency of pricing strategies which ultimately contributes to the creation of a more transparent and user-friendly ecosystem for all the participants in the e-commerce market, including both consumers and sellers.

2. Analysis of existing approaches and formulation of product matching problem in marketplaces

2.1. Product matching based on attribute value similarity and set-theoretic methods

Product attribute matching based on value similarity can be referred to fundamental matching approaches relying on comparison of textual and numerical fields across various characteristics to determine how closely they align [7]. Textual attributes are typically assessed using metrics such as Levenshtein distance, Jaro-Winkler distance or TF-IDF (Term Frequency – Inverse Document Frequency), which allows for determination of how similar two text fragments are while accounting for differences in spelling [8]. Numerical attributes (e.g., price, weight, dimensions, etc.) are usually evaluated by calculating absolute or relative deviations. Once individual similarity metrics are created and computed for each attribute, they are aggregated (commonly through a weighted sum) to produce an overall similarity evaluation between products being compared. If the indicator obtained exceeds a predefined threshold, the products are considered similar. Additionally, in the set-theoretic approach each product listing is designed as a set of atomic elements (of features, n-grams, tokens, etc.). Similarity between items is determined using classical metrics such as the Jaccard, Sørensen–Dice, or the Simkovich–Simpson index. Due to the computational simplicity, this approach provides high speed and easy scalability. However, it ignores word order and context, which significantly limits its accuracy. Key advantages of this approach include its relative simplicity, transparency and ease of implementation. However, its main limitations lie in handling large and diverse product categories, dealing with low-quality or noisy data and its severely restricted ability to capture semantic meaning.

2.2. Product matching based on a rule-based expert system

A natural extension of the attribute value comparison approach is the development of rule-based systems [9, 10]. Rule-based product matching is built using expert-defined logical constructs that describe the functioning of a specific domain and help determine the degree of similarity between products being assessed. Each rule typically evaluates a subset of well-defined attributes and applies simple logical conditions or threshold values (e.g., "If the brand name is identical and the Levenshtein distance between model names is less than two, then the two products can be considered similar"). Since this approach relies on system of rules directly encoding the expert knowledge, it is generally quick and easy to understand. However, when applied to broad and complex product categories, the rule system can become cumbersome and difficult to maintain. It also requires constant manual updates and enhancement. Moreover, because each rule must be determined manually by domain experts, the system is prone to human error and ultimately turns out to be insufficiently adaptive and excessively overloaded.

2.3. Product matching based on taxonomies and ontologies

These methods rely on in-depth contextual and semantic analysis of domain-specific information and use structured representations to map relationships between products and their attributes in the form of a knowledge graph [11, 12]. These relationships can indicate shared characteristics (e.g., belonging to a certain brand or product family), hierarchical inheritance or descriptive associations. Modeling data through a dependency graph allows for deeper exploration of available information by incorporating context, internal relationships between entities and hidden patterns. Instead of relying solely on pairwise attribute comparison, this approach enables the identification of complex dependencies, the use of intricate logical relationships and the evaluation of structural features in aggregate. Additionally, the constructed ontologies and taxonomies can make a significant contribution to the standardization or alignment of fragmented information, which is particularly important when integrating data from multiple sources [13]. Key advantages of this approach include its holistic perspective, which enables the discovery of clusters and related components of equivalent products by analyzing entire subgraphs sometimes revealing matches missed by simpler methods. However, building and maintaining a comprehensive knowledge graph as a rule represents a labor-intensive and resource-demanding process that requires regular updates as a product assortment expands or new data sources emerge. Furthermore, incomplete or inconsistent taxonomies and ontologies can dramatically affect matching accuracy, which potentially negates the benefits of this approach.

2.4. Product matching based on machine learning

Machine learning and deep neural network approaches treat the product matching as an objective functional optimization task [14, 15]. In this formulation, a pre-trained model evaluates a set of product attributes and determines whether they correspond to an equivalent item. Classical machine learning methods begin with choice and engineering a feature collection to construct a representative description of a product. These engineered features are then passed into a binary classifier (such as logistic regression, random forest or support vector machines (SVM), etc.) which is trained to distinguish similar products based on dependencies in the original attribute space. Application of these methods requires a labeled dataset with indication of target class labels, enabling the model to learn how similarities and differences in features affect the probability of a match. While classical machine learning methods can achieve high predictive accuracy, they rely on labels, strong domain expertise and meticulous engineering of initial attributes, which necessitate ongoing adaptation in case of changes in data or their distribution. Deep learning methods significantly reduce or even eliminate the need for costs related to manual feature engineering by automatically learning patterns from raw data. Common base models include recurrent neural networks (RNNs), long short-term memory networks (LSTMs) and convolutional neural networks (CNNs). These models process the input attributes to generate embedding representing characteristic vector which is then used to assess product similarity. Key advantages of deep learning models include their ability to more accurately handle noisy or heterogeneous data, absence of need for manual feature engineering, and greater robustness when processing previously unknown inputs. However, they also face challenges in capturing and interpreting semantic meaning and still require availability of labeled data.

2.5. Formulation of the task for product matching in marketplaces

Based on the analysis and the operational specifics of digital marketplaces, it becomes clear that typical challenges in product matching include: ambiguous descriptions and the use of marketing terms, data duplication and missing information, multi-format inputs (text descriptions, images, tabular attributes) and varying levels of details. All these factors increase the risk of incorrect matches or, conversely, missed potential matches. Therefore, it is necessary to develop a multimodal model capable of processing both the visual component and product characteristics while accounting for their semantic meaning. This task formulation reflects the heterogeneous nature of the data and the high variability of the digital environment. Relying solely on attributes is often insufficient and can result in matching errors, especially when key product differences are visible in images. In a similar manner, images alone can be ambiguous or appear almost identical. Thus, by integrating these components into a single system, the model will be able to perform product matching more effectively and comprehensively.

3. Multimodal product matching model for marketplaces (MMMP)

Let there be a set of product information cards $\Omega = \{p_1, p_2, p_3, \ldots, p_n\}$ containing information about each product. The task is to design a model Ψ (1) that maps the set Ω into a space $\mathbb{R}^d$ such that the resulting vector representation enables the assessment of similarity between any pair of products from Ω according to (2), where S is a similarity measure:

$$\Psi(p_i) = (x_1, x_2, x_3, \ldots, x_d);\ p_i \in \Omega, \quad (1)$$

$$S : \mathbb{R}^d \times \mathbb{R}^d \to [0, 1]. \quad (2)$$

Since the model must process multiple modalities, its conceptual representation can be decomposed into component parts as shown in (3):

$$\Psi(p_i) = \mathrm{Concat}\,(\psi_{title}(p_i), \psi_{image}(p_i), \psi_{metadata}(p_i));\ p_i \in \Omega, \quad (3)$$

where ψ_{title} encodes the product name into a vector representation;

ψ_{image} encodes the product image into a vector representation;

$\psi_{metadata}$ encodes the product attributes into a vector representation.

Since it is necessary to account for context and semantic meaning, this paper proposes using the transformer architecture introduced in [16] as the foundation. It is known that this architecture can represent textual data in vector form and, due to its attention mechanism combined with deep neural networks, it effectively captures semantic and contextual features [17]. Unlike traditional embedding generation models such as Word2Vec [18] and GloVe [19], which assign a fixed vector to each word, the transformer generates dynamic, context-dependent embeddings. This means that the same word can have different representations depending on the context in which it appears in a sentence. Given that this paper focuses on vector representation, only the encoder block of the transformer is relevant. The encoder architecture consists of three key components: a multi-head self-attention mechanism, a residual connection and normalization layer and a feedforward neural network. Let us examine the main transformations involved.

Let there be a set of input tokens $t = \{t_1, t_2, t_3, \ldots, t_n\}$, each of which is associated with its own embedding as defined in (4):

$$\text{Embed: } t_n \rightarrow e_i \in \mathbb{R}^{d_{model}}, \quad (4)$$

where d_{model} corresponds to the dimensionality of the model's embeddings.

To account for the order of elements, positional encoding of the form (5, 6) is used:

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right), \quad (5)$$

$$PE\left(pos, 2i+1\right) = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right), \quad (6)$$

where *pos* is the position in sequence;

i is the dimensionality.

Then, the input to the first encoder block is given by (7):

$$X = \left(e_1 + \rho e_1, e_2 + \rho e_2, \ldots, e_n + \rho e_n\right) \in \mathbb{R}^{n \times d_{model}}. \quad (7)$$

Each encoder block contains an attention mechanism component that uses query (Q), key (K) and value (V) matrices, which are defined as follows in (8):

$$Q = \mathrm{X}W^Q,\ K = XW^K, V = XW^V; \\ W^Q, W^K, W^V \in \mathbb{R}^{d_{model} \times d_k}, \quad (8)$$

where W^Q, W^K and W^V are the trainable parameters of the model;

d_k is typically defined as the ratio of d_{model} to the number of attention heads.

The attention-based importance score is computed according to (9):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (9)$$

The transformer architecture uses multiple attention heads, which are concatenated as shown in (10) and (11):

$$head_i = \text{Attention } (Q_i, K_i, V_i), \quad (10)$$

$$\text{MultiHead}(X) = \\ = \text{Concat } (head_1, head_2, head_3, \ldots\ head_n)W^O, \quad (11)$$

where the matrix W^O is also a trainable parameter of the model.

An important component of the transformer architecture is the residual connection [20] and normalization layer [21], applied after the attention mechanism as shown in (12):

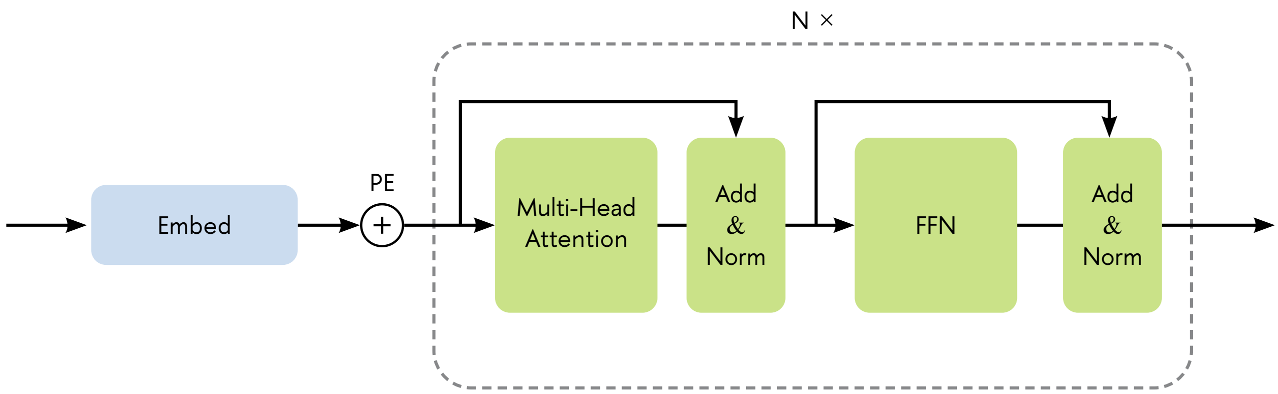
$$X_{attn} = \text{LayerNorm}\,(X + \text{MultiHead}(X)). \quad (12)$$

Each encoder block also includes a feedforward neural network, typically in the form of a two-layer MLP (13) with a nonlinearity such as ReLU (or GeLU):

$$\text{FFN}(X_{attn}) = \max(0, X_{attn}\, W_1 + b_1)W_2 + b_2, \quad (13)$$

after which, the residual connection and normalization layer is applied again. The output from one encoder block is then transferred sequentially through N similar encoder blocks, producing the final output. A conceptual representation of the encoder architecture is shown in *Fig. 1*.

The task of determining the similarity between product titles requires a reliable representation of sentence semantics that allows for effective and accurate comparison of textual content. Traditional transformer-based models, such as BERT (Bidirectional Encoder Representations from Transformers) [22], have dem-



onstrated significant success across a wide range of natural language processing (NLP) tasks. However, BERT is optimized for token-level tasks and does not generate fixed-size, holistic sentence-level embeddings. To overcome this limitation, Sentence-BERT [23] is proposed, which is specifically designed to generate sentence-level embeddings and employs a Siamese neural network architecture. Unlike BERT, which processes each pair independently, Sentence-BERT encodes input into a dense fixed-size vector enabling efficient computation of cosine similarity between sentences in vector space. Moreover, Sentence-BERT can learn using contrastive loss functions of the form (14), which explicitly optimize the model for identifying semantic similarity between sentences rather than just capturing contextual token-level relationships:

$$L = (1 - y_k)\left\| f(x_i) - f(x_j) \right\|_2^2 + \\ + y_k (\max\left(0, m - \left\| f(x_i) - f(x_j) \right\|_2\right))^2, \quad (14)$$

where x_i and x_j are a pair of items;

y_k is a binary label that takes the value 0 if the objects are similar, and 1 if they are different;

m defines the margin that separates the items.

Based on the above and in line with the stated task, this study proposes using Sentence-BERT to generate high-quality sentence embeddings that enable effective assessment of similarity on the basis of semantic relationships between product titles presented in textual form. In this way, we obtain a robust implementation of the ψ_{title} component. Using the generated embeddings, we apply cosine similarity as the similarity metric (15):

$$\text{Sim}(e_i, e_j) = \frac{e_i \cdot e_j}{\|e_i\| \|e_j\|}. \quad (15)$$

To define the visual component of the model ψ_{image} it is important to consider that digital platforms host a large variety of products with significant differences in image quality, angle, lighting, shadows, and the positioning of elements. These factors collectively render traditional similarity assessment methods insufficiently effective. It is known that traditional convolutional neural networks, such as ResNet [24] and EfficientNet [25] are widely used for image similarity search. However, these models rely on local convolutions and therefore have a fixed receptive field, which can limit their ability to capture distant or disjoint relationships between elements within an image. In product matching, fine-grained relationships between elements such as logos, labels, and shapes are crucial, and CNNs often struggle to effectively process such dependencies. Furthermore, CNNs tend to be highly sensitive to the aforementioned variations in images, which raises further concerns about their suitability for solving the current task. To address this issue, this study proposes the use of the vision transformer (ViT) architecture, as described in [26]. The vision transformer (ViT) processes images as a sequence of patches and uses an attention mechanism to capture both local and global relationships within the image. Unlike CNNs, which extract features in a hierarchical manner, the vision transformer analyzes the entire image simultaneously, allowing it to dynamically focus on the most relevant areas. This distinctive property makes ViT particularly useful for assessing the similarity of clothing items, electronics and consumer goods, where differences in texture, brand placement, and various distortions can significantly impact the final similarity score. Another important advantage of the vision transformer in the context of this task is its high robustness to occlusions in images. For example, ViT can correctly identify a product model even if the brand logo is partially obscured or the image is captured from a different angle, something traditional CNNs typically struggle with. It is also worth noting that vision transformers offer strong potential for integrating image and its textual description to compute semantic similarity between them [27]. This is especially valuable in product search scenarios, where different textual attributes (such as title, description, brand, model, etc.) and the corresponding image can be associated to check compliance.

Based on the above, this study proposes defining ψ_{image} as a multistage pipeline consisting of pretraining, fine-tuning and subsequent evaluation.

In the first stage, we perform initial training on a large dataset to learn generalized representations and extract fundamental features. Let f_θ denotes a vision transformer (ViT) parameterized by θ and let there be a dataset containing images of all products $I = \{x_1, x_2, x_3, ..., x_n\}$. Using the self-supervised distillation method DINO [28], we optimize according to (16) and generate an initial embedding for each product image as shown in (17):

$$\theta^* = \arg\min_\theta L_{SSL}(f_\theta, I). \tag{16}$$

$$e_i = f_{\theta^*}(x_i) \forall x_i \in I. \tag{17}$$

In the second stage, we adapt the vision transformer. To do this, we use subsets of product images obtained by dividing I into categories and price segments, and by performing fine-tuning according to (18):

$$L = \max\begin{pmatrix} \left\| f_{\theta^*}(x_a) - f_{\theta^*}\left(x_p\right)\right\|_2^2 - \\ -\left\| f_{\theta^*}(x_a) - f_{\theta^*}(x_n)\right\|_2^2 + \alpha, 0 \end{pmatrix}, \tag{18}$$

where x_a is the anchor object;

x_p denotes the object similar to the anchor;

x_n represents the object that is different from both x_a and x_p.

After completing the fine-tuning process, we use (15) to evaluate the similarity between any pair of vector representations e_i and e_j. The described pipeline ensures high efficiency and accuracy of the ψ_{image} component.

When defining the third component of the model $\psi_{metadata}$ it is important to consider that product characteristics in digital marketplaces include various types of data, such as numerical variables (e.g., weight, volume, or dimensions) and categorical variables (e.g., configuration, color, or composition). Traditional approaches often rely on multilayer perceptrons (MLPs) or ensemble methods, which frequently struggle to detect complex interactions in a multitude of diverse features, in the absence of prior feature engineering. Given the nature and specifics of the domain, it is reasonable to assume that certain features within the tabular representation of product attributes may interact with each other in a non-

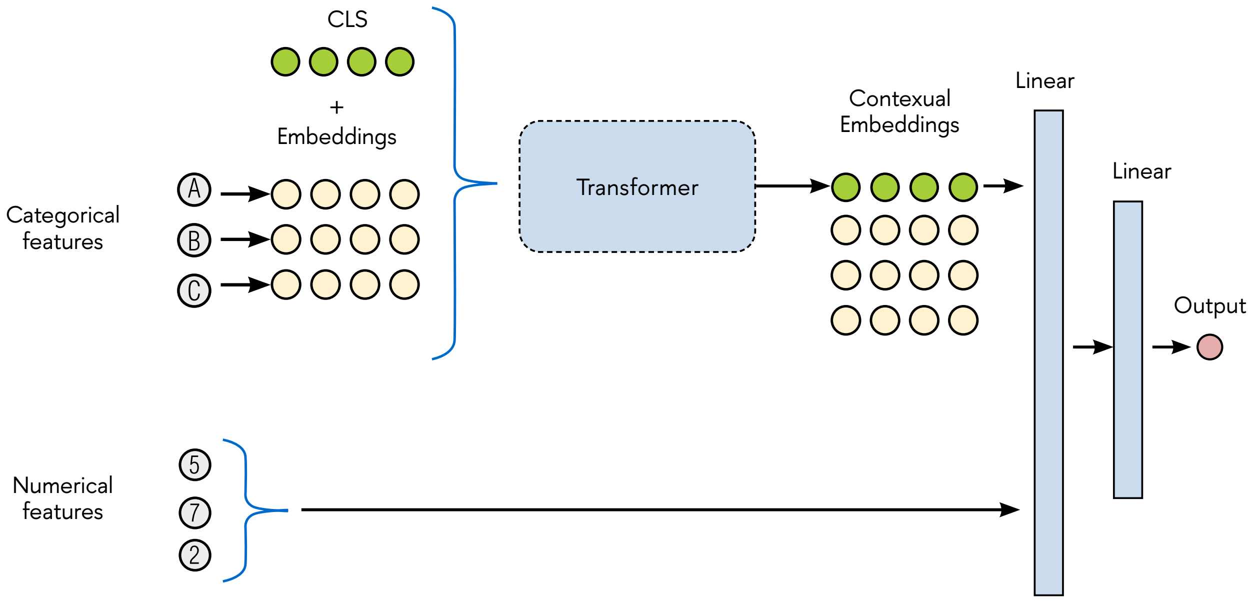


Fig. 2. Architecture of the $\psi_{metadata}$ component.

trivial way. These interactions are crucial for the task at hand, as even minor differences (such as between "cotton" and "poly-cotton" in composition) can significantly impact the perceived identity of characteristics. As previously discussed, the core of the transformer architecture is the attention mechanism, enabling detection of how the attributes are interrelated. The context-dependent embeddings generated by the transformer form vector representations that position similar elements close to each other in the feature space, taking their interrelationships into account. This enables more accurate differentiation of similar products based on their available attributes. Moreover, the attention mechanism effectively identifies feature interactions in an automated manner, eliminating the need for domain experts and manual feature engineering. In addition to its high flexibility with respect to input features, the transformer is also effective at handling missing or distorted data, which is a common issue on large e-commerce platforms. Furthermore, several studies [29, 30] exploring the application of transformers to tabular data have shown that such architectures can outperform traditional approaches.

Based on the above, the $\psi_{metadata}$ component of the model is defined using the concept of a tabular transformer. The architecture adapted for the task at hand, as proposed in this study, is shown in *Fig. 2.* To evaluate the similarity between any pair of products using the tabular transformer, we apply (15).

The proposed tabular transformer architecture effectively generates vector representations of product attributes by applying context-dependent embeddings for categorical features, specialized processing of numerical variables and an attention mechanism to capture complex interactions. This is particularly relevant in large-scale and heterogeneous catalogs, which are typical of digital marketplaces.

After defining all components, the final similarity measure is specified using a weighted scoring method as shown in (19):

$$S\left(p_i, p_j\right) = w_1 \mathrm{Sim}\left(\psi_{title}(p_i), \psi_{title}(p_j)\right) + \\ + w_2 \mathrm{Sim}\left(\psi_{image}(p_i), \psi_{image}(p_j)\right) + \\ + w_3 \mathrm{Sim}\left(\psi_{metadata}(p_i), \psi_{metadata}(p_j)\right), \sum w_i = 1. \quad (19)$$

Here the weighting coefficients w_1, w_2 and w_3 can be flexibly adjusted to account for the characteristics and specific requirements of the matching task.

4. MMMP application for product similarity assessment in Wildberries marketplace

As part of the experiment, a decision was made to conduct a study on products listed on the Wildberries platform. The primary focus is on the product category "Smartphones," which was chosen due to several important factors. First, smartphones are complex technical devices with a wide range of features, including both numerical attributes (e.g., memory size, battery capacity, etc.) and categorical attributes (e.g., brand, operating system, etc.). This makes them a suitable object for testing in heterogeneous features environment. Second, the assortment of such devices in marketplaces is quite extensive and diverse, allowing the model operation to be examined on highly variable data. Third, this category is one of the most popular and in-demand in marketplaces, which adds practical significance to the results of the study. In combination these factors provide an effective setting for testing the model's ability to assess product similarity. Thus, selecting the "Smartphones" category enables evaluation of the proposed model under real-world conditions, offering sufficient complexity and diversity to assess its performance. It follows that a successful application of the model in this product group can be extended to other categories, including technically complex goods.

During the experiment, a dataset was collected consisting of 12233 product listings. Each listing contains information about a specific item available on the platform and includes the following features: X1 – product image, X2 – product title, X3 – color, X4 – model, X5 – SIM card type, X6 – operating system, X7 – operating system version, X8 – warranty period, X9 – dust and water protection rating, X10 – display type, X11 – screen size, X12 – screen resolution, X13 – screen refresh rate, X14 – screen protective coating, X15 – internal storage capacity, X16 – RAM, X17 – main camera, X18 – front camera, X19 – lens features, X20 – built-in flash, X21 – battery capacity, X22 – processor model, X23 – number of processor cores, X24 – network standard, X25 – wireless interfaces, X26 – satellite navigation, X27 – connector type, X28 – additional features, X29 – package contents, X30 – product description, X31 – country of manufacture, X32 – number of SIM cards, X33 – product condition, X34 – processor clock speed, X35 – product weight with packaging, X36 – service life.

A sample of the collected data is shown in *Table 1.*

Before using the collected data to train the model, thorough preprocessing was carried out to ensure data quality, consistency and suitability for further analysis. These procedures included cleaning, duplicate removal, handling of missing values, correction of inconsistencies, normalization of numerical data and tokenization of textual information. Additionally, outlier detection and removal were performed to minimize the impact of anomalous values.

After training the model, similarity scores were obtained for the products shown in *Table 1*, with the results presented in *Table 2.* The effectiveness of the proposed method was also evaluated by comparing it with a naive baseline approach based on Jaccard index, applied without prior tokenization.

The experimental part of the study confirmed the effectiveness of the proposed approach. Specifically, the data in *Table 2* illustrate the example of similarity scores calculated by the MMMP model for four product listings (Product 1, Product 2, Product 3, and

Table 1.

Sample of collected data

	Product 1	Product 2	Product 3	Product 4
Image				
Title	Apple iPhone 16 Pro Max Gold/Desert 512 GB	Apple iPhone 16 Pro Max White 1 TB	Galaxy S24 Ultra 512 GB Yellow	Xiaomi 14 12/256 GB 5G White RST
Price	168 750	182 267	159 800	77 585
Model	iPhone 16 Pro Max	iPhone 16 Pro Max	S24	Xiaomi 14
OS version	iOS 18	IOS 18	Android 14	Android 14
Warranty period	1 year	1 year	1 year	1 year
Display type	Super Retina XDR OLED ProMotion	Super Retina XDR OLED ProMotion	Dynamic AMOLED 2X	LTPO AMOLED
Screen size	6.9"	6.9"	6.8"	6.36"
Screen resolution	2868x1320	2868x1320	3120x1440	2670x1200
Internal storage capacity	512 GB	1 TB	512 GB	256 GB
RAM	8 GB	8 GB	12 GB	12 GB
Wireless interfaces	Wi-Fi 7; NFC; Bluetooth 5.3	Wi-Fi 7; NFC; Bluetooth 5.3	Wi-Fi; NFC; Bluetooth	Wi-Fi; IR-Port; NFC; Bluetooth
…	…	…	…	…

Table 2.

Product similarity assessment

		Product 1	Product 2	Product 3	Product 4
Product 1	MMMP	1	0.98657	0.877564	0.544109
	Jaccard	1	0.308	0.11	0.077
Product 2	MMMP	0.98657	1	0.883621	0.549463
	Jaccard	0.308	1	0.10	0.11
Product 3	MMMP	0.877564	0.883621	1	0.670598
	Jaccard	0.11	0.10	1	0.13
Product 4	MMMP	0.544109	0.549463	0.670598	1
	Jaccard	0.077	0.11	0.13	1

Product 4). The highest similarity score (0.98657) is observed between Product 1 and Product 2, indicating successful identification of nearly identical items despite differences in descriptions and data modalities. Simultaneously less similar items (such as Product 1 and Product 4) received significantly lower similarity score (0.544109), demonstrating the model's ability to correctly distinguish between groups of products with different sets of attributes and descriptions. Similar results (0.549463 between Product 2 and Product 4, and 0.670598 between Product 3 and Product 4) show that the model is capable of recognizing even subtle differences in product characteristics while also detecting meaningful patterns in their descriptions. Thus, the proposed approach reliably identifies both clear and nuanced matches and avoids incorrect grouping of products that are, in fact, different.

The results obtained confirm that the transformer-based multimodal model is capable of effectively solving the product matching task in the presence of a wide variety of product listings and heterogeneous data sources. Accurate identification of product records in marketplaces reduces the risk of duplication and analytical errors, positively impacting all stages of economic activity within the digital platform ecosystem. It also helps both sellers and buyers to navigate the product assortment more accurately, simplifying the decision-making process. Moreover, the results obtained demonstrate the model's correct functioning in relation to the defined task making possible its application in intelligent product matching in marketplaces.

Conclusion

In the course of this study, the task of intelligent product matching in digital marketplaces was proposed and solved. This task requires comprehensive analysis and processing of multimodal data, as well as the application of modern instrumental methods within the framework of an economic and mathematical model (MMMP). The first section of the article highlighted the potential of product matching in marketplaces as typical representatives of digital platforms and demonstrated the relationship between intelligent matching capabilities and key economic performance indicators within the e-commerce market. The second section analyzed existing approaches to entity resolution (ER) and multimodal data analysis, revealing the core challenge of product record matching in marketplaces: the fact that products are described by different sellers with varying levels of details and in different formats (text, images, tabular attributes, etc.). The third section described the detailed development process of the multimodal product matching model (MMMP) whose core is composed of transformer-based modules for processing textual, visual, and tabular data. The proposed architecture takes into account the flexibility of the attention mechanism and is capable of self-learning as product assortments expand and new unstructured descriptions emerge with consideration of visual data components. MMMP enables effective integration of diverse input data modalities and captures complex contextual relationships which are critical for accurate determination of product similarity or dissimilarity, considering pricing characteristics.

Finally, in the fourth section, the application of the proposed model for assessing product similarity in the Wildberries marketplace was demonstrated, where the results confirmed the high accuracy and stability of the MMMP operation.

Thus, the main conclusions of the article can be summarized in an axiomatic form:

- the high relevance of multimodal analysis for product matching tasks has been substantiated in the context of the growing diversity of product listings in marketplaces;
- the potential of transformer architectures for the integrated processing of textual, visual and tabular features has been established;
- the need for further adaptation of such models to evolving market conditions has been identified;
- the MMMP model has been developed, which can be used both in academic research of intelligent product record identification in marketplaces and for practical purposes by participants in the e-commerce market — where matching accuracy is crucially important, for instance, in pricing, and thus for establishing an equilibrium model of supply and demand.

The results of this work demonstrate that a well-designed deep learning architecture that integrates multiple data modalities provides a significant advantage over simpler and more narrowly focused existing solutions. Moreover, the application of such models contributes to enhancing the transparency and effectiveness of analytical tools, which ultimately helps strengthen the trust of sellers and more importantly, the trust of buyers in a given marketplace as a whole.

Future research directions within the chosen subject area may include expanding the set of processed modalities (e.g., incorporating video reviews or audio information) and developing self-learning mechanisms that allow the model to automatically reconfigure itself in response to changes in data structure. Additionally, the continued development of the MMMP model suggests the integration of active learning methods (in real time), as this would enable faster accumulation and processing of relevant examples to refine matching criteria, including in the environment of marketplaces. ■

References

1. Fletcher A., Ormosi P. L., Savani R. (2023) Recommender systems and supplier competition on platforms. *Journal of Competition Law & Economics*, vol. 19, no. 3, pp. 397–426. https://doi.org/10.1093/joclec/nhad009

2. Hussien F.T.A., Rahma A.M.S., Abdulwahab H.B. (2021) An e-commerce recommendation system based on dynamic analysis of customer behavior. *Sustainability*, vol. 13, no. 19, article 10786. https://doi.org/10.3390/su131910786

3. Chen F., Liu X., Proserpio D., et al. (2020) Studying product competition using representation learning. Proceedings of the *43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20)*, pp. 1261–1268. https://doi.org/10.1145/3397271.3401041

4. Hu S., Wei M.M., Cui S. (2023) The role of product and market information in an online marketplace. *Production and Operations Management*, vol. 32, no. 10, pp. 3100–3118. https://doi.org/10.1111/poms.14025

5. Cheung M., She J., Sun W., Zhou J. (2019) Detecting online counterfeit-goods seller using connection discovery. *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 15, no. 2, article 35. https://doi.org/10.1145/3311785

6. Sun J., Zhang X., Zhu Q. (2020) Counterfeiters in online marketplaces: Stealing your sales or sharing your costs. *Journal of Retailing*, vol. 96, no. 2, pp. 189–202. https://doi.org/10.1016/j.jretai.2019.07.002

7. Köpcke H., Thor A., Rahm E. (2010) Evaluation of entity resolution approaches on real-world match problems. *Proceedings of the VLDB Endowment*, vol. 3, nos. 1–2, pp. 484–493. https://doi.org/10.14778/1920841.1920904

8. Cohen W.W., Ravikumar P., Fienberg S.E. (2003) A Comparison of string distance metrics for name-matching tasks. Proceedings of *Workshop on Information Integration (IJCAI-03)*, pp. 73–78.

9. Singh R., Meduri V.V., Elmagarmid A., et. al. (2017) Synthesizing entity matching rules by examples. *Proceedings of the VLDB Endowment*, vol. 11, no. 2, pp. 189–202. https://doi.org/10.14778/3149193.3149199

10. Wang J., Li G., Yu J.X, Feng J. (2011) Entity matching: How similar is similar. *Proceedings of the VLDB Endowment*, vol. 4, no. 10, pp. 622–633. https://doi.org/10.14778/2021017.2021020

11. Angermann H. (2022) TaxoMulti: Rule-based expert system to customize product taxonomies for multi-channel e-commerce. *SN Computer Science*, vol. 3, article 177. https://doi.org/10.1007/s42979-022-01070-8

12. Mao M., Chen S., Zhang F., et. al. (2021) Hybrid ecommerce recommendation model incorporating product taxonomy and folksonomy. *Knowledge-Based Systems*, vol. 214, article 106720. https://doi.org/10.1016/j.knosys.2020.106720

13. Aanen S. S., Vandic D., Frasincar F. (2015) Automated product taxonomy mapping in an e-commerce environment. *Expert Systems with Applications*, vol. 42, no. 3, pp. 1298–1313. https://doi.org/10.1016/j.eswa.2014.09.032

14. Ristoski P., Petrovski P., Mika P., Paulheim H. (2018) A machine learning approach for product matching and categorization: Use case: Enriching product ads with semantic structured data. *Semantic Web*, vol. 9, no. 5, pp. 707–728. https://doi.org/10.3233/SW-180300

15. Shah K., Kopru S., Ruvini J. D. (2018) Neural network based extreme classification and similarity models for product matching. Proceedings of the *2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans – Louisiana*, vol. 3, pp. 8–15. Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-3002

16. Vaswani A., Shazeer N., Parmar N., et. al. (2017) Attention is all you need. Proceedings of the *31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA*, pp. 6000–6010. https://dl.acm.org/doi/pdf/10.5555/3295222.3295349

17. Zhang H., Shafiq M.O. (2024) Survey of transformers and towards ensemble learning using transformers for natural language processing. *Journal of Big Data*, vol. 11, article 25. https://doi.org/10.1186/s40537-023-00842-0

18. Mikolov T., Chen K., Corrado G., Dean J. (2013) Efficient estimation of word representations in vector space. *arXiv:1301.3781*. https://doi.org/10.48550/arXiv.1301.3781

19. Pennington J., Socher R., Manning C. D. (2014) GloVe: Global vectors for word representation. Proceedings of the *2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar*, pp. 1532–1543.

20. He K., Zhang X., Ren S., Sun J. (2016) Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

21. Ba J.L., Kiros J.R., Hinton G.E. (2016) Layer normalization. *arXiv:1607.06450*. https://doi.org/10.48550/arXiv.1607.06450

22. Devlin J., Chang M. W., Lee K., Toutanova K. (2019) Bert: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the *2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, Minnesota*, vol. 1, pp. 4171–4186. Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

23. Reimers N., Gurevych I. (2019) Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the *2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China*, pp. 3982–3992. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1410

24. Wu Z., Shen C., van den Hengel A. (2019) Wider or deeper: Revisiting the ResNet model for visual recognition. *Pattern Recognition*, vol. 90, pp. 119–133. https://doi.org/10.1016/j.patcog.2019.01.006

25. Tan M., Le Q. (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the *36th International Conference on Machine Learning*, vol. 97, pp. 6105–6114.

26. Dosovitskiy A., Beyer L., Kolesnikov A., et al. (2021) An image is worth 16x16 words: transformers for image recognition at scale. *arXiv:2010.11929*. https://doi.org/10.48550/arXiv.2010.11929

27. Radford A., Kim J. W., Hallacy C., et. al. (2021) Learning transferable visual models from natural language supervision. Proceedings of the *38th International Conference on Machine Learning*, vol. 139, pp. 8748–8763.

28. Caron M., Touvron H., Misra I., et. al. (2021) Emerging properties in self-supervised vision transformers. *arXiv:2104.14294*. https://doi.org/10.48550/arXiv.2104.14294

29. Huang X., Khetan A., Cvitkovic M., et. al. (2020) TabTransformer: Tabular data modeling using contextual embeddings. *arXiv:2012.06678*. https://doi.org/10.48550/arXiv.2012.06678

30. Gorishniy Y., Rubachev I., Khrulkov V., et. al. (2021) Revisiting deep learning models for tabular data. Proceedings of the *35th International Conference on Neural Information Processing Systems (NIPS'21)*, article 1447, pp. 18932–18943.

About the authors

Artem Yu. Varnukhov

Assistant, Department of Business Informatics, Ural State University of Economics, 62, 8 Marta St., Yekaterinburg 620144, Russia;

E-mail: varnuhov_ayu@usue.ru

Dmitry M. Nazarov

Doctor of Sciences (Economics);

Head of Department, Department of Business Informatics, Ural State University of Economics, 62, 8 Marta St., Yekaterinburg 620144, Russia;

E-mail: slup2005@mail.ru

ORCID: 0000-0002-5847-9718

DOI: 10.17323/2587-814X.2025.2.25.40

Development of recommendation systems to improve the efficiency of regulated procurement in the electric power industry

Anna I. Denisova [a]
E-mail: a.i.denisova@inbox.ru

Dzhamilya A. Sozaeva [a,b]
E-mail: dasozaeva@gmail.com

Konstantin V. Gonchar [a,c]
E-mail: goncharkv@gmail.com

[a] State University of Management, Moscow, Russia
[b] Synergy University, Moscow, Russia
[c] Bauman Moscow State Technical University, Moscow, Russia

Abstract

This article considers ways to improve the efficiency of the regulated procurement market by implementing recommender systems into the existing procurement IT infrastructure. Using state, municipal and commercial procurement of electric power products as an example, the article considers promising classes of recommender systems for implementation, proposes a methodology for developing such services, and discloses algorithms for processing, configuring and interpreting data necessary for their operation. The difference between the author's approach to creating services and previously published works is substantiated, testing and A/B testing are carried out, and an assessment of the effectiveness is presented. The results obtained have scientific novelty (the methodology of using neural networks in relation to the procurement industry has been substantiated) and practical significance (the customer's time saved on searching for suppliers by up to 40%; the pool of potential suppliers has been expanded; supplier risks have been diversified by selecting relevant procedures from new areas and from new customers; suppliers have been provided with the opportunity to find up to 2–3 new customers for 1 recommendation mailing with a frequency of 1–2 times a week). We proposed to implement the developments in the practice of the operator of public procurement tenders. The authors see further development of recommendation services and solutions for the procurement industry in improving the analysis of semantic (text, logical) content of procurement documents, as well as the behavioral strategies of suppliers. The risks and limitations are associated with the high cost of maintaining a staff of developers-practitioners in neural networks, possible hallucinations of neural networks and their high sensitivity to errors and original data sets.

Keywords: recommender systems, efficiency of regulated procurement, probability of winning in public procurement, personalized recommendations, "non-closing" of tenders, competition in procurement

Citation: Denisova A.I., Sozaeva D.A., Gonchar K.V. (2025) Development of recommendation systems to improve the efficiency of regulated procurement in the electric power industry. *Business Informatics*, vol. 19, no. 2, pp. 25–40. DOI: 10.17323/2587-814X.2025.2.25.40

Introduction

Improving IT solutions and developing a digital ecosystem for public procurement, procurement of companies with state participation (hereinafter referred to as public procurement, regulated procurement) are priority tasks reflected in the state program of the Russian Federation "Public Finance Management and Regulation of Financial Markets" for the period up to 2030 [1]. In the early stages of the program's implementation, digitalization of procurement solved a particular number of problems: organizing electronic interaction between customers and suppliers; accelerating procurement procedures; reducing barriers for entrepreneurs to enter the public procurement market; increasing competition and the transparency degree of tenders; and minimizing the corruption elements. In the past few years, the development of information technologies in this field has faced new challenges which are reflected in the priorities for the development of the industry in the period from 2025 to 2030. According to the authors of the study, this is due to the trends emerging in e-commerce in general. Thus, at present, the following performance indicators like system speed, fault tolerance, storing and processing significant volumes of data, and the confidentiality guarantee are no longer competitive advantages of information systems: this is a mandatory set that guarantees survival in the market. The strengths of modern IT solutions lie in built-in algorithms and services that make it possible to predict the consequences of management decisions, while not requiring special programming skills from the user. Hence, the implementation of such developments allows one to reduce the risks and negative consequences of decisions made for the financial and economic activities of the organization. With the use of predicting services, it becomes possible to minimize transaction costs for operational tasks, ensuring the sustainable functioning of the enterprise and the achievement of its objectives. As for the field of regulated procurement, over the past three years the problems of effective "closing" of tenders (successful conduct of the procedure, with determination of the winner and signing a contract with him) have been acute issues, with high risks of contract termination, as well as maintaining an optimal level of competition in procurement. "Failure to close" procurement procedures, suppliers' absence at tenders are often associated with unfavorable price and terms and conditions of contracts established by government bodies (buyers) [2]. The risks of current contract termination might be caused not only by force majeure on the supplier's side, as is often believed, but also by entrepreneurs' overestimation of their production capabilities and resource constraints. This problem can be considered following the example of state and municipal procurement, where the statistics are most accurate. According to a report from the Ministry of Finance of Russia, 323 131 contracts were terminated in 2023 for a total of RUB 708.5 billion, which is 9.2% of the total number and 5.9% of the total value of contracts signed in 2023. Similarly, in 2022, the number of terminated contracts constituted 295 072 units, which amounted to 9.2%of the total number and 5.9% of the total value of contracts [3]. According to the report for the 1st and 2nd quarters of 2024, the termination statistics slightly improved in terms of the number of contracts (fewer) and worsened in terms of value (the total value of terminated contracts increased [4]). Thus, there have been no significant improvements in the termination statistics yet. As for competition in tenders, this indicator in regulated procurement formally answers questions about the presence of a sufficient number of participants in commodity and industry markets, as well as the presence of biases towards purchases from a single supplier (contractor, performer), cartels and other anti-competitive agreements. For a more accurate illustration of this phenomenon, we could see the statistics of government procurement since 2022. The level of competition also fluctuates within 2–3 applications per one tender against 4.2 in the period before the pandemic caused by the COVID-19 coronavirus [5]. The Ministry of Finance notes an increase in the number of purchases from a single supplier and sounds the alarm about negative trends in terms of competition.

Analyzing all of the above, it is quite difficult to determine which problem has a higher priority. There are two aspects which need to be considered in cases of "non-closure" of tenders. The first one is low competition due to the passivity of suppliers, their ignorance of the conditions and opportunities of the market, fears of failing to cope. Another one is high demands of customers, leading to the signing of contracts with very weak players, a single supplier, or, to the refusal of suppliers to enter the regulated procurement market in general. In such a situation, it seems that there is an imbalance between the goals pursued by government bodies and suppliers in the government procurement market and the information they have. This is confirmed by a large-scale study by the team from HSE University which highlights "a discrepancy between the criteria for the efficiency of procurement which government and suppliers are guided by in their current activities and the goals that the procurement regulation system is currently focused on" [6]. Government and state-financed organizations strive to demonstrate high competition in their purchases with an enhanced cost efficiency, a minimum number of contract terminations, while suppliers are focused on minimal competition and keeping the price reasonable. However, both parties agree on the contract termination policy: they aim to fulfill their contracts without negative consequences. According to the author, the importance of regulated procurement for the country's economy determines the need to maintain a balance of actors' interests in this market by increasing their awareness. It comprises clear understanding about the nature and state of the market, prospects and feasibility of conducting procurement procedures or participating in them at any given time, in a particular region and under other specific conditions relevant to the current level of technological development. We claim that whole market awareness can be raised through active digitalization and introduction of *recommendatory predictive systems* that will stimulate suppliers to participate in tenders. The recommendatory predictive systems allow one to select procurement procedures relevant to the scale, experience and resource availability of companies for successful participation in them. For the government, the development of such services in procurements brings many advantages: their procedures are applied for by those suppliers who have the potential to fulfill contracts at a real level of competition among worthy market players, and, therefore, the risks of unfair performance or termination are significantly reduced.

Thus, the purpose of this study is to justify the use of a specific class of recommender systems to improve the efficiency of entrepreneurs' participation in regulated procurement and to demonstrate the operation of the service using a separate industry/sphere as an example. To show the novelty of this approach, the following theoretical aspects of the issue should be considered.

1. Methodology for the development and implementation of recommendation services for the regulated procurement sector

A recommender system (RecSys) is an algorithm that selects and offers relevant content to the user based on available information about the content, the user, his behavior and the behavior of other users, as well as about their utilization of the content [7].

Recommender systems are playing an increasingly important role in today's information and business world. With the ever-growing volume of data and content available to users, it is becoming increasingly difficult to navigate and find the most relevant information. Recommender systems can solve these problems: by analyzing user preferences and context, they can provide personalized recommendations that match the interests and needs of users.

The basic principles of decision support systems (preference modeling, decision making under uncertainty) [8–10] have found partial application in modern recommender system algorithms, but some important reasons (for example, psychological and cognitive aspects) are usually left out. It is worth emphasizing that decision support systems and recommender systems are created to fulfill different goals. A recommender system is focused on predicting what content might be interesting to a user based on his past behavior, preferences and actions of other users, thus achieving personalization of offers through the use of various data processing algorithms and the identification of hidden patterns, often without highlighting any specific choice. A classic decision support system, in turn, is created to help the user make an informed decision under uncertainty, providing an explanation for why one or another alternative is better than others.

In recent years, there has been an explosive growth in the interest of developers and users in recommendation systems in various fields, including e-commerce, social networks, music, movies, news and much more. This can be partly linked to how the theory of choice has changed: for many years, economists talked about "rational choice," when consumer behavior was measurable. Budget constraints often served as measures of the effectiveness of decisions made. The development of the "irrational theory of choice" and the work of Thaler [11], and, in general, interest in behavioral economics, led to the popularization of emotional, social and personal components in decision-making, shopping, sales, etc. It became obvious: it is thanks to recommendations that users receive personalized content that matches their individual interests, all of which significantly increases the usability of most IT services.

As for such a specific area as regulated procurement, the recommendation generation system should work as follows: suppliers are regularly offered a list of the most relevant tenders for them, in which they are expected to participate. In this case, the recommendation can be received through various communication channels: notifications in the user's personal account, email newsletters, notifications via messengers, etc.

In the information field, there are several attempts known to develop and implement recommender systems in the procurement sphere both in the Russian and foreign markets [12–14]. However, they have a number of limitations.

The main approaches on the development of recommendation services in procurement can be divided into five groups [15–17]:

1. Content-based filtering: the algorithm analyzes the characteristics of elements that the user has already "worked with" and offers him similar ones. In the context of a procurement system, this may mean, for example, that the supplier will be offered those tenders in which the customer is located in one of the regions where the supplier has already worked.
2. Collaborative filtering: the algorithm analyzes the history of user activities and searches for groups with similar preferences among them in order to offer new users exactly what others liked: this system is based on the history of user interactions with elements. In the context of a procurement system, this may mean that pools of similar suppliers will be formed, for example, according to the principles of work in one region and within one sphere. Then, if supplier A takes part in a tender, supplier B from the same pool can be recommended to take part in the same tender.
3. Popularity-based recommendations: the algorithm recommends the elements that are most popular among users. This approach can be made more complex by dividing users into clusters and determining the most popular elements based on them. It is most appropriate to use this technology in the case where there is a lack of data on a specific user. In the context of a procurement system, we can recommend, for example, procedures that relate to the most popular areas according to the Russian classification of products by type of economic activity.
4. Recommendations based on subject area knowledge (knowledge-based): the algorithm offers the user elements that are related to those he has already shown interest in. Since the authors rely primarily on the needs of the supplier in this study, this approach is hardly applicable. However, for the customer, its implementation could look as follows: customer A purchased laser printers, and after that the system suggests that he purchase A4 paper, proposing a list of suitable suppliers for this.
5. Hybrid systems (hybrid systems) offer a combination of approaches listed (mainly content and collaborative filtering) to provide the most personalized recommendations.

The authors see the goals of the recommender system in the procurement sphere primarily in raising the number of procurement participants, increasing their activity in tenders and, due to this, improving competition and reducing the number of failed purchases. We have already devoted a number of our works to the problem of failed procedures in tenders [18, 19] and consider it one of the key issues in the efficiency of procurement activities. In this regard, it is inappropriate to use the approach based on popularity (3) since in this way suppliers will be recommended procedures with an already high level of competition. In addition, the procurement process is clearly linked to the time: the opportunity to submit applications and participate in tenders lasts 1–2 weeks on average. Recommendations should also be generated with appropriate frequency. For the generation of recommendations by this approach, tenders which are already accomplished or going to be accomplished soon are considered as making the participation of new suppliers impossible or meaningless since such tenders already become irrelevant. Thus, it can be concluded that in this case the most suitable method for forming a recommender system in the procurement sphere is *content filtering*. Based on this thesis, we will justify the novelty of the approach: until now, recommender systems have not been used on a large scale in the sphere of public procurement, remaining exclusively the prerogative of e-commerce and B2B services.

2. Principles of selecting initial data for forming recommendations

The explosive growth of interest in recommender services is due to the fact that almost all public procurement procedures are now conducted electronically. Electronic trading platforms (ETP) ensure the conduct of electronic trades. To develop their thesis, the authors selected "federal trading operators" from among all the platforms, "federal trading operators" are selected that have the right to conduct purchases of government customers under 44-FZ: Sberbank AST, RTS-Tender, National Electronic Platform (Fabrikant), ETP GPB (Gazprombank), AGZ RT, JSC "EETP" (Roseltorg), Russian Auction House (ETP RAD), TEK-Torg [20]. Another platform can be added to this group – ETP AST GOZ, where state defense order trades are conducted.

The information base of the study consists of data on the activity of suppliers on the platform of JSC EETP (Roseltorg) since 2020 (historical sample) considering open data posted in EIS[2]. The model was trained using participation data from 2020 divided into two-week intervals (in accordance with the average duration of collecting applications for tenders). Testing was conducted for the period October–December 2023. Only electronic purchases in the energy sector published in EIS were considered. That meant that at least one code assigned to the procedure was included in the OKPD2 27 group "Electrical equipment". The choice of goods for the electric power industry was due to their high importance for the life support of customers. It was necessary to preliminarily compile "profiles" of suppliers based on their preferences according to historical data. The following aspects were of interest: supplier characteristics, including what areas of activity they were engaged in (according to the OKPD2 classifier), where (in what regions and on what sites), on what regulatory framework (in accordance with what federal laws tenders were held) and with whom the supplier interacted.

Several types of organizations could act as customers:

- authorities, budgetary network institutions that spend budget funds in accordance with Federal Law No. 44-FZ of 05.04.2013 "On the contract system in the sphere of procurement of goods, works, services to meet state and municipal needs";

[2] Unified information system in the sphere of procurement, https://zakupki.gov.ru/epz/main/public/home.html

- companies with state and municipal participation (such as PJSC Gazprom, PJSC Sberbank, PJSC VTB, etc.), as well as state and municipal unitary enterprises operating on the basis of Federal Law No. 223-FZ of 18.07.2011 "On Procurement of Certain Types of Legal Entities";
- commercial customers, the conduct of purchases of which is determined only by the Civil Code of the Russian Federation and the rules established by customers.

If the need to collect and process formal characteristics for the supplier's profile (region, main field of activity, etc.) was obvious in the context of making recommendations, then the issue of evaluating interaction in customer–supplier pairs has not been sufficiently studied. This is due to the fact that information about companies' participation in tenders from 2022 is not intended for publication in open sources, which significantly complicates the identification of preferences and behaviors in customer–supplier pairs. Meanwhile, based on their expert experience in the procurement industry, the authors of the study suggested that non-economic relations between customers and suppliers (friendly, kinship, national, religious, political, and others) supplier preferences may also be affected.

First of all, it is necessary to illustrate that the connection between customers and suppliers really takes place: *Table 1* presents an estimate of their interaction frequency for 2022 and 2023. Even leaving out narrower areas of activity (more detailed OKPD2 codes), the share of stable interactions is significant, especially against the background of a reduction in the number of customers (according to data from open sources (EIS in the field of procurement) using the Roseltorg platform as an example), in 2023, compared to 2020, their number decreased by 35%).

The algorithm for generating recommendations consists of the following steps.

1. Formation of a profile of each supplier[3] based on a statistical assessment of their preferences.

2. Collecting information on all relevant published procedures over the past two weeks.

Table 1.

Evaluation of interaction between suppliers and customers within the OKPD2 group 27

Indicator	2022	2023
The share of customer–supplier pairs identified earlier (starting from 2020) on the Roseltorg platform, among all pairs formed during the year, %	19.9	21.1
The ratio of the number of customer–supplier pairs identified previously (starting from 2020) using the Roseltorg platform as an example, to the number of unique customers who showed activity during the corresponding year, %	0.89	0.96
Average level of competition at auctions (average number of applications per tender), %	2.05	2.01

[3] In the context of the task at hand, only those suppliers who have participated at least once in the procedures for purchasing goods according to OKPD2 27 are analyzed here.

3. Preliminary filtering of procedures by price categories.

Today, the practice of segmenting purchases depending on the initial (maximum) contract price [3] has become established in the work of large federal tender operators. Moreover, depending on the segment, different pricing segments are determined, which cannot but affect participation. Usually, 12 segments are distinguished: "up to 100 thousand rubles," "from 100 to 500 thousand rubles", etc. up to the segment "from 100 million rubles," as well as a separate case when the price is not determined. It has been empirically established that the supplier is not recommended for tenders which price does not fall into their price category or the nearest neighboring ones.

4. Calculation of feature values using formula (1):

$$x_i = \frac{\overline{x}_i \cdot n}{N}, \quad (1)$$

where $\overline{x}_i$ – the average share of a given feature in the supplier's history;

n – the number of unique matches of a feature with the supplier's history;

N– the total number of unique values of a feature in the procedure.

5. Calculation of the weighted sum of all the features of the tenders. The weight determines the importance of each feature in the final assessment, and their determination is a separate task that directly affects the quality of the prediction.

6. Ranking of procedures by weighted sums of feature values. The higher the value, the more suitable this procedure is for a specific supplier. Ten procedures with the maximum value are recommended to the supplier.

The algorithm was implemented in Python, mainly using the *numpy* libraries, and partly *sklearn*, *catboost*, *pytorch*.

3. Development and testing of a prototype of a recommender system

Here is a final list of features on the basis of which the recommendation for a specific supplier will be ranked:

- the presence of a condition that the purchase is intended for small and medium-sized businesses;
- customer;
- the fact of at least one win of the supplier with a given customer;
- customer region;
- the fact of at least one win of the supplier in the customer's region;
- regulatory framework (44-FZ, 223-FZ, commercial procurement);
- the fact of at least one win of the supplier within the framework of the relevant regulatory framework;
- sphere of activity (according to the full OKPD code);
- the fact of at least one win within the relevant field of activity (according to the full OKPD code);
- combining together the sphere of activity and the region of the customer (interpreted as the participation of the supplier in the relevant region in a tender in a specific sphere and, similarly, the fact of at least one win);
- combining together the sphere of activity and the customer (interpreted as the supplier's participation in the tender from the corresponding customer in a specific sphere and, similarly, the fact of at least one win);
- combining together the sphere of activity and the region of the customer (interpreted as the participation of the supplier in the relevant region in a tender in a specific sphere and, similarly, the fact of at least one win);
- combining together the area of activity and the site

on which the notice is published (interpreted as the supplier's participation in a tender on the relevant site in a specific area and, similarly, the fact of at least one win).

The main quality metric of the content filtering model in this case is *recall at K* (*r@K*), that is, sensitivity (completeness) on *K* elements [9]. We will consider such an outcome as a "positive forecast" when the supplier has participated in the recommended tender out of 10 recommended ones (*r*@10). Using sensitivity as the target metric, we strive to increase the number of real recommendations to different suppliers. In this case, these 10 procedures are determined by ranking by the highest probability of participation for a specific supplier. From the point of view of business logic, the most important tender on this list is the very first one as it is the one that the user wants to look through in most cases.

Determining the weights of features plays a key role in calculating the rating. The training data contains information about the participation of suppliers in potentially interesting (pre-filtered) procedures: "1" – participated, "0" – the opposite case. The task of the ranking algorithm is to rate the tenders for a specific supplier so that the probability of participation is maximum among the first ten. In other words, it evaluates which tenders out of the first ten recommended the supplier at least participated in, and then *r*@10 is calculated. Two approaches were used to calculate the rating:

- modeling the probability of participation of a specific supplier in a specific procedure and ranking according to the probability estimates obtained;
- ranking by a weighted sum of values, and the methods for determining the weights may be different.

In the basic version, when building a decision tree model using *sklearn*, the *r*@10 value is 0.21[4].

The following experiments were conducted using the basic model:

1) calculation of the weighted sum (the method for obtaining weights is the sequential exclusion of each feature to identify the most significant ones and setting weights in accordance with the subsequent change in the key metric (*r*@10 = 0.23));

2) similar to the first point the initial values of the features were pre-normalized (r@10 = 0.25);

3) weighted sum calculation (weighting method – Bayesian optimization, target metric – probability of participation in the first ten tenders (*r*@10 = 0.22)).

There was no significant change in the metric; the best option was the second one. Then experiments were conducted with different classification models to improve the metric:

1) random forest (sklearn [21]), *r*@10 = 0.306;

2) gradient boosting on decision trees (catboost [22]), *r*@10 = 0.331;

3) fully connected neural network with one hidden layer, built on *PyTorch* [23], *r*@10 = 0.355.

Random Forest algorithm is an ensemble method based on many decision trees. Each tree is built on a random subsample of the training data (with repetitions), and a random set of features at each node is used for splitting. Gradient Boosting Decision Trees (GBDT) is a modification of the algorithm in which trees are built one after another, each new tree correcting the errors of the previous one [24].

Figure 1 shows an illustration of a neural network built in accordance with point (3). For clarity, the number of inputs was taken to be three. The hidden layer was designed to highlight the most significant features. As the means of activation, ReLU (Rectified Linear Unit) was used. The number of neurons in the hidden layer can be random; in this case their number

[4] Here and below, similar estimates were obtained on a training sample with cross-validation $k = 5$.

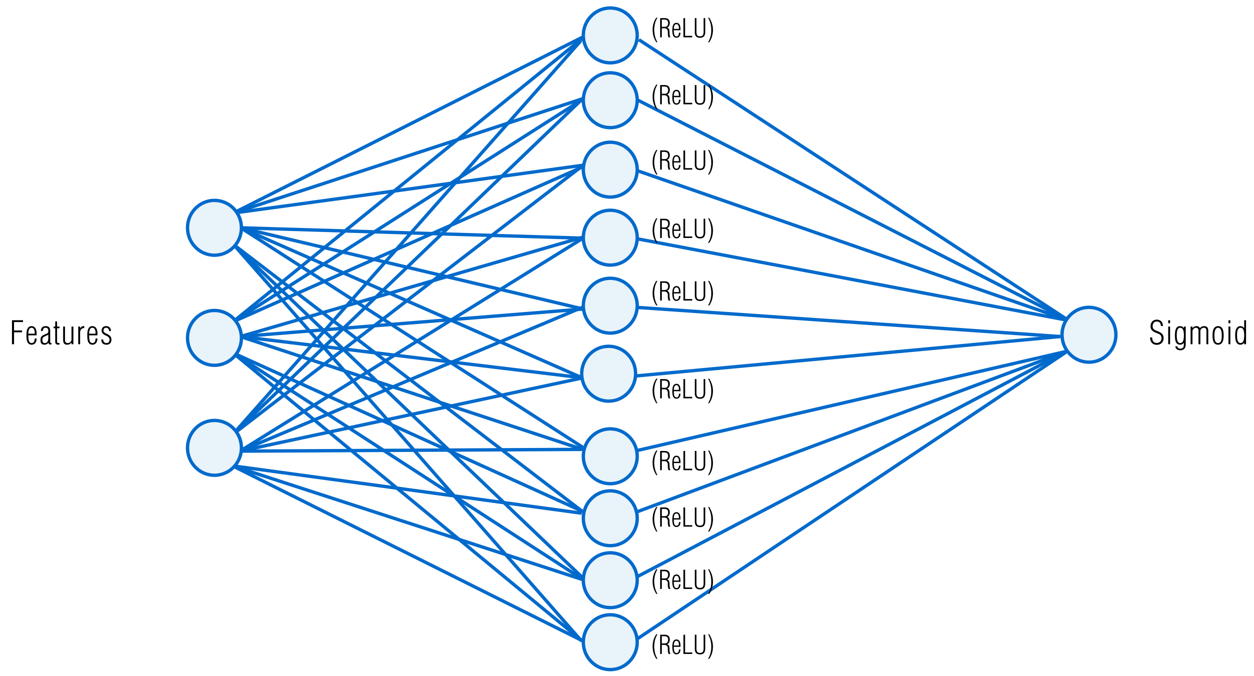


Fig. 1. Architecture of a fully connected neural network for predicting supplier participation.

based on the results of the experiments was taken to be 10. The activation function of the output layer is sigmoid (logistic function) and returns a number in the range from 0 to 1.

The neural network was trained using the Focal Loss function. This is a modification of the cross-entropy function usually used in problems with highly unbalanced classes which reduces the weight for easily classified ones [25]. For each batch (data batch), this function was calculated using formula (2):

$$\text{FocalLoss} = -\frac{1}{N}\sum_{i=1}^{N}\alpha_{t_i}\left(1-p_{t_i}\right)^{\gamma}\log\left(p_{t_i}\right), \tag{2}$$

where N is the number of examples in the batch;

α_t is the weighting coefficient for the correct class;

p_t is the predicted probability of the correct class;

γ is the focusing parameter.

The relatively low values of the metric can be explained by the specific field of activity: since the click-rate is of primary interest for the research, it can be estimated that the average supplier interacts with the "content" approximately two times a year. To compare it with the Netflix service in the USA: it is known that in 2019, its catalog contained [26] 47 000 episodes of TV series, 4 000 films and the number of subscribers in the second quarter was 60.1 million people [27]. In 2024, about 7 000 films and TV shows were available [28] (the exact number of episodes of the series was not disclosed, but more than 10 000 new episodes were added in 2024 [29]) to 90 million subscribers (USA and Canada) [30]. At the same time, it is unknown how many "requests" a user makes on average per year, but it can be assumed that at least one (in reality the number must be much higher). Even with such a rough estimate, the average number of interactions with content here is 118 times a year.

It should be emphasized that the behavior of users in the procurement market has its own specifics, which differs quite significantly from the standard areas of recommender systems application. The distribution of activity among suppliers is uneven — within some OKPD2 categories, the number of participations per year can exceed 1 000. At the same time, any actions of the supplier on the tender page (views, downloads, etc.) can be considered as a fact of "activity."

As confirmed in experiments, the larger the pool of purchases potentially suitable for a supplier, the more difficult it is for relevant tenders to get into the top 10. On average, a supplier can actually be suitable for quite a few procedures, so the complexity of ranking increases. The decisive factor is the interaction between suppliers and customers.

Nevertheless, the benefits of implementing a recommendation system in the work of the platform were substantiated as part of A/B testing. For this purpose, two similar in number groups of suppliers working on the trading platform according to OKPD2 27 were identified. Testing was carried out at 10 weekly intervals at the end of 2023. Each member of the groups received an e-mail containing the top ten recommended tenders. For each e-mail, the number of openings and transitions from the e-mail to the platform was counted. At intervals 6–9, recommendations were sent out according to the described methodology. The rest of the time, the mailing was also carried out, but the recommendations were formed in a "naïve" way: the supplier was offered 10 random purchases that were announced in his region, with activity profiles in which he already worked and the price category of which suited him.

It was necessary for the composition of these groups to be homogeneous. That is why they were selected so that each had approximately the same proportion of active and inactive clients and approximately the same proportion of preferred price categories of purchases. It was also necessary that the average frequency of opening e-mails with "naïve" recommendations sent before the experiment was not supposed to be significantly different (in fact, the difference was no more than 2%), and the groups themselves were practically identical in number.

Group A suppliers were emailed every two weeks with a list of 10 tenders recommended for each to participate.

The results are presented in *Figs. 2* and *3*.

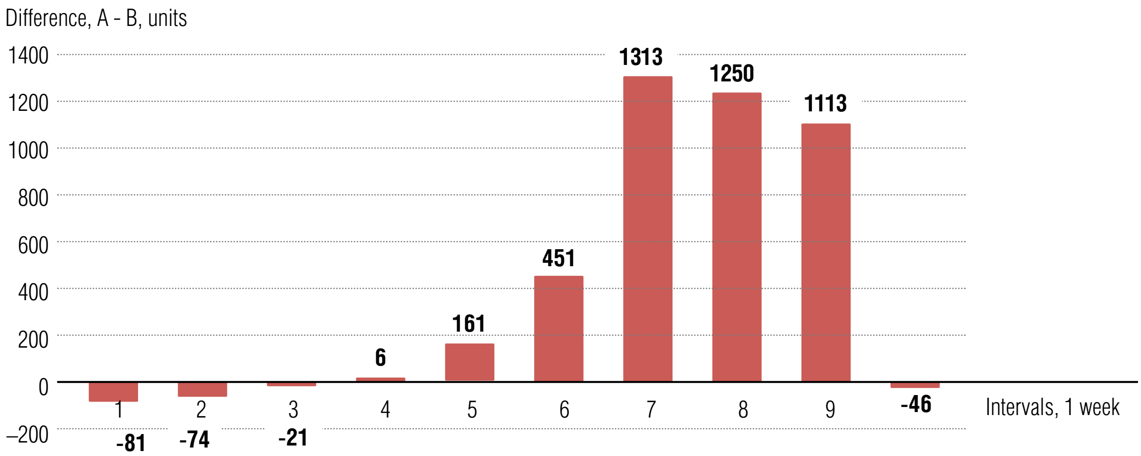


Fig. 2. Difference between the number of views of the letter with recommendations for group A and the same indicator for group B.

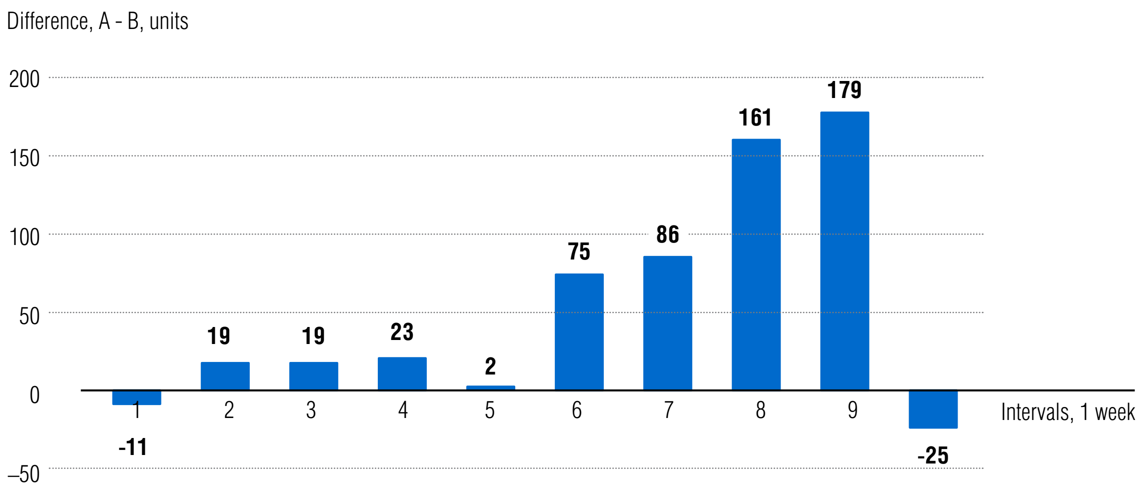


Fig. 3. Difference between the number of clicks to the site from a letter with recommendations for group A and the same indicator for group B.

Thus, it can be concluded that the recommendation emails were perceived positively and with interest by clients, stimulating them to additional actions, and contributing to their information. Due to this, positive effects were obtained for the competitive environment in the tenders (auctions).

4. Results and discussion

The proposed developments have already been partially implemented in the work of the trading operator JSC EETP (Roseltorg), which ensured an increase in the number of successfully completed procedures by 3.7% [31]. In general, customers improved the effectiveness of supplier searches by 40%. Of course, this also works in the opposite direction: if customers find suppliers for contracts and agreements more often, then entrepreneurs are more likely to open new relevant sales markets for themselves and are less likely to encounter problems with contract termination. However, despite the positive experience and optimistic prospects, the authors also note the risks of implementing recommendation services. Firstly, with the rise in the number of parameters on the basis of which recommendations are built, there is a risk of increasing the "noise level" of the model. Secondly, developers can spend significant time and computing power on collecting, processing and storing secondary characteristics without identifying a priority group of parameters due to which the most accurate recommendation is formed. To avoid such risks, industry experts should be involved to adjust the substantive part of the development. Another group of risks is associated with overtraining of the system, which can occur due to an imbalance in the initial data (for example, due to the popularity of a number of categories of the OKPD2 classifier), excessive complexity of the model (then the algorithm will "memorize" individual preferences of the most active clients instead of identifying common features).

It is also impossible to underestimate the risks of the "human factor" – errors during A/B testing, which is aggravated by the high sensitivity of a system to error; the high cost and complexity of technical support and staffing of such developments.

Conclusion

Thus, as a result of the research carried out, the authors' team completed the following tasks:

- formed a hypothesis on the feasibility of using recommender systems to improve the efficiency of state, municipal and corporate procurement (regulated procurement);
- studied the typologies of systems and substantiated the choice of a specific class of systems that is most relevant for developing recommendations for participants in the regulated procurement market;
- developed a prototype of a recommender system, for which he substantiated the methodology for its construction and the data structure for its content;
- tested the prototype on procurement for the electric power industry by sending out recommendations to participate in specific tenders to entrepreneurs who could potentially, due to their market position, both win the procedure and effectively fulfill the contract.

The research carried out allowed us, firstly, to expand the problems of assessing and improving the efficiency of regulated procurement. Currently, efficiency in this area is determined by the level of competition (the number of applications per 1 procedure) and the resulting savings (how much cheaper than the initial declared price it was possible to purchase). However, qualitative indicators such as personal preferences of procurement participants, factors for choosing certain procedures or customers remained poorly studied. Personalized recommendations allow for a better study of market participants' moods and capabilities, and make it possible to improve the efficiency of trades.

Secondly, the results of the study were significant for science and practice. In particular, a methodology was developed for building recommendation services for government needs, rather than for solving purely commercial problems. In the future, such a methodology can be replicated in other areas where the government is a counterparty. As for practice, it is planned to transform the prototype into a full-fledged system and subsequently fully integrate it into the work of the trade operator.

Further research areas, according to the authors, may be more applied in nature, focused on customizing recommendation services for the specific industries, as well as studying the capabilities of other classes of recommendation systems. ■

Acknowledgments

This study was carried out with financial support from the Russian Science Foundation (project 23-28-01644).

References

1. Garant System (2014) *Resolution of the Government of the Russian Federation of April 15, 2014 No. 320 "On approval of the state program of the Russian Federation "Public Finance Management and Regulation of Financial Markets" (with amendments and additions)*. Available at: https://base.garant.ru/70644234 (accessed 10 October 2024) (in Russian).

2. Denisova A.I., Sozaeva D.A., Gonchar K.V., Aleksandrov G.A. (2024) Analysis of factors for failure to close tenders for public procurement of drugs. *Problems of Forecasting*, no. 2, pp. 178–191 (in Russian). https://doi.org/10.47711/0868-6351-203-178-191

3 Ministry of Finance of Russia (2024) *Consolidated analytical report on the results of monitoring the*

procurement of goods, works, services to meet state and municipal needs, as well as the procurement of goods, works, services by certain types of legal entities for 2023. Available at: https://minfin.gov.ru/common/upload/library/2024/05/main/Svodnyy_2023.pdf (accessed 10 October 2024) (in Russian).

4. Ministry of Finance of Russia (2024) *Quarterly report on the results of monitoring the procurement of goods, works, services to meet state and municipal needs, as well as the procurement of goods, works, services by certain types of legal entities for the second quarter of 2024*. Available at: https://minfin.gov.ru/common/upload/library/2024/07/main/II_kvartal_2024_g.pdf (accessed 10 October 2024) (in Russian).

5. Kuznetsov M. (2024) What awaits the public procurement market in 2024. *Rossiyskaya Gazeta – Economy of the Central District*, no. 4(9246). Available at: https://rg.ru/2024/01/11/reg-cfo/delo-za-malym.html (accessed 10 October 2024) (in Russian).

6. Yakovlev A.A., Balaeva O.N., Rodionova Yu.D., Tkachenko A.V. (2020) Problems and efficiency of public procurement through the eyes of their participants. *All-Russian economic journal ECO*, no. 11, pp. 83–103 (in Russian). https://doi.org/10.30680/ECO0131-7652-2020-11-83-103

7. Falk K. (2020) *Recommender systems in practice*. Moscow: DMK Press (in Russian).

8. Baranov V.V. (2005) *Processes of making management decisions motivated by interests*. Moscow: Fizmatlit (in Russian).

9. Kravchenko T.K., Isaev D.V. (2024) *Decision support systems: textbook and practical training for universities. 2nd edition, revised and additional edition*. Moscow: Yurait (in Russian).

10. Keeney R.L., Raiffa H. (1976) *Decisions with multiple objectives: Preferences and value tradeoffs*. John Wiley & Sons.

11. Thaler R. (2022) *New behavioral economics. Why people break the rules of traditional economics*. Moscow: Eksmo (in Russian).

12. García Rodríguez M.J., Montequín V., Ortega-Fernández F., Balsera J. (2020) Bidders recommender for public procurement auctions using machine learning: data analysis, algorithm, and case study with tenders from Spain. *Complexity*, article 8858258. https://doi.org/10.1155/2020/8858258

13. Hickok M. (2024) Public procurement of artificial intelligence systems: new risks and future proofing. *AI & Society*, vol. 39, pp. 1213–1227. https://doi.org/10.1007/s00146-022-01572-2

14. Seliverstov Y.A., Komissarov A.A., Lesovodskaya A.A., et al. (2022) Development and research of models of multi-class classifiers for a recommendrd system for preparing applications on the e-procurement. *Computing, Telecommunications and Control*, vol. 15, no. 2, pp. 43–62 (in Russian). https://doi.org/10.18721/JCSTCS.15204

15. Chen J., Dong H., Wang X., et al. (2023) Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, vol. 41, no. 3, article 67. https://doi.org/10.1145/3564284

16. Ko H., Lee S., Park Y., Choi A. (2022) A survey of recommendation systems: Recommendation models, techniques, and application fields. *Electronics*, vol. 11, no. 1, article 141. https://doi.org/10.3390/electronics11010141

17. Andrianov N.A., Atakhodjaeva M.R., Borodin E.I. (2022) Mathematical modeling of a recommender system and data processing of a telecommunications company using machine learning models. *Bulletin of the South Ural State University. Series: Computer technologies, management, radio electronics*, vol. 22, no. 2, pp. 17–28 (in Russian). https://doi.org/10.14529/ctcr220202

18. Denisova A.I., Sozaeva D.A., Gonchar K.V., Aleksandrov G.A. (2023) Improving the methodology for assessing the economic efficiency of public e-procurement of medicines. *Financial Journal*, vol. 15, no. 4, pp. 63–81 (in Russian). https://doi.org/10.31107/2075-1990-2023-4-63-81

19. Denisova A.I., Sozaeva D.A., Gonchar K.V., Aleksandrov G.A. (2024) Analysis of factors for non-closure of tenders for public procurement of medicines. *Problems of Forecasting*, no. 2, pp. 178–191 (in Russian). https://doi.org/10.47711/0868-6351-203-178-191

20. ConsultantPlus (2024) *Resolution of the Government of the Russian Federation of 08.06.2018 N 656 (as amended on 23.12.2024) "On the requirements for operators of electronic platforms, operators of specialized electronic platforms, electronic platforms, specialized electronic platforms and the functioning of electronic platforms, specialized electronic platforms, confirmation of compliance with such requirements, on the loss of the status of an operator of an electronic platform, operator of a specialized electronic platform by a legal entity"*. Available at: https://www.consultant.ru/document/cons_doc_LAW_300361/ (accessed 10 October 2024) (in Russian).

21. scikit-learn: Machine Learning in Python (2025) *RandomForestClassifier*. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html (accessed 11 February 2025).

22. CatBoost library (2025) *CatBoostClassifier*. Available at: https://catboost.ai/docs/en/concepts/python-reference_catboostclassifier (accessed 11 February 2025).

23. PyTorch (2025) *torch.nn*. Available at: https://pytorch.org/docs/stable/nn.html (accessed 11 February 2025).

24. scikit-learn: Machine Learning in Python (2025) *Ensembles: Gradient boosting, random forests, bagging, voting, stacking*. Available at: https://scikit-learn.org/stable/modules/ensemble.html (accessed 11 February 2025).

25. Lin T.-Y., Goyal P., Girshick R., He K., Dollár P. (2017) Focal loss for dense object detection. *2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy*, pp. 2999–3007. https://doi.org/10.1109/ICCV.2017.324

26. Statista (2025) *Content available on Disney's new streaming service's (Disney+) compared to Netflix in the United States as of April 2019*. Available at: https://www.statista.com/statistics/1041104/disney-catalog-netflix-comparison-us/ (accessed 10 October 2024).

27. Safronov A. (2019) Netflix US subscribers fall for first time in 8 years. *RBC Investments*. Available at: https://quote.ru/news/article/5d3013649a794752c4a31fee (accessed 10 October 2024).

28. Morley M. (2025) Netflix library by the numbers 2024: 589 new originals released and library swells to over 7,000 titles. *What's on Netflix*. Available at: https://www.whats-on-netflix.com/news/netflix-library-by-the-numbers-2024/ (accessed 10 May 2025).

29. Moore K. (2025) Netflix added over 10,000 episodes of TV in 2024. *What's on Netflix*. Available at: https://www.whats-on-netflix.com/news/netflix-added-over-10000-episodes-of-tv-in-2024/ (accessed 10 May 2025).

30. Statista (2025) *Number of Netflix paying streaming subscribers worldwide as of 4th quarter 2024, by region*. Available at: https://www.statista.com/statistics/483112/netflix-subscribers/ (accessed 10 May 2025).

31. Alliance in the field of artificial intelligence (2020) *Recommender system in the field of electronic procurement*. Available at: https://ai-russia.ru/library/roseltorg-system (accessed 10 October 2024).

About the authors

Anna I. Denisova

Candidate of Sciences (Economics);

Senior Lecturer, Institute of Information Systems, Department of Mathematical Methods in Economics and Management, State University of Management, 99, Ryazansky Ave., Moscow 109542, Russia;

E-mail: aidenisova@inbox.ru

ORCID: 0000-0002-2296-237X

Dzhamilya A. Sozaeva

Doctor of Sciences (Economics), Associate Professor;

Head of Laboratory, Project-training Laboratory for Research and Development in the Field of Digitalization of Management, State University of Management, 99, Ryazansky Ave., Moscow 109542, Russia;

Associate Professor, Department of Valuation Activities and Corporate Finance, Synergy University, 80, Bldg. G, Leningradsky Ave., Moscow 125190, Russia;

E-mail: dasozaeva@gmail.com

ORCID: 0000-0003-3955-891X

Konstantin V. Gonchar

Junior Researcher, Project-training Laboratory for Research and Development in the Field of Digitalization of Management, State University of Management, 99, Ryazansky Ave., Moscow 109542, Russia;

Senior Lecturer, Department of IU-3, Bauman Moscow State Technical University, 5, 2-nd Baumanskaya St., Moscow 105005, Russia;

E-mail: goncharkv@gmail.com

ORCID: 0000-0001-7345-3340

DOI: 10.17323/2587-814X.2025.2.41.53

A customer avatar model based on Kolmogorov–Arnold networks

Fedor V. Krasnov
E-mail: krasnov.fedor2@rwb.ru

Fedor I. Kurushin
E-mail: kurushin.fedor@rwb.ru

Research Center of LLC WILDBERRIES SK, Moscow, Russia

Abstract

The increasing pace of development of e-commerce continues to present new challenges in terms of personalizing product search and recommendations. Monolithic search and recommendation systems have become cumbersome and are unable to effectively address the need for a deeper understanding of users on electronic trading platforms (ETPs) despite having access to comprehensive information about their interests and purchase histories. Collaborative filtering mechanisms which are widely used suffer from a lack of diversity in offerings and a reduced capacity to surprise users. Additionally, the low frequency of recommendation updates and the replacement of "personalized" with "similar to others" concepts contribute to these issues. We have approached the resolution of these issues by developing a shopping assistant named "Ellochka" that is individual for each user of ETP. The digital avatar model of the user continually searches for relevant products based on their history of interaction with ETP. We were guided by the principle of independence – avatar models do not share information with each other. When a new user joins, they are assigned a unique avatar model that evolves independently. Each avatar has its own language to generate search queries. The level of complexity of each avatar can vary depending on the intensity of its interaction with ETP. Continued interaction with the avatar allows for tracking of optimal purchase conditions, reminding users of expiration dates and the need for re-purchasing frequently purchased items. Isolating the avatar allows it to be retrained after each event, without significantly impacting the overall search and recommendation system. The use of neural network architecture-based and Kolmogorov–Arnold networks in the avatar-model has led to improvements in the main indicators of search and recommendation effectiveness, namely, novelty and diversity.

Keywords: large language models, product search, search query recommendations, search query transformation, user intent determination, text analysis, machine learning, e-commerce

Citation: Krasnov F.V., Kurushin F.I. (2025) A customer avatar model based on Kolmogorov–Arnold networks. *Business Informatics*, vol. 19, no. 2, pp. 41–53. DOI: 10.17323/2587-814X.2025.2.41.53

Introduction

Artificial intelligence is increasingly penetrating scientific methodology and everyday life. On the one hand, approaches to the classification of scientific knowledge are being revised; for example, a section "28.23: artificial intelligence" has appeared in the Code of State Categories Scientific and Technical Information. On the other hand, in an online store of women's clothing, the science-intensive processes are comparable to office work in geology [1].

This paper presents an avatar model of a personal assistant for making purchases on electronic trading platforms (ETP) called "Ellochka," in honor of the heroine of the work by Ilf and Petrov "12 Chairs". In this novel, Ellochka is a beautiful and spoiled young woman who lives for her own pleasure at her husband's expense and is mainly engaged in purchasing things.

Personal assistants based on artificial neural networks of deep learning have become a familiar attribute of customer service in the logistics, banking and e-commerce industries. However, the "personality" of such an assistant is limited only by the temporary desire to save resources during the maintenance session. The user's personal preferences and moods do not affect the operation of the system in any way. The personal assistant will not be able to continue the interrupted conversation from the same place and, most likely, will forget its beginning.

Personal assistants and search and recommendation models in e-commerce are based on large language models (LLM) based on the transformer architecture [2].

However, each large language model contains a dictionary file which it uses further in neural network structures in the form of ordinal numbers of tokens from the dictionary. This is because linear algebra works with models of numbers rather than string variables.

The size of the LLM dictionary largely determines the number of model parameters, because it is one of the main dimensions of the LLM along with the size of the model context window. *Table 1* shows the sizes of the most effective dictionaries according to the MTEB benchmark [3].

Table 1.

Industry research indicators

Model	Vocabulary size	Number of Russian words
alan-turing-institute/mt5-large-finetuned-mnli-xtreme-xnli	250 100	26 427
ai-forever/rugpt3large_based_on_gpt2	50 257	43 213
aeonium/Aeonium-v1-Base-4B	128 000	102 913

Search and recommendation systems are highly reliable information systems that ensure the continuity of the service process and minimal delays in interacting with buyers and sellers.

Speed in serving search queries is achieved through horizontal scaling of computing resources. A specific computing resource is not assigned within a single user service session. This leads to the need to synchronize user actions that change, for example, the balance, stock balances or the contents of the customer's shopping cart. These digital objects exist in the singular and are managed centrally.

The main contribution of this study is the formulation and testing of a new, decentralized approach to building search and recommendation systems in e-commerce, focusing on the avatar model of the buyer (*Fig. 1*).

The second most important contribution of this study is the expansion of the approach to building artificial deep learning networks for time series processing by a new class of Kolmogorov–Arnold networks using wavelet transformations.

1. Methodology

In order to avoid the well-known problems with novelty and diversity in product recommendations [4–7], in this study, we propose to consider the rec-

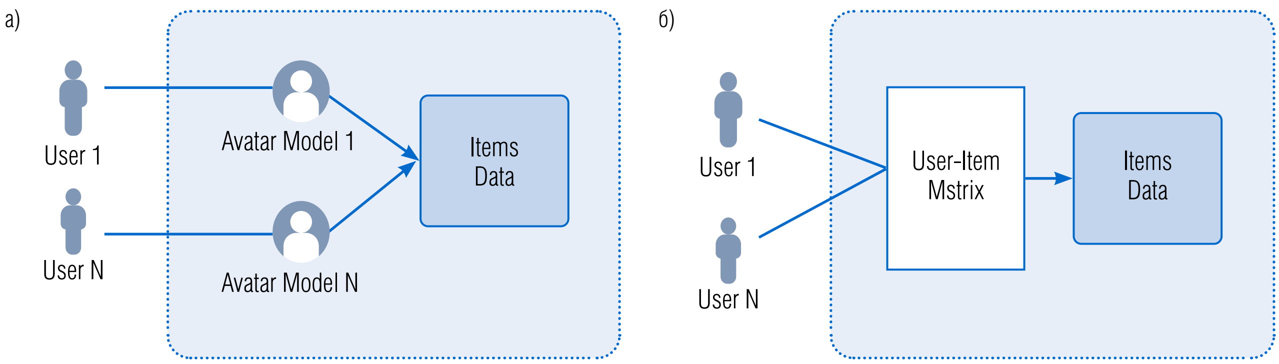


Fig. 1. The scheme of the recommendations:
(a) Decentralized (proposed in this study), (b) Centralized.

ommendations of search queries that lead to the right products. In everyday terms, you recommend how to search for a product, rather than trying to guess the desired product. A similar approach to recommendations using text is proposed in [8], in which the authors receive a significant improvement in metrics by rejecting product recommendations by their number, as in [9, 10].

Another well-known problem of generative models based on LLM is hallucination. In order to overcome the negative effects of hallucination, various methods are used to regularize the output of the model. For example, the reflection of the model on itself [11] or the use of adaptive spectral normalization [12]. The authors of this study have reduced the negative effect of hallucination due to the fact that the generated search query text is not intended for a human, but for a search engine, in other words, for another machine learning model. This made it possible to reduce the requirements for vocabulary and coherence of the generated search query text, since the presence of tokens is sufficient for a correct semantic search for products; in turn, the inconsistency of tokens by case or repetition affects the relevance of the search only slightly.

The avatar model is based on the paradigm of information search – the composition of candidate and ranking models [13]. The authors of this study developed this approach and, instead of a ranking model, created a model for selecting pairs of candidate queries with fixed five levels of relationships: "narrowing," "expanding," "paraphrasing," "different characteristics," "related products." This made it possible to create more customer-focused interaction scenarios, depending on the presence or absence of purchases in the interaction session with the ETP.

Modern tokenization models in LLM are based on compressed dictionaries from character n-grams [14, 15]. This approach solves the problem of missing a word in the dictionary and reduces memory consumption, but tokenization is ambiguous – the same sentence can be tokenized into different sets of tokens. This property is used in LLM training as bootstrapping for a more uniform distribution of the back propagation of the error. For the avatar model, there is no need to solve the problem of tokenizing new words, since the avatar model uses all tokens from the user's vocabulary. Moreover, the avatar model uses dictionary n-grams up to four words long to receive events from token structures. For example, from brand names consisting of several words, such as "Vans off the Wall." Obviously, the search query "Vans off the Wall" should not be used to search for products with the tokens "vans" and "wall."

When updating the avatar model, the new tokenization model is created, since the avatar model has 103 times fewer training parameters than, for example, the GPT2 (120 million) model.

2. Kolmogorov–Arnold networks

Kolmogorov–Arnold networks (KAN) show promising results in various generative models [16] and time series models [17]. We will show the advantages of KAN in comparison with the Multilayer Rumelhart Perceptron (MLP). Let's give an MLP with dimension n at the input and m at the output, containing fully connected layers of an artificial neural network from l to $l + 1$. Then the MLP equation in matrix form is given by formula (1):

$$x^{(l+1)} = W_{l+1,l}\, x^{(l)} + b^{(l+1)}, \quad (1)$$

where $W_{l+1,l}$ is the weight matrix connecting the layer and the layer ;

$x^{(l)}$ is the input vector;

$x^{(l+1)}$ is the output vector;

$b^{(l+1)}$ is the bias for $l + 1$.

Weight matrix $W_{l+1,l}$ can be represented as follows:

$$W_{l+1,l} = \begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,2} & \cdots & w_{m,n} \end{pmatrix}, \quad (2)$$

where $w_{i,j}$ is the weight between the i node in the $l+1$ layer and the j node in the l layer, $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$.

Then the bias vector $b^{(l+1)}$ can be presented as follows:

$$b^{(l+1)} = \begin{pmatrix} b_1^{(l+1)} \\ b_2^{(l+1)} \\ \vdots \\ b_m^{(l+1)} \end{pmatrix}. \quad (3)$$

Thus, in the expanded form, equation (1) looks like this:

$$\begin{pmatrix} x_1^{(l+1)} \\ x_2^{(l+1)} \\ \vdots \\ x_m^{(l+1)} \end{pmatrix} = \begin{pmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,2} & \cdots & w_{m,n} \end{pmatrix} \begin{pmatrix} x_1^{(l)} \\ x_2^{(l)} \\ \vdots \\ x_n^{(l)} \end{pmatrix} + \begin{pmatrix} b_1^{(l+1)} \\ b_2^{(l+1)} \\ \vdots \\ b_m^{(l+1)} \end{pmatrix}. \quad (4)$$

Now let's assume that we have L layers, each of which has a structure (4). Let's say $\sigma(\cdot)$ is an activation function. Then the compact formula for the entire MLP network $f(x)$, where x is the input vector, and $f(\cdot)$ is the MLP, is given as follows (5):

$$f(x) = x^{(L)}. \quad (5)$$

where

$$x^{(l+1)} = \sigma(W_{l+1,l}\, x^{(l)} + b^{(l+1)}), \quad (6)$$

for $l = 0, 1, 2, ..., L-1$, $x^{(l+1)}$ denotes the output vector.

This is equivalent to the following representation (7):

$$f(x) = \sigma\begin{pmatrix} W_L \sigma(W_{L-1} \cdots \sigma(W_2 \sigma(W_1 x + b_1) + \\ + b_2) \cdots + b_{L-1}) + b_L \end{pmatrix}. \quad (7)$$

Now let's look at how relationships between layers are created in KAN. Let $x^{(l)}$ be a vector with dimension n. Then the transposed vector can be represented as follows:

$$x^{(l)} \in R^n \Rightarrow (x^{(l)})^{\mathrm{T}} \in R^{1\times n}.$$

Let's denote the matrix of transposed vectors $x^{(l)}$ with m rows and n columns as X_l:

$$X_l = \begin{pmatrix} (x^{(l)})^{\mathrm{T}} \\ (x^{(l)})^{\mathrm{T}} \\ \vdots \\ (x^{(l)})^{\mathrm{T}} \end{pmatrix} \in R^{m\times n}.$$

Each row of the matrix X_l – is a transposed vector $(x^{(l)})^{\mathrm{T}}$. Now let's define the operator A_o, that acts on the matrix $\Phi_{l+1,l}(X_l)$. This operator summarizes the elements of each row of the matrix and outputs the resulting vector r. Thereby definition formula of A_o:

$$A_o(\Phi_{l+1,l}(X_l)) = r,$$

where r vector, derived from the following expression (8):

$$r_i = \sum_j \left[\Phi_{l+1,l}\left(X_l\right)\right]_{ij} = \sum_{j=1}^{n} \phi_{i,j}\left(x_j^{(l)}\right), \text{ for } i = 1, 2, ..., m. \quad (8)$$

In the expression (8), $\left[\Phi_{l+1,l}\left(X_l\right)\right]_{ij}$ corresponds to the element in the i row and j column of the matrix $\Phi_{l+1,l}(X_l)$. Thus, the operator A_o can be written as follows:

$$A_o\left(\Phi_{l+1,l}\left(X_l\right)\right) = \sum_j \left[\Phi_{l+1,l}\left(X_l\right)\right]_{ij}.$$

By definition A_o performs an action on a matrix $\Phi_{l+1,l}(X_l)$, summarizes the elements in each row, and outputs the resulting vector r. Indeed, Φ_{l+1} receives a vector at the input $x^{(l)}$ and outputs data at the output, where each element of $x^{(l)}$ is the sum of one element (9):

$$X_{l+1,l} = \Phi_{l+1,l}(X_l), \quad (9)$$

where:

$$\Phi_{l+1,l}\left(X_l\right) = \begin{pmatrix} \phi_{1,1}(x_1^{(l)}) & \phi_{1,2}(x_2^{(l)}) & \cdots & \phi_{1,n}(x_n^{(l)}) \\ \phi_{2,1}(x_1^{(l)}) & \phi_{2,2}(x_2^{(l)}) & \cdots & \phi_{2,n}(x_n^{(l)}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m,1}(x_1^{(l)}) & \phi_{m,2}(x_2^{(l)}) & \cdots & \phi_{m,n}(x_n^{(l)}) \end{pmatrix}. \quad (10)$$

In expression (10) $\Phi_{l+1,l}$ is the activation function between the layers l and $l + 1$. Each element $\phi_{i,j}(\cdot)$ denotes an activation function that connects the j neuron in the l layer with the neuron in the $l + 1$ layer. Instead of multiplication, equation (10) calculates a function with trainable parameters. Therefore, if we consider X_o as input data that contains only the input vector in the form of strings, then for the entire KAN, the output data after the L layers will be as follows (11):

$$
\begin{aligned}
f_{KAN}\left(X_0\right) &= x^{(L)} = A_o\left(\Phi_{L,L-1}\left(X_{L-1}\right)\right) = \\
&= A_o\left(\Phi_{L,L-1}\left(\begin{pmatrix} \left(A_o\left(\Phi_{L-1,L-2}(X_{L-2})\right)\right)^{\mathrm{T}} \\ \left(A_o\left(\Phi_{L-1,L-2}\left(X_{L-2}\right)\right)\right)^{\mathrm{T}} \\ \vdots \\ \left(A_o\left(\Phi_{L-1,L-2}\left(X_{L-2}\right)\right)\right)^{\mathrm{T}} \end{pmatrix}\right)\right) = \\
&= A_o\left(\Phi_{L,L-1}\left(\begin{pmatrix} \left(A_o\left(\Phi_{L-1,L-2}\cdots\left(A_o\left(\Phi_{1,0}\left(X_0\right)\right)\right)\right)\right)^{\mathrm{T}} \\ \left(A_o\left(\Phi_{L-1,L-2}\cdots\left(A_o\left(\Phi_{1,0}\left(X_0\right)\right)\right)\right)\right)^{\mathrm{T}} \\ \vdots \\ \left(A_o\left(\Phi_{L-1,L-2}\cdots\left(A_o\left(\Phi_{1,0}\left(X_0\right)\right)\right)\right)\right)^{\mathrm{T}} \end{pmatrix}\right)\right)
\end{aligned} \quad . \quad (11)
$$

Thus, traditional MLPs use fixed nonlinear activation functions at each node, linear weights and biases to transform input data across layers. The output data on each layer is calculated using a linear transformation followed by a fixed activation function. During the backpropagation of the error, the gradients of the loss function relative to the weights and biases are calculated to update the model parameters. In contrast, KAN replaces linear weights with one-dimensional functions that can be trained by placing them on edges rather than on nodes. Wavelet transformations are most often used as one-dimensional functions: MHAT wavelet ("Mexican hat"), Shannon wavelet, Morlet wavelet, Gauss wavelet (*Table 2*).

One-dimensional functions from previous levels are summed up in the nodes. Each function can be adapted, which allows KAN to train both activation and transformation of input data. This change leads to increased accuracy and interpretability, since KAN can better approximate functions using fewer parameters. During backpropagation of an error in KAN, gradients are calculated relative to one-dimensional functions, updating them to minimize the loss function. This leads to more efficient learning for complex and multidimensional functions.

Table 2.

Wavelets used as one-dimensional functions

Name	Formula
MHAT-wavelet ("Mexican hat")	$\psi(t) = \frac{-d^2}{dt^2} e^{-t^2/2} = \left(1 - t^2\right) e^{-t^2/2}$
Shannon wavelet	$\psi_k^n(t) := 2^{n/2} \psi^{(Sha)}\left(2^n t - k\right)$, where $\psi^{(Sha)}(t) := \frac{\sin\left(\pi t\right)}{\pi t}$
Morlet wavelet	$\psi_\sigma(t) = c_\sigma \pi^{-1/4} e^{-t^2/2}\left(e^{i\sigma t} - k_\sigma\right)$, where $k_\sigma = e^{-\sigma^2/2}, c_\sigma = \left(1 + e^{-\sigma^2} - 2e^{-3\sigma^2/4}\right)^{-\frac{1}{2}}$
Gauss wavelet	$\psi(t) = \frac{-d}{dt} e^{-t^2/2}$

3. A model for reducing the diversity of search queries

In the field of e-commerce, there is the problem of the "long tail" of search queries, which makes it difficult to create effective models. This is due to a wide variety of queries, typos, synonyms and slang.

Modern approaches to query reformulation use neural network methods. By highlighting text factors, vector representations of search queries are constructed, and then the "nearest neighbors" are searched in vector space using Retrieval Augmented Generation (RAG) language models [18] and BERT transformer models [19].

Product search allows you to use user behavioral data from making purchases. Thus, a collaborative paradigm is used. The user-product relationship determines the sequences that led to the purchase and uses them as options for the current user [20]. In the present study, the problem of query reformulation (QR) is considered from the perspective of the "search query – product" relationship for the application of a surrogate function [22] in the avatar model.

In addition to the reformulation of requests, the task is also to determine the type of connection "request – request" and "request – product". As an example, demonstrating the various types of relationships identified by the authors, we present *Table 3.*

For a more visual representation, let's denote the set of all search queries as Q. The approximate number of elements in this set is $Q_v \approx 10^{10}$, which is comparable to the stream of events analyzed in the search for the Higgs boson [21].

It is known from the study [23] that 98% of all purchases are made using a limited set of queries, which we will denote as Q_{HPQ}. Without limiting generality, we can assume that Q_{HPQ} does not depend on time. The remaining set of queries is denoted as Q_{LPQ}, where HPQ and LPQ are common abbreviations for high—performing and low-performing queries, respectively.

Based on the definition of the set Q, we can write the following equation:

$$Q = Q_{HPQ} \cup Q_{LPQ}, \tag{12}$$

$$\varnothing = Q_{HPQ} \cap Q_{LPQ}. \tag{13}$$

Table 3.

Examples of relation between the pairs of queries

Source Query	Suggest	Scenario
Iphone 16 pro max	Apple iPhone 16 Pro Max	Paraphrased
Women's red Dress	Women's Dress	Expanding
Dress	Women's red Dress	Narrowing
Women red dress	Women's blue dress	Other characteristics
Women's evening dress	High-heeled shoes	Substituting
High-heeled shoes	Running shoes	Not relevant

It follows from the nature of the Q distribution that $|Q_{LPQ}| \gg |Q_{HPQ}|$. The task of reformulating queries can then be considered as a search for a function F such that $F(Q_{LPQ}) \Rightarrow Q_{HPQ}$. Consider the requests $q^{p_i}_{HPQ} \in Q_{HPQ}$ and $q^{p_i}_{LPQ} \in Q_{LPQ}$ resulting purchases of a product $p_i \in P$ from the entire product catalog P. Then we can write that there is a surrogate function F that results $q^{p_i}_{LPQ} q^{p_i}_{HPQ}$ for the product p_i.

$$F\left(Q^{p_i}_{LPQ}\right) \Rightarrow Q^{p_i}_{HPQ}, \text{ for } \forall p_i \in P. \tag{14}$$

Expression (14) allows us to formulate the conditions for obtaining a data set for using numerical methods of obtaining F, for example, using artificial neural networks of deep learning. To do this, you need to collect pairs $\left\{q^{p_i}_{LPQ}, q^{p_i}_{HPQ}\right\}$ from user search logs.

The Negative Batch Sampling approach was used to generate negative examples [10], with the number of negative examples varying from 5 to 12. Using a large number of negative examples improved the model's performance slightly, but required significantly more resources to train the model.

In *Fig. 2* shows a diagram of the avatar model in the form of a sequence of tokenization models, query reformulation (QR) models, and KAN.

4. Research questions

This research is applied in nature. The described methodology serves as a source for research questions and verification using numerical methods. When creating the methodology, the authors formulated the following research questions:

RQ-1: What is the size distribution of the dictionary used by users when forming search queries?

RQ-2: How does the use of KAN quantitatively improve the avatar model compared to MLP?

5. Experiment

To test our methodology, we collected a sequence of purchase search queries for random customers over the year. The result is a dataset D with a unique key "user" and a time series ["request1," ..., "purchase1," "end of session," "request N"]. Each purchase is presented as a text consisting of the name, brand and product characteristics. Typos have been corrected in search queries, synonyms have been added, and low-performance queries have been mapped to high-performance queries. This way, a time-ordered text consisting of queries and descriptions of purchased goods is received for each user in D. Based on this text, autoregressive training of the avatar model of each user is performed.

Using the obtained avatar model, a predictive model was built for each user: based on the entered search queries, the next search query from the search query index is predicted. Pessimization is performed on words from purchased goods, taking into account their remoteness in terms of user interaction time. Multiple search queries are predicted for the user.

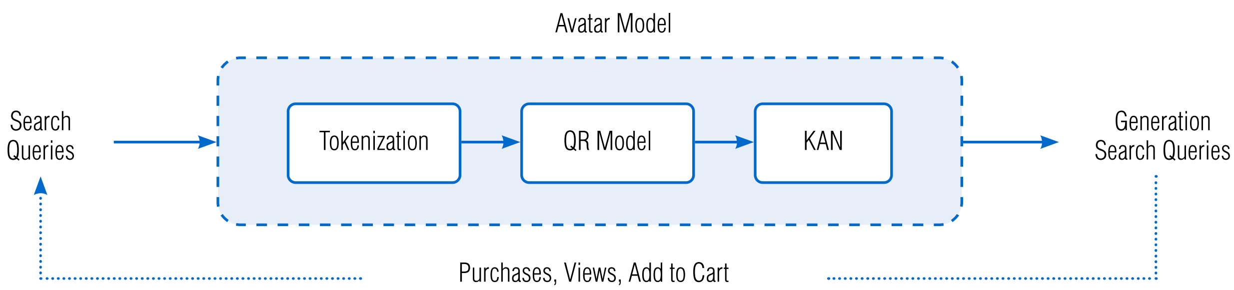


Fig. 2. Avatar model scheme.

Here is a scenario for using the results of this study: A user visits an ETP, does not enter a search query, sees search results for products that are close to him in vocabulary and do not contain recent purchases in the first place. The search output reflects the current status of the product catalog, inventory, region, popularity and other conditions that enhance relevance. The search results are highly diverse and new, since they contain products from several predicted queries for the user. The issue does not contain the purchased goods in the first positions.

For an experimental answer to the RQ-1 research question, a distribution of the number of unique words for each user was constructed from the dataset *D* (*Fig. 3*).

From the distribution in *Fig. 3*, we obtain that the maximum dictionary size does not exceed 1500 words. Compared to the size of the dictionary in the LLM shown in *Table 1*, the size of the dictionary of an individual user is an order of magnitude smaller. This fact provides a basis for a significant acceleration of the avatar model compared to the LLM.

To find an answer to the research question RQ-2, a series of experiments were conducted to train an avatar model on a data set *D*. The following options were considered as hyperparameters of the avatar model:

Tokenisation model:

Unigram, byte pair encoding (BPE), **Word**

Vocabulary size: 100, 1000, **"Unlimited"**

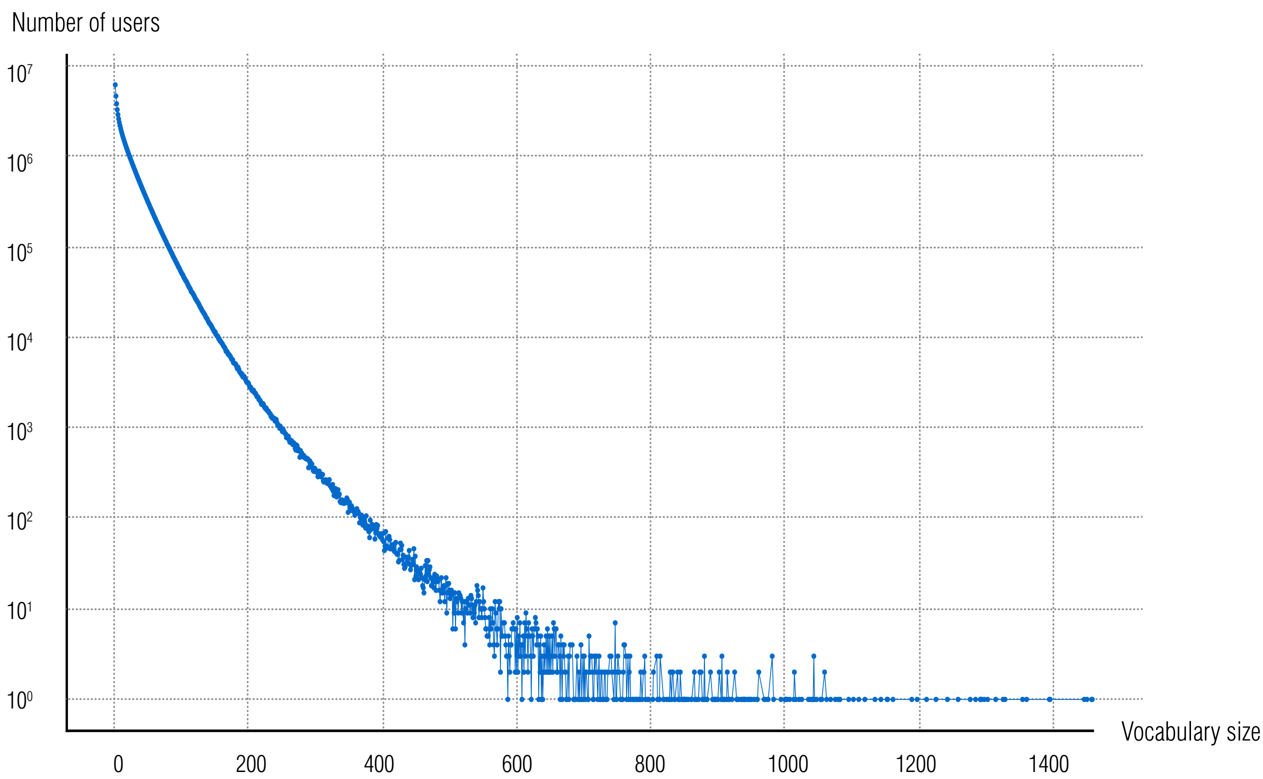


Fig. 3. Distribution of the number of unique words for each user.

Number of examples for subword tokenization: 3, 5, 7, 9, **without sampling**

QR Model:

The dimension of the vector space of the token representation: 32, **64**, 128, 256

Recurrent Neural Network (RNN) parameters:

Number of layers: **2**, 3, 4, 5, 6

Bidirectional: yes, **no**

Type: LSTM, **GRU**

In-batch negative sampling: 3, **5**, 7, 12

Dual Margin Cosine Embedding Loss: positive margin **0.9**, negative margin **0.2**

Shared vector embeddings: no, **yes**

KAN model:

Wavelet type: **"Mexican hat",** Shannon wavelet, Morlaix wavelet, Gauss wavelet.

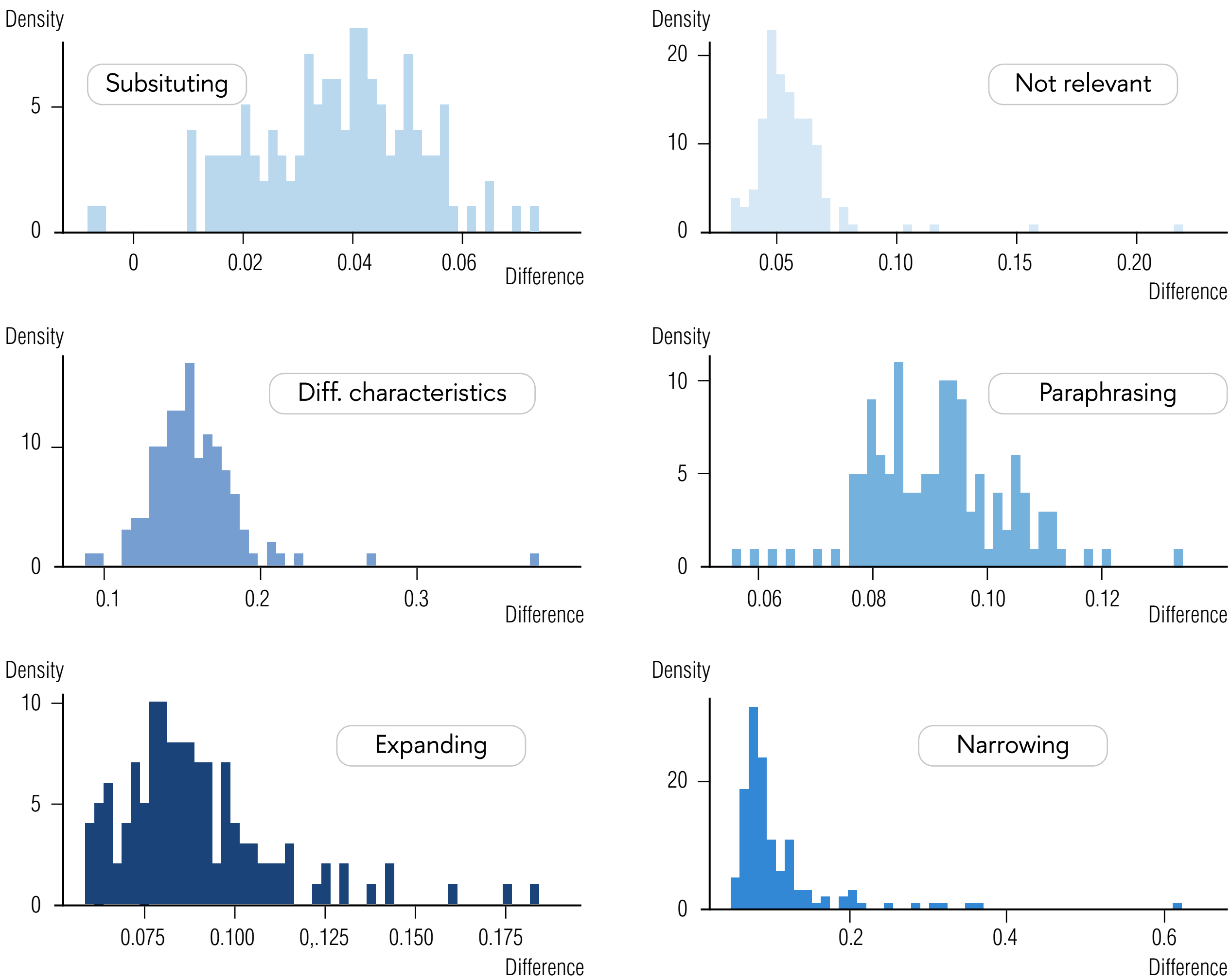


Fig. 4. Distribution of differences in the Accuracy metric between SKA and MPR.

A total of 120 experimental sessions were conducted, each of which lasted from 12 to 18 hours. The study identified the parameters for which the avatar model demonstrated the best results in terms of convergence rate and minimal learning error.

All other things being equal, the distribution shown in *Fig. 4* shows positive changes in accuracy when using KAN in the avatar model in comparison with MLP.

Conclusion

This study suggests a new approach to creating a recommendation system for users in the field of e-commrce, named "Ellochka" in honor of the heroine of the novel "12 Chairs" with a small vocabulary, who successfully coped with any communication tasks.

The authors have developed and tested a methodology based on the following principles:

1. Abandoning the use of a monolithic, unified recommendation system for all users of the electronic trading platform in favor of creating separate recommendation models for each user.

2. Building language models for recommending search query texts based on small-size token dictionaries instead of huge dictionaries with character tokenization.

3. Application of the mathematical apparatus of Kolmogorov–Arnold networks to improve the convergence rate of models during training.

The methodology proposed in the article has been successfully tested on the data of a working electronic trading platform and has allowed us to improve the autonomous indicators of the recommendation system of tips. ■

References

1. Butorin A.V., Murtazin D.G., Krasnov F.V. (2020) *Method and system for predicting effective thicknesses in the interwell space when constructing a geological model based on the method of clustering spectral curves*. Patent for invention RU 2718135 C1, 03/30/2020. Application No. 2019128334 dated 09/09/2019.

2. Krasnov F. (2023) Query understanding via Language Models based on transformers for e-commerce. *International Journal of Open Information Technologies*, vol. 11, no. 9, pp. 33–40 (in Russian).

3. Muennighoff N., Tazi N., Magne L., Reimers N. (2022) MTEB: Massive text embedding benchmark. *arXiv:2210.07316*. https://doi.org/10.48550/arXiv.2210.07316

4. Li P., Tuzhilin A. (2023) When variety seeking meets unexpectedness: Incorporating variety-seeking behaviors into design of unexpected recommender systems. *Information Systems Research*, vol. 35, no. 3. https://doi.org/10.1287/isre.2021.0053

5. Wang Y., Banerjee C., Chucri S., et al. (2024) Beyond item dissimilarities: Diversifying by intent in recommender systems. *arXiv:2405.12327*. https://doi.org/10.48550/arXiv.2405.12327

6. Castells P., Hurley N., Vargas S. (2022) Novelty and diversity in recommender systems. *Recommender Systems Handbook* (eds. F. Ricci, L. Rokach, B. Shapira). Springer, New York, NY, pp. 603–646. https://doi.org/10.1007/978-1-0716-2197-4_16

7. Ding Q., Liu Y., Miao C., et al. (2021) A hybrid bandit framework for diversified recommendation. Proceedings of the *AAAI conference on artificial intelligence*, vol. 35, no. 5, pp. 4036–4044. https://doi.org/10.1609/aaai.v35i5.16524

8. Li J., Wang M., Li J., et al. (2023) Text is all you need: Learning language representations for sequential recommendation. Proceedings of the *29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1258–1267. https:// doi.org/10.1145/3580305.3599519

9. Sun F., Liu J., Wu J., et al. (2019) BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. Proceedings of the *28th ACM international conference on information and knowledge management*, pp. 1441–1450. https://doi.org/10.1145/3357384.3357895

10. Klenitskiy A., Vasilev A. (2023) Turning dross into gold loss: is BERT4Rec really better than SASRec? Proceedings of the *17th ACM Conference on Recommender Systems*, pp. 1120–1125.

11. Ji Z., Yu T., Xu Y., et al. (2023) Towards mitigating LLM hallucination via self reflection. *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 1827–1843.

12. Egorov E.A., Rogachev A.I. (2023) Adaptive spectral normalization for generative models. *Doklady Mathematics*, vol. 108(suppl. 2), pp. S205–S214. https://doi.org/10.1134/S1064562423701089

13. Chang W.C., Jiang D., Yu H.F., et al. (2021) Extreme multi-label learning for semantic matching in product search. Proceedings of the *27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 2643–2651.

14. Sennrich R., Haddow B., Birch A. (2016) Neural machine translation of rare words with subword units. Proceedings of the *54th Annual Meeting of the Association for Computational Linguistics*, pp. 1715–1725. https://doi.org/10.18653/v1/P16-1162

15. Kudo T., Richardson J. (2018) SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. Proceedings of the *2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 66–71.

16. Liu Z., Wang Y., Vaidya S., et al. (2024) KAN: Kolmogorov-Arnold networks. *arXiv:2404.19756.* https://doi.org/10.48550/arXiv.2404.19756

17. Vaca-Rubio C. J., Blanco L., Pereira R., Caus M. (2024) Kolmogorov-Arnold networks (KANs) for time series analysis. *arXiv:2405.08790.* https://doi.org/10.48550/arXiv.2405.08790

18. Ma X., Gong Y., He P., et al. (2023) Query rewriting for retrieval-augmented Large Language Models. *arXiv:2305.14283.* https://doi.org/10.48550/arXiv.2305.14283

19. Chen Z., Fan X., Ling Y. (2020) Pre-training for query rewriting in a spoken language understanding system. *I2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020), Barcelona, Spain*, pp. 7969–7973. https://doi.org/10.1109/ICASSP40776.2020.9053531

20. Bhandari M., Wang M., Poliannikov O., Shimizu K. (2023) RecQR: Using Recommendation Systems for Query Reformulation to correct unseen errors in spoken dialog systems. *17th ACM Conference on Recommender Systems (RecSys'23), Singapore*, pp. 1019–1022.

21. Kim T.H., Neubauer M., Sfiligoi I., et al. (2004) The CDF central analysis farm. *IEEE Transactions on Nuclear Science*, vol. 51, no. 3, pp. 892–896. https://doi.org/10.1109/TNS.2004.829574

22. Kuleshov A.P. (2008) Cognitive technologies in adaptive models of complex objects. *Informatsionnye Tekhnologii i Vychslitel'nye Sistemy*, vol. 1, pp. 18–29.

23. Luo C., Lakshman V., Shrivastavaet A., et al. (2022) ROSE: Robust caches for Amazon product search. *WWW '22: Companion Proceedings of the Web Conference 2022*, pp. 89–93. https://doi.org/10.1145/3487553.3524213

About the authors

Fedor V. Krasnov

Candidate of Sciences (Technology);

Senior Data Scientist, Research Center of WB SK LLC based on the Skolkovo Innovation Center, 5, Nobel St., Mozhaisk District, Western Administrative District, Moscow, Russia;

E-mail: krasnov.fedor2@rwb.ru

ORCID: 0000-0002-9881-7371

Fedor I. Kurushin

Doctoral Student;

Data Scientist, Research Center of WB SK LLC based on the Skolkovo Innovation Center, 5, Nobel St., Mozhaisk District, Western Administrative District, Moscow, Russia;

E-mail: kurushin.fedor@rwb.ru

ORCID: 0009-0007-5126-4507

DOI: 10.17323/2587-814X.2025.2.54.76

Mathematical model and intelligent system for analyzing the intensity of megaproject changes: the role of temporary change management hubs

Pavel A. Mikhnenko
E-mail: pmihnenko@bmstu.ru

Bauman Moscow State Technical University, Moscow, Russia

Abstract

Megaprojects represent large-scale investment programs with complex organizational structures, uniting a multitude of stakeholders whose interactions lead to the redistribution of power and the creation of temporary management centers. In conditions of unstable and uncertain external environments, such stakeholder behavior can result in the failure to achieve the set goals of the megaproject. An important scientific task is the development of mathematical models and methods for managing changes in megaprojects caused by the integrative actions of stakeholders under complex external conditions. The present study is aimed at creating a mathematical model and developing an information system for neural network analysis of the intensity of changes in megaprojects. Megaproject management is described using a vector-matrix model of a dynamic system with feedback based on the results of changes. To identify recurring patterns of negative events, the event-oriented analysis method was used. This allows for justifying new approaches to management aimed at reducing uncertainty and enhancing the effectiveness of megaproject implementation. Based on the proposed tools, a retrospective neural network analysis of the intensity of changes in the "Nord Stream 2" megaproject was conducted. Within the study, key groups of stakeholders were identified whose interactions significantly impacted the project's implementation: Group 1 – Gazprom PJSC, European companies and the governments of Russia and Germany supporting the project; Group 2 – the governments of transit countries, the USA, environmental organizations and Baltic region countries opposing the project or expressing concern about its consequences. It was demonstrated that the integration of separate stakeholder groups contributes to the formation of temporary management centers with varying interests, leading to an increase in both positive and negative changes within the project. The outcome of the work was the development of an information system for analyzing the intensity of changes in megaprojects in the form of a prototype, which includes: a mathematical model for managing changes in megaprojects; a neural network analysis methodology based on the use of a large language model for processing textual information and generating quantitative assessments; as well as a software interface for uploading documents, automated data processing, and visualization of results. The primary neural network used was the large language model Qwen 2.5-Plus, which, while not specifically adapted for this task, had its parameters calibrated for analyzing the intensity of changes in megaprojects. The system prototype provides users with the ability to analyze stakeholder interactions, assess the intensity of changes and forecast potential risks based on historical data. A promising direction for further research involves applying the model we developed and neural network analysis methodology for comparative studies of various types of megaprojects.

Keywords: megaproject, "Nord Stream 2", stakeholders, integration activity, uncertainty, intensity of changes, temporary change management center, mathematical model, large language model, neural network analysis, information system

Citation: Mikhnenko P.A. (2025) Mathematical model and intelligent system for analyzing the intensity of megaproject changes: the role of temporary change management hubs. *Business Informatics*, vol. 19, no. 2, pp. 54–76. DOI: 10.17323/2587-814X.2025.2.54.76

Introduction

In this study, we define megaprojects as large-scale investments that require significant resource allocation and have long implementation times, with far-reaching implications for the economy, society, and environment [1]. The increasing prevalence of these projects is closely linked to the transformation of the global economic system [2]. Key stakeholders in megaprojects include government agencies, private companies, financial investors, and civil society organizations [3], whose interactions determine the outcomes of these projects. Due to their size and complexity, megaprojects not only affect the livelihoods of millions, but also generate varied and often conflicting expectations among different stakeholders [4, 5].

The key challenges of megaprojects include high costs, budget overruns and schedule delays [6], with the likelihood of cost and time overruns increasing alongside the scale of the project [7]. Frequently, mega-projects fail to achieve their expected outcomes due to uncertainty about economic benefits and complexity in stakeholder behavior [8].

In Russia, megaprojects are seen as tools for socioeconomic development and temporary management hubs that help coordinate the actions of various participants [9, 10]. They are also viewed as mechanisms for promoting technological sovereignty based on the concept of technological development until 2030. However, their informational support is still fragmented, making it difficult to compare projects on a unified methodological basis [10].

A distinguishing feature of megaprojects is the ambiguity of goals and the redistribution of stakeholder influence, which leads to the formation of temporary management hubs [11, 12]. Participants interpret project goals through symbolic frameworks that shape perceptions of the project's legitimacy [5]. Reducing uncertainties can be achieved by analyzing relationships between uncertainty types, incorporating stakeholder expectations and fostering dialogue with stakeholders [13].

From an organizational design perspective, megaprojects can be characterized as complex, multilayered networks [14–16]. As adaptive systems, they demonstrate properties of self-organization and emergence, wherein stakeholders dynamically adjust their behaviors through interconnected network interactions [3]. To address tasks within constrained timeframes, various coordination mechanisms establish temporary management hubs [17, 18]. However, a cohesive and universally accepted understanding of their fundamental nature remains elusive [17].

Megaprojects are characterized by flexible boundaries that facilitate the interchangeability of internal and external components, enabling the emergence of diverse configurations throughout their lifecycle [19, 20]. The concept of "organizational capabilities" is employed to describe their organizational design, encompassing a blend of knowledge, skills, resources, and processes for effective stakeholder integration [16, 21, 22]. Informal stakeholders often leverage coordination and adaptation mechanisms, which can lead to their incorporation into temporary management hubs, thereby amplifying their influence [23]. Ultimately, the success of megaprojects hinges on the nature and quality of interactions among stakeholders [24].

The object of this research is megaprojects – large-scale investment programs aimed at addressing socioeconomic development challenges which require significant resources and time. The subject is the change management processes of megaprojects, driven by the integration of stakeholders under conditions of external environmental uncertainty.

The goal of the study is to develop a mathematical model for managing changes in megaprojects and an intelligent information system for analyzing the intensity of these changes.

This research addresses a critical business problem associated with the fundamental challenges of megaproject management [6, 8]: budget overruns, delays, incomplete achievement of results and the complexity of stakeholder coordination. These issues arise from the high uncertainty of the external environment, intricate participant interactions and intersubjective factors. We characterize change management in megaprojects as a "wicked problem," defined by the absence of clear-cut solutions, dependence on context and participants and the necessity to consider multiple perspectives.

We identify several limitations in traditional methods of project change analysis. For instance, PERT/CPM fails to account for stakeholder interactions, Earned Value Management (EVM) struggles to adapt to uncertainty, and tools like IBM Rational Focal Point and SAP Portfolio and Project Management lack the capability to analyze temporary management hubs and process large volumes of data.

The main outcome of our research is a prototype of an information system. This system includes a mathematical model for change management in vector-matrix form, a methodology for neural network analysis based on the large language model Qwen 2.5-Plus (Alibaba Cloud) and a software interface for document uploads and automated data processing. The prototype enables the analysis of stakeholder interactions, assessment of the intensity of changes and forecasting of consequences based on historical data.

The research hypothesis is that the integration activity of stakeholders leads to the formation of temporary change management hubs (TCMHs), the dynamics of which determine the intensity and direction of changes in the megaproject, while the success of implementation depends on the ability of key stakeholders to establish sustainable cooperative relationships.

1. Research methods

This study employs a mixed-methods research design combining systematic literature analysis with empirical validation. Our methodological approach addresses three research gaps identified in megaproject management literature: (1) stakeholder coordination inefficiencies, (2) environmental uncertainty prediction, and (3) emergent governance structures (Temporary Collaborative Management Hubs – TCMHs).

The empirical investigation comprised three complementary phases:

First, we conducted primary data collection through semi-structured interviews with 12 senior project managers across the energy infrastructure sector, supplemented by participant observation at three international megaproject conferences (2021–2023). This qualitative approach enabled us to capture practitioner insights on stakeholder dynamics.

Second, we performed document analysis of 47 megaproject artifacts, including progress reports, meeting minutes, and stakeholder agreements. This made possible systematic examination of interaction patterns and TCMH formation processes in real-world contexts.

Third, we developed and validated a neural network analysis framework, benchmarking its performance against traditional content analysis methods using a corpus of 3 200 project documents. The comparative assessment focused on change detection sensitivity and pattern recognition accuracy.

The study progressed through two sequential research stages:

1. The developmental stage involved creating a novel mathematical model for change management, operationalized through a dedicated information system. Our vector-matrix formulation incorporates feedback loops to capture non-linear stakeholder influences.

2. The application stage featured retrospective analysis of the "Nord Stream 2" megaproject, selected as a critical case due to its complex stakeholder ecosystem and well-documented implementation challenges. We processed 1 850 project documents spanning 2018–2022 using our neural network architecture.

Key operational definitions guide our analysis:

- Megaproject stakeholders: Formal and informal actors exerting influence through direct participation or external pressure.
- Management actions: Deliberate interventions altering project trajectories.
- Integration Activity: Proactive coalition-building efforts measured through communication frequency and commitment levels.

Our findings demonstrate that effective integration requires either formal authority delegation or emergent leadership recognition. These integration processes catalyze TCMH formation, creating distinct governance nodes that significantly impact project outcomes.

1.1. Mathematical model for managing changes in a megaproject

Grounded in the methodological frameworks of control theory and organizational change theory, we propose a vector-matrix mathematical model to describe the dynamics of megaproject management. This model represents a dynamic system with negative feedback based on the results of changes (*Fig. 1*).

Changes in the external environment EXT make the vector G of project goals relevant, the comparison of which with the vector of results of the next stage R_C forms a vector of misalignment:

$$D = G - R_C + N.$$

The vector D is a vector of operational tasks for managing project changes. By the intensity of megaproject changes, we mean the extent of transformation of its actual goals and the structure of stakeholder interactions in response to changes in external environmental factors.

The intensity changes matrix $\mathbf{Q}$ $(s \times g)$ is the result of the matrix product:

$$\mathbf{Q} = \mathbf{AP},$$

where $\mathbf{A}$ is a symmetric square matrix $(s \times s)$ representing the degrees of pairwise stakeholder integration:

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1s} \\ \vdots & \ddots & \vdots \\ a_{s1} & \cdots & a_{ss} \end{bmatrix},$$

$\mathbf{P}$ is the matrix $(s \times g)$ representing the degrees of stakeholder impact on the achievement of the project's actual goals:

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1g} \\ \vdots & \ddots & \vdots \\ p_{s1} & \cdots & p_{sg} \end{bmatrix}.$$

The matrix $\mathbf{Q}$ indicates the extent to which each stakeholder, while influencing the achievement of project goals, involves other stakeholders in solving tasks.

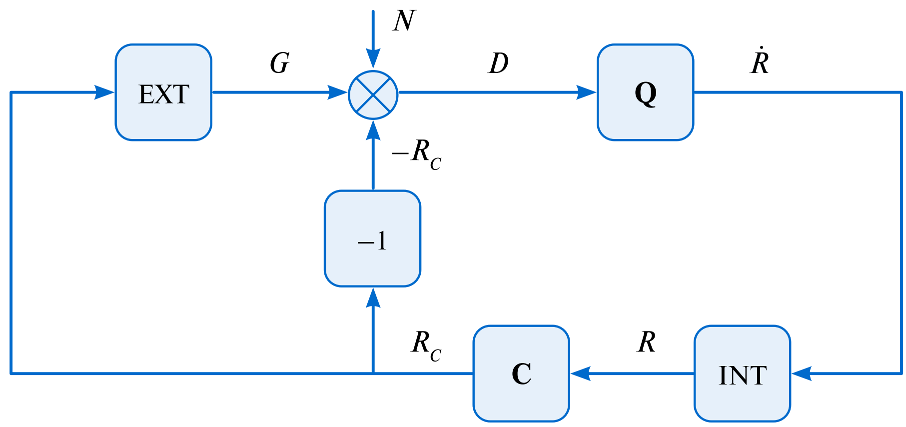


Fig. 1. Diagram of the mathematical model for managing changes in a megaproject[1].
Source: developed by the author.

[1] Here and hereafter, matrices are denoted by bold uppercase letters, with their dimensions indicated in parentheses. Vectors are denoted by italicized uppercase letters (a column vector is implied by the term "vector").

To assess the degree of stakeholder integration in the matrix **A**, we use a scale ranging from 0 to 1. Here, 0 indicates no integration, while 1 represents maximum interaction. The elements on the main diagonal reflect the degree of self-organization of each stakeholder. These values may be less than 1 if the stakeholder represents a group of participants that does not employ forms of maximum coordination.

The forms of stakeholder integration are ranked according to the degree of coordination (*Table 1*).

Analysis of the matrix **A** enables the identification of TCMHs to assess their role in achieving project goals at each stage. In the matrix **P**, the degree of a stakeholder's impact is evaluated on a scale from 0 to 1 (from complete lack of integration to maximum potential). Depending on the stakeholder's role, their potential can be directed either toward supporting or opposing the project goals.

To analyze the intensity of changes, a normalized matrix is formed:

$$\mathbf{Q}_N = \frac{100}{\sum_i \sum_j |q_{ij}|}\mathbf{Q},$$

where q_{ij} are the elements of matrix **Q**.

The normalization of matrix $\mathbf{Q}_N$ ensures that the sum of its elements equals 100%, with each element representing the percentage share of change intensity caused by the impact of a stakeholder, taking into account their interactions with other stakeholders. Multiplying the vector D on the left by matrix **Q** yields an s-dimensional vector $\dot{R}$, which describes the rate of project changes caused by the actions of stakeholders:

$$\dot{R} = \mathbf{Q}D.$$

The element INT (integral) describes the formation of an s-dimensional vector of cumulative change results for the project:

$$R(t) = \int_{t_0}^{t} \dot{R}(t)dt.$$

We introduce a control matrix **C** ($g \times s$), which allows us to compare the cumulative results of the project with its goals:

$$\mathbf{C}(t) = \mu \mathbf{Q}^{\mathrm{T}},$$

where μ is the coefficient of dynamics for eliminating misalignment, and "T" denotes transposition.

Table 1.

Forms of stakeholder integration

Degree of integration	Form of integration	Explanation
0.0–0.2	Independent management	Stakeholders work independently without coordinating decisions or interacting with each other.
0.2–0.4	Coordination (weak integration)	Stakeholders regularly exchange information and coordinate plans but remain responsible for their own tasks.
0.4–0.6	Collaboration (moderate integration)	Stakeholders collaborate to achieve project goals, hold regular meetings, and use shared tools and platforms.
0.6–0.8	Unification (high integration)	Stakeholders work closely together, form working groups and teams, make joint decisions, and develop strategies.
0.8–1.0	Consolidation (maximum integration)	All aspects of the project are fully integrated; stakeholders have a single management and control center, using unified standards and processes.

The result of left multiplication of vector R by matrix $\mathbf{C}$ is a g-dimensional vector $R_C = \mathbf{C}R$, which reflects the cumulative results aligned with the project goals. This vector impacts the project's external environment EXT and is used for comparison with the goal vector G. The matrix $\mathbf{C}$ ensures dimensional consistency among the vectors R_C, G, N and D according to the number of project goals.

An important aspect of modeling is the error vector N, caused by the uncertainty of the external environment and the inaccuracy of stakeholders' result evaluations. The model with negative feedback allows for analyzing the impact of change intensity on the alignment of results with project goals (vector D). In this study, we modeled the vector N as a set of uniformly distributed random variables centered around the variables of vector V, with variance corresponding to the expert assessment of informational uncertainty in the external environment.

The stochastic model enables the optimization of change intensity by minimizing the mathematical expectation and standard deviation of the misalignment vector D. The optimal intensity depends on the rate of changes in the external environment and the level of informational uncertainty. An increase in the pace of external environmental changes requires higher management intensity to enhance adaptability; however, a rise in uncertainty reduces the need for stakeholder integration [25]. Artificially intensifying stakeholder integration may divert the project's progress from the plan, increasing the variance of misalignments and reducing the predictability of outcomes. Nevertheless, insufficient change intensity under conditions of a dynamic external environment leads to a deviation of the project's trajectory's mathematical expectation from the planned path.

In this model, the dynamics of goal changes as a function of the dynamics of external environmental factors were modeled by the expression

$$g = \alpha_1 + \alpha_2 \sin(2\pi\alpha_3 t),$$

where g is an element of vector G;

α_1 is the offset coefficient along the ordinate axis, taking conditional values of 1, 2, 3, and 4 according to the number of goals for each stage;

α_2 is the amplitude of change;

α_3 is the frequency coefficient of changes, whose value equals the expert assessment of the degree of external environment dynamics under which the goal was set;

t is the time instant – the step for generating change results during one project stage.

The process of forming TCMHs is described as follows:

1. The process begins when the integration threshold is reached:

$$a_{ij} > \alpha,$$

where a_{ij} is an element of the integration matrix $\mathbf{A}$ – the degree of interaction between stakeholders i and j;

α is the threshold value (assumed to be 0.5).

2. A group of stakeholders is classified as a TCMHs if the following condition is met:

$$\sum_{j=1}^{s} a_{ij} > \beta \text{ for all } i \in \text{TCMHs},$$

where β is the minimum total degree of integration for TCMHs participants (assumed to be 2).

3. The impact strength of a TCMHs:

$$p_{\text{TCMHs}} = \sum\nolimits_{i,j \in \text{TCMHs}} a_{ij} \cdot p_{ij}\ .$$

The dynamics of TCMHs development are described by changes in key parameters over time:

1. The change in integration is described by the equation:

$$\frac{da_{ij}(t)}{dt} = f\left(a_{ij}(t), p_i(t), p_j(t), n(t)\right),$$

where f is a function that takes into account the current degree of integration, the impact potential of stakeholders, and the level of informational uncertainty.

2. Adaptation to external conditions:

$$g_i(t+1) = g_i(t) + \gamma \cdot c_{ij}(t) \cdot \left(g_i(t) - r_i(t)\right),$$

where γ is the adaptation rate coefficient.

3. The disintegration of a TCMHs occurs when $a_{ij}(t) < \delta$ or there is a significant increase in informational uncertainty $n(t) > \varepsilon$, where δ and ε are the threshold values for integration and uncertainty, respectively.

4. New TCMHs are formed if:

$$a_{ij}(t) > \alpha \text{ and } \sum_{j=1}^{s} a_{ij} > \beta.$$

Methods for analyzing the evolution of TCMHs:

1. Stakeholder clustering:

$$\text{Claster } k = \left\{ i \,|\, a_{ij} > \alpha_k \ \forall j \in \text{claster } k \right\}.$$

2. Identification of key parameters of the interaction network:

a) stakeholder centrality[2]:

$$c_i = \sum_{j=1}^{s} a_{ij},$$

b) clustering coefficient:

$$\phi_i = \frac{\sum_{j,k} a_{ij} \cdot a_{ik} \cdot a_{jk}}{\sum_{j,k} a_{ij} \cdot a_{ik}}.$$

3. Evaluation of the TCMHs life cycle:

a) average degree of integration:

$$\overline{a}_k(t) = \frac{1}{|k|} \sum_{i,j \in k} a_{ij}(t).$$

b) total impact:

$$p_k(t) = \sum_{i \in k} p_i(t).$$

Using the mathematical model for managing megaproject changes in the information system should address the following key questions:

1. What is the degree of stakeholder integration at each stage? What factors impact changes in the level of integration?
2. What TCMHs are formed during the project implementation? What is their role and nature of interaction?
3. What is the overall intensity of changes? How do positive and negative changes compare?
4. How do the speed of changes and informational uncertainty affect project management?
5. What is the level of misalignment between goals and results? What factors impact its dynamics?
6. How will changes in stakeholder activity impact future outcomes? What scenarios are possible?
7. What level of change intensity is optimal? How can the impact of uncertainty be minimized?
8. How does the model respond to estimation errors and changes in input data?

Thus, the mathematical model we developed differs from traditional models in the following key aspects:

1. Stakeholder consideration: The concept of "stakeholder integration activity" has been introduced, enabling the description of TCMHs formation mechanisms; the integration matrix **A** quantitatively evaluates the interaction among stakeholder groups [12, 18].
2. Detailed feedback mechanism: A negative feedback mechanism based on change results is used, providing an accurate description of goal adjustments based on actual outcomes [3, 4].
3. Intensity assessment: Intensity is determined

[2] "Stakeholder centrality" is a quantitative measure of their importance, impact, or role within the interaction network.

through the matrix product of **A** and **P**; accounting for the signs of the elements in matrix **Q** ensures a comprehensive analysis of changes [3, 21, 23].

4. Uncertainty consideration: The error vector *N* models informational uncertainty, allowing the system's behavior to be studied under varying levels of comple-xity [13].

5. New variable interpretation: The vector *D* is considered as an operational management task; the control matrix **C** aligns cumulative results with goals, which is crucial for long-term projects.

6. Megaproject specificity: Goal dynamics are modeled using a periodic function, reflecting the extended implementation timelines; integration accounts for the complex relationships among participants [19, 23].

7. Change optimization: The interrelationship between the rate of external environment change, informational uncertainty, and change intensity is considered, enabling project management adaptation to evolving conditions [25].

8. Practical applicability: The model was developed with real-world megaprojects in mind, such as the "Nord Stream 2" project.

These features ensure an accurate description of change management processes in megaprojects and the practical applicability of the model.

1.2. Intelligent information system and methodology for retrospective neural network analysis of megaproject change intensity

We conducted a retrospective neural network analysis to systematize information about the megaproject and obtain numerical estimates for calculating the intensity of changes and the role of TCMHs. By "retrospective," we refer to the analysis of a completed project, while "neural network-based" highlights our use of a large language model – an artificial transformer-type neural network built on machine learning technology.

We based our methodology on the use of the large language model Qwen 2.5-Plus (freely licensed) and implemented it through the Qwen Cloud API. We developed the software interface in Python 3.10. The main characteristics of the system are outlined in *appendix 1*, and the data processing algorithm along with the evaluation generation process is described in *appendix 2*.

The input data included documents containing information about stakeholders, goals, budgets, timelines and other aspects of the megaproject. We sourced this information from publicly available project charters, feasibility studies, scientific articles, official websites of organizations, statistical reports and industry reviews.

We formulated the user query to the model to obtain information about project stages, stakeholders, project goals, and numerical estimates of matrices and vectors. To enhance reliability, we utilized prompt engineering templates (*appendix 3*).

The adequacy of the generated content depends on the choice of a model capable of processing both textual and numerical data, the quality of the input data, the effectiveness of the query constructed using prompt engineering principles, and the analytical and statistical processing of the output data.

To improve the quality of the results, we applied a procedure of sequential statistical processing of numerical estimates. Within a single session, we processed batches of documents step by step, expanding the project's informational base. After loading each new document, we generated intermediate estimates while taking previous steps into account. This approach allowed us to bypass interface limitations related to the volume of uploaded files and improve the accuracy of estimates through sequential statistical processing, including the calculation of arithmetic means and standard deviations of the results. The number of steps was determined by the available volume of documents and the need to achieve convergence of mean values and stabilization of deviations.

2. Results of the retrospective intelligent analysis of changes in the "Nord Stream 2" megaproject

In this study, we analyzed 63 documents (2011–2024) to form the information base. These documents include media publications (RBK, TASS, Financial Times, etc.), scientific articles (eLIBRARY, Scopus, Web of Science) and data from official websites (PJSC "Gazprom", "Nord Stream 2", etc.). To refine numerical estimates, we processed the documents in eight steps by merging files.

Table 2 presents the list of key project stakeholders generated based on the results of the neural network analysis of textual data.

General goal of the "Nord Stream 2" project is to diversify gas supply routes, eliminate transit risks, meet the growing demand for energy resources in European countries and strengthen the continent's energy security. The project is exclusively commercial in nature[3].

Table 3 provides a list of the key goals of the project by stages, along with expert assessments of the dynamics of change (V) and informational uncertainty (N) in the external environment, which impact the setting of these goals.

In *appendix 4*, the stakeholder integration matrices by project stages are presented, constructed using the methodology of neural network analysis. The analysis

Table 2.

Key stakeholders of the "Nord Stream 2" project

Code	Stakeholder	Explanation
S1	PJSC Gazprom	The initiator and main beneficiary of the project, aiming to increase gas exports and reduce transit risks. The model considers the role and activities of Nord Stream 2 AG, whose sole shareholder is PJSC Gazprom.
S2	European energy ompanies	Investors and partners in the project interested in stable gas supplies and profits from participation: Uniper SE (Germany), Wintershall Dea GmbH (Germany), OMV AG (Austria), Engie SA (France), Royal Dutch Shell (Netherlands).
S3	Government of Russia	Supports the project to ensure revenues from gas exports and accelerate socio-economic development in Russia.
S4	Government of Germany	Initially supported the project as economically beneficial but later changed its position under pressure from the USA.
S5	Governments of transit countries	Opponents of the project, concerned about losing transit revenues and increased Russian impact.
S6	Government of the USA	An active opponent of the project, viewing it as a threat to Europe's energy security and a tool for Russian impact. It imposed sanctions on companies involved in the construction.
S7	Environmental organizations	Opposed the project, expressing concerns about its impact on the Baltic Sea environment.
S8	Governments of the Baltic region countries	Have mixed attitudes toward the project: some segments of the population support it due to potential economic benefits, while others are concerned about environmental risks and geopolitical consequences.
S9	Gas consumers in Europe	Interested in stable and affordable gas supplies but also concerned about the opinions of some politicians regarding possible dependence on Russia.

[3] Formulated based on: RIA "News" (official website), Lavrov stated the goal of the "Nord Stream 2" project, 28.08.2018 (updated 03.03.2020) [online resource]. URL: https://ria.ru/20180828/1527333450.html?ysclid=m3ic7exqtj972069035. Accessed on 15.11.2024.

Table 3.

Key objectives by project stages with assessments of dynamics of change (*V*) and information uncertainty (*N*) of the external environment[4]

Project stage	Key objectives	*V*	*N*	Comment
Stage 1. Concept and Planning (2011–2015)	G1.1. Planning to double gas supplies to Europe compared to the "Nord Stream" project.	0.22 (0.05)	0.31 (0.04)	Gas consumption growth in Europe was relatively predictable, though subject to fluctuations.
	G1.2. Developing the concept of reducing dependence on transit countries.	0.73 (0.06)	0.64 (0.06)	Political instability in transit countries created supply risks. High dynamics and uncertainty.
	G1.3. Developing the concept of European energy security.	0.50 (0.03)	0.44 (0.05)	The concept of European energy security was discussed, but its specific content and relation to "Nord Stream 2" were ambiguous.
	G1.4. Planning to attract European investments and partners.	0.43 (0.10)	0.53 (0.05)	European companies showed interest in the project, but the stance of individual countries and the EU as a whole remained unclear.
Stage 2. Preparation and Start of Construction (2015–2018)	G2.1. Obtaining necessary permits and approvals.	0.60 (0.09)	0.72 (0.04)	The process of obtaining permits and approvals across different jurisdictions was complex, lengthy, and highly uncertain.
	G2.2. Signing contracts with contractors and suppliers.	0.27 (0.11)	0.19 (0.12)	Signing contracts with contractors was a relatively standard procedure, albeit with certain risks.
	G2.3. Financing the project.	0.52 (0.08)	0.60 (0.07)	Attracting financing depended on political factors and sanction risks, creating uncertainty.
	G2.4. Starting the construction of the offshore pipeline section.	0.33 (0.10)	0.35 (0.07)	Technical challenges of constructing the offshore section were predictable and manageable.
Stage 3. Active Construction Phase and Increased Sanctions Pressure (2018–2021)	G3.1. Completing the pipeline despite U.S. sanctions.	0.90 (0.04)	0.81 (0.07)	U.S. sanctions pressure constantly increased, creating high dynamics and uncertainty for project completion.
	G3.2. Minimizing the impact of sanctions on project timelines and costs.	0.80 (0.11)	0.72 (0.10)	Finding ways to minimize the impact of sanctions was a challenging task with high uncertainty.
	G3.3. Certification and launch of the pipeline.	0.70 (0.05)	0.83 (0.09)	Certification and project launch faced political pressure and regulatory obstacles, creating high uncertainty.
	G3.4. Maintaining dialogue with European partners and regulators.	0.81 (0.08)	0.92 (0.07)	Political dialogue amid sanctions and changing geopolitical conditions was extremely difficult and unpredictable.
Stage 4. Project Suspension and Geopolitical Consequences (2022–present)	G4.1. Preserving the infrastructure of the "Nord Stream 2" project.	0.20 (0.09)	0.72 (0.06)	Preserving the infrastructure is technically possible, but the future of the project remains uncertain.
	G4.2. Assessing damages and exploring potential uses of the pipeline.	0.13 (0.06)	0.90 (0.05)	Potential uses of the pipeline under geopolitical instability are highly uncertain.
	G4.3. Minimizing financial losses.	0.52 (0.10)	0.81 (0.08)	Assessing and minimizing financial losses is complicated due to uncertainty about the project's future.
	G4.4. Analysis and lessons learned.	0.12 (0.07)	0.20 (0.05)	Analysis and lessons learned are internal processes, relatively independent of external factors.

[4] The mean value and the 95% confidence interval (Student's t-distribution) of the assessments from six experts are shown.

of their structure revealed two TCMHs (Temporary Change Management Hubs):

- TCMH-1: S1–S4 (Gazprom, European companies, Russia, Germany) – a stable integration during the first three stages and disintegration at the fourth stage;
- TCMH-2: S5–S8 (transit countries, the USA, environmental organizations, Baltic region countries) – an increase in integration across the stages of the project.

Examples of TCMH evolution:

TCMH-1 (S1–S4):

- Formation: High initial integration (a_{ij} > 0.8) and shared goals for project implementation.
- Development: Stable existence during the first three stages due to the support of key stakeholders.
- Dissolution: Significant decrease in integration at the fourth stage due to changes in the political climate (a_{ij} < 0.4).

TCMH-2 (S5–S8):

- Formation: Gradual increase in integration (a_{ij} > 0.4 by the third stage) in response to growing opposition to the project.
- Development: Strengthening of cooperation among project opponents (a_{ij} > 0.7 by the fourth stage).
- Current state: Maintaining a high level of integration even after the project's suspension.

The graphs of the integration dynamics of these TCMHs across the project stages form the "integration scissors" (*Fig. 2*).

Examples of the most illustrative dynamics of integration at various stages of the project include the following pairs of stakeholders (*Fig. 3*):

- S1/S4: Gazprom, Germany – a decline and a sharp drop at the final stage;
- S3/S4: Russia, Germany – a decline and a sharp drop at the final stage;
- S4/S6: Germany, USA – a decline during the first three stages and a sharp increase in "mutual understanding" at the final stage;
- S5/S6: Transit countries, USA – steady growth;
- S1/S9: Gazprom, Consumers – a noticeable decline.

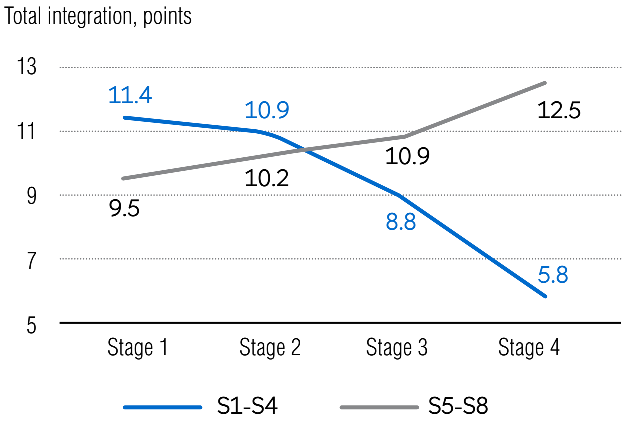


Fig. 2. "Integration scissors" of two groups of stakeholders, points.

In *appendix 5*, we present matrices explaining the stakeholders' impact on project objectives at different stages, derived using neural network analysis. Negative values denote adverse impacts that hinder the achievement of these objectives.

Figure 4 illustrates the relationship between the levels of integration, total impact, and negative influence of stakeholders across project stages (in points).

The decline in integration was accompanied by a decrease in the absolute value of stakeholders' influence. Negative influence became most pronounced during the second and third stages of the project.

Appendix 6 presents matrices of the intensity of project changes across stages. Negative values correspond to the intensity of adverse changes. *Figure 5* illustrates the total absolute values of overall change intensity (in points) and the percentage of negative change intensity across project stages.

The most intense changes occurred during the first two stages of the project, with a minimum of negative changes observed in the first stage. By the third stage, nearly 46% of all changes were directed toward opposing the project.

Figure 6 illustrates the dynamics of the generation of both total and negative changes as described by the aforementioned TCMHs.

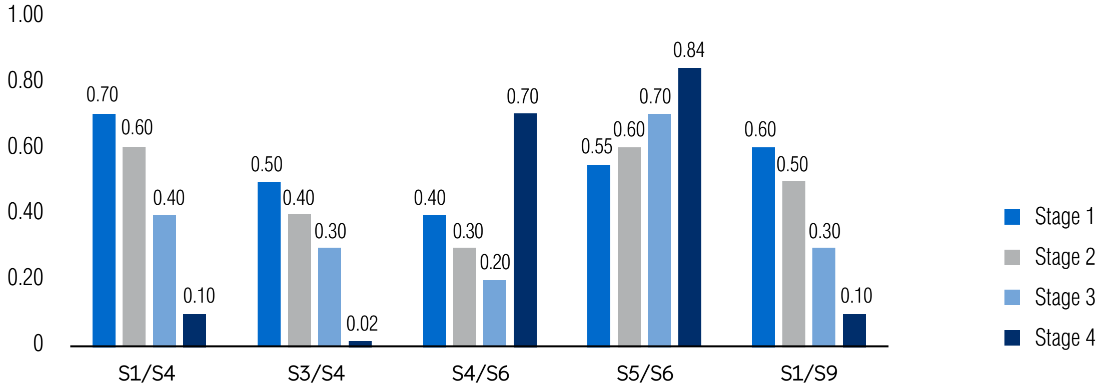


Fig. 3. Examples of integration dynamics for selected pairs of stakeholders across project stages.

At the first stage, TCMH-1 generated approximately 75% of positive changes, whereas at the third stage, TCMH-2 triggered over 46% of changes, nearly all of which were aimed at opposing the project.

Classification of TCMH types based on the analysis of the "Nord Stream 2" project data:

A. Stable TCMHs are characterized by a high degree of integration ($a_{ij} > 0.8$) and resilience to changes in the external environment (e.g., TCMH-1).

B. Dynamic TCMHs exhibit a moderate degree of integration ($0.4 < a_{ij} < 0.8$) and respond quickly to changes in external conditions (e.g., TCMH-2).

C. Transitory TCMHs emerge temporarily to address specific tasks and dissolve shortly after their completion ($a_{ij} < 0.4$) (e.g., temporary working groups for obtaining permits).

Figure 7 illustrates the graphs of the dynamics of cumulative results of R_C, constructed based on matrices **A** and **P**. The x-axis reflects the conditional duration of the stage, divided for clarity into 20 segments. The y-axes indicate the goal numbers. The direction of the shift of the R_C curve relative to the horizontal axis illustrates the degree of goal achievement and the deviation from planned values.

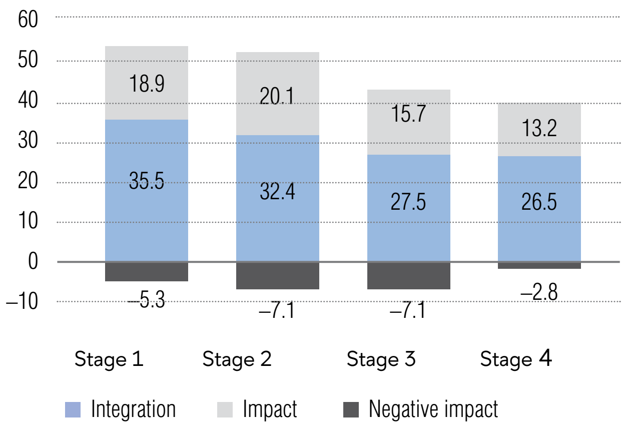


Fig. 4. Integration, total impact, and negative influence across project stages, in points.

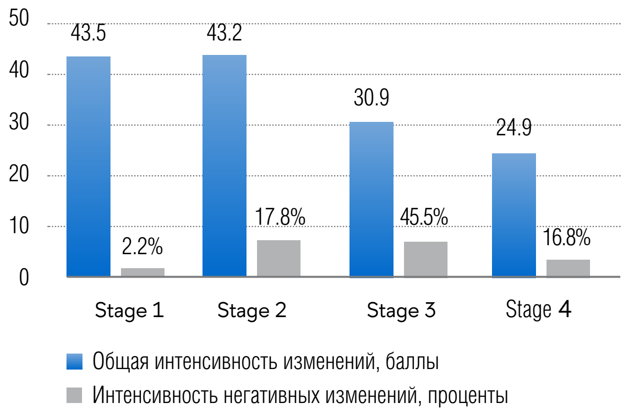


Fig. 5. Intensity of changes and percentage of negative changes.

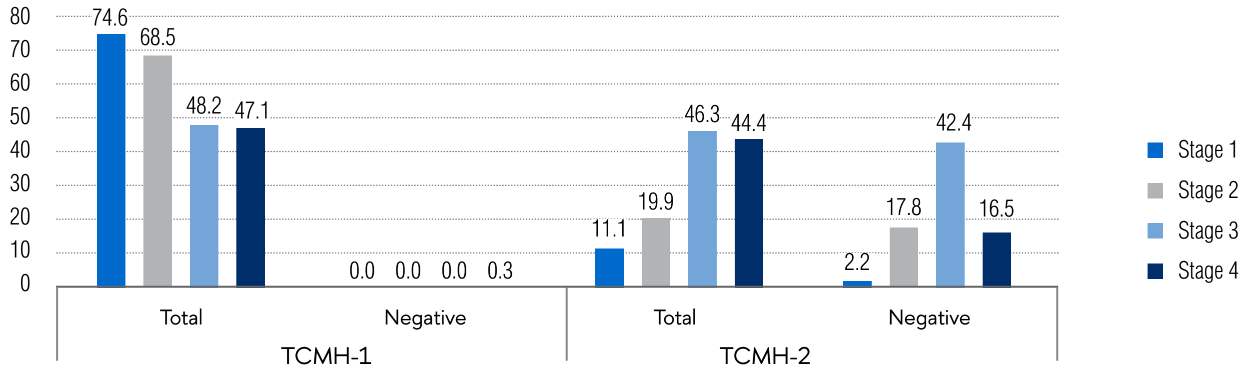


Fig. 6. Dynamics of change generation by two TCMHs, in percentages.

According to the simulation results, the goals of the first two stages (conceptual planning and construction preparation) were achieved thanks to TCMH-1, despite the negative influence of TCMH-2. Experimental modeling[5] demonstrated that an increase in the integration of TCMH-2 during the first stage would not have altered the outcomes. However, the maximum negative impact of TCMH-2 would have reduced the dynamics of goal achievement, particularly affecting the following objectives: reducing dependence on transit countries (G1.2), establishing European energy security (G1.3), concluding contracts (G2.2), securing financing (G2.3), and initiating construction (G2.4).

The implementation of the goals for the active construction phase encountered increasing sanction pressure: the construction pace (G3.1) and dialogue with regulators (G3.4) fell below the planned levels. The goals of minimizing the impact of sanctions (G3.2) and certifying the pipeline (G3.3) were not achieved due to the influence of TCMH-2. Experimental modeling demonstrated that maximum integration of TCMH-1 would have facilitated the achievement of goals G3.2 and G3.3, whereas an intensification of the negative impact from TCMH-2 would have resulted in the complete failure of the objectives for this stage.

Evaluating the geopolitical consequences requires time, with the analysis and lessons learned (G4.4) identified as the primary focus of the current stage. The simulation results indicate that achieving full integration of TCMH-1 is essential for successfully accomplishing all objectives at this stage. Furthermore, reducing the intensity of changes during the third and fourth stages diminishes the impact of informational "noise" but leads to greater deviations from the desired target outcomes [25].

3. Discussion and conclusion

The results of the retrospective analysis of changes in the "Nord Stream 2" megaproject are consistent with expert assessments and existing literature, confirming the suitability of the analytical framework we developed. These findings also support the hypothesis that the integrative activities of stakeholders lead to the formation of temporary change management hubs (TCMHs), which significantly influence the intensity of project changes. The dynamics of TCMH integration influence both the intensity and direction of changes. Additionally, the success of megaproject implementation hinges on stakeholders' ability to establish sustainable cooperative relationships.

[5] By experimental modeling, we refer to the "hypothetical" modification of certain values in matrices **A** and **P** to identify the nature of the impact that such scenarios would have on the cumulative results of R_C.

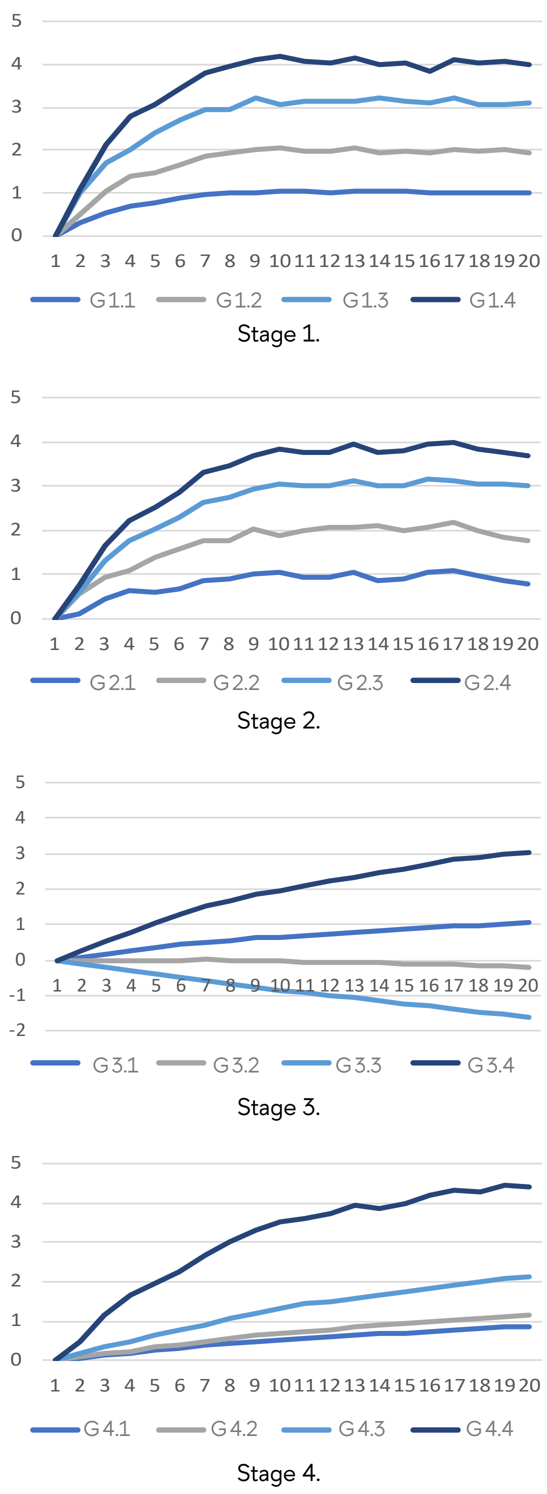


Fig. 7. Dynamics of the cumulative results of project content changes by stages.

The proposed mathematical model and neural network analysis methodology make it possible to assess change intensity and the impact of stakeholder integration activities on megaproject implementation. These insights could prove useful for project managers and participants, government authorities and researchers in the field of large-scale project management.

This research has identified the limitations of rational decision-making methods in managing megaproject changes:

1. The mathematical model demonstrates high sensitivity to input parameters, a characteristic typical of "wicked" problems.

2. The analysis of the "Nord Stream 2" megaproject revealed the significant influence of intersubjective factors on the dynamics of changes.

3. TCMHs were predominantly formed through informal mechanisms of coordination and adaptation.

The intelligent information system (IIS) we developed was compared with existing systems such as SAP Portfolio and Project Management (SAP), Total Organizational Risk Engine (TORE), Microsoft Power BI with Azure Machine Learning (MPBI) and IBM Watson Discovery (IBM). A comparative analysis was conducted by a panel of five experts using a scale ranging from 0 to 5.

The criteria for comparison included: document processing time (C1), data volume (C2), consideration of external environmental uncertainty (C3), analysis of temporary change management hubs (C4), evaluation of change intensity (C5), adaptability to new data (C6), interface usability (C7) and the need for customization (C8). *Figure 8* shows the distribution of expert scores across information systems and criteria.

According to expert evaluation, the information system we developed outperforms other systems in addressing the tasks under consideration.

The mathematical model, information system and neural network analysis methodology developed in this study can be utilized for both retrospective analysis of completed projects and management of ongoing projects. The key approaches include:

1. Forecasting and optimization of TCMHs dynamics:

Neural network data analysis predicts the formation and transformation of temporary change management hubs (TCMHs), identifying potential conflicts and suggesting preventive measures. Managing the intensity of changes through stakeholder integration levels and accounting for external factors helps avoid both excessive and insufficient activity.

2. Evaluation of management effectiveness and strategy adjustment:

Regular updates of integration and influence matrices enable the assessment of TCMH performance, adjustment of management strategies and adaptation of project goals based on current data and forecasts.

3. Scenario modeling and decision support:

Modifying parameters (integration, uncertainty, external environment) within scenario modeling assists in evaluating the consequences of decisions and selecting the optimal strategy. Integrating analytical data into decision support systems provides an objective foundation for effective management.

4. Risk monitoring and uncertainty management:

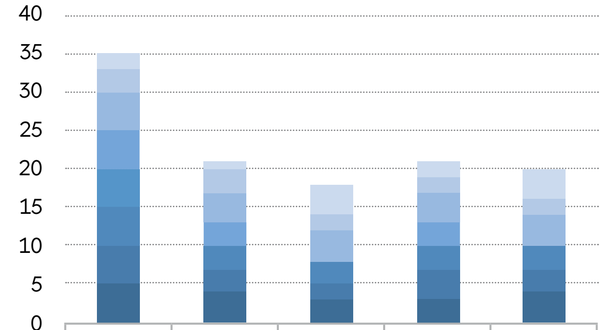
Analysis of key parameters (matrices **A**, **P**, **Q**, and vector *D*) allows for timely identification of risks and prevention of their escalation. Assessing informational uncertainty and external environmental impacts minimizes risks and enhances project adaptability.

5. Coalition building and stakeholder engagement:

Analyzing opportunities for enhancing integration among key participants facilitates the formation of effective coalitions and strengthens dialogue with stakeholders to achieve shared goals.

Thus, the proposed approaches provide comprehensive change management, enhancing the adaptability and effectiveness of projects.

A promising direction for further research is the application of the neural network analysis methodology to compare megaprojects of different types. This will help identify patterns in TCMH dynamics, taking into account industry-specific characteristics, scale, cultural context and other factors. Based on the data extracted by the neural network from a large document corpus, conclusions can be drawn regarding the success and failure factors of megaprojects. Additionally, universal recommendations for change management can be developed. ■



	IIS	SAP	TORE	MPBI	IBM
C8	2	1	4	2	4
C7	3	3	2	2	2
C6	5	4	4	4	4
C5	5	3	0	3	0
C4	5	0	0	0	0
C3	5	3	3	3	3
C2	5	3	2	4	3
C1	5	4	3	3	4

Fig. 8. Expert scores by criteria.

Acknowledgments

The work was performed within the framework of the state assignment of the Ministry of Science and Higher Education of the Russian Federation (topic No. 225013004474-0 “Development of methods and tools for multimodal business intelligence”).

The author thanks the anonymous reviewer for valuable comments and recommendations that contributed to improving the quality of the article, as well as the experts who participated in the comparative analysis of information systems.

References

1. Abdikeev N.M., Bogachev Yu.S., Bekulova S.R. (2019) Institutional mechanisms for ensuring a scientific and technological breakthrough in the Russian economy. *Management Sciences*, no. 9(1), pp. 6–19 (in Russian). https://doi.org/10.26794/2304-022X-2019-9-1-6-19

2. Mitrofanova I.V., Zhukov A.N., Batmanova V.V., Mitrofanova I.A. (2014) Megaprojects of territorial development: The experience of the United States of America and the Russian Federation. *National interests: priorities and security*, no. 31(268), pp. 28–40 (in Russian).

3. Daniel E., Daniel P.A. (2019) Megaprojects as complex adaptive systems: The Hinkley point C case. *International Journal of Project Management*, vol. 37, no. 8, pp. 1017–1033. https://doi.org/10.1016/j.ijproman.2019.05.001

4. Gil N.A. (2023) Cracking the megaproject puzzle: A stakeholder perspective? *International Journal of Project Management*, vol. 41, no. 3, article 102455. https://doi.org/10.1016/j.ijproman.2023.102455

5. Floricel S., Brunet M. (2023) Grandstanding? The elusive process of shaping megaproject symbolism. *International Journal of Project Management*, vol. 41, no. 5, article 102498. https://doi.org/10.1016/j.ijproman.2023.102498

6. Flyvbjerg B. (2014) What Pou should know about megaprojects and why: An overview. *Project Management Journal*, vol. 45, no. 2, pp. 6–19. https://doi.org/10.1002/pmj.21409

7. Baerenbold R. (2023) Reducing risks in megaprojects: The potential of reference class forecasting. *Project Leadership and Society*, vol. 4, article 100103. https://doi.org/10.1016/j.plas.2023.100103

8. Favari E., Cantoni F. (eds.) (2020) *Megaproject management: A multidisciplinary approach to embrace complexity and sustainability*. Springer Cham. https://doi.org/10.1007/978-3-030-39354-0

9. Gusev A.B., Yurevich M.A. (2022) The state management system of megaprojects as a model for restoring sovereignty. *Voprosy teoreticheskoj ekonomiki*, no. 3, pp. 62–76 (in Russian).

10. Balatsky E.V., Ekimova N.A. (2021) Phenomenon of megaprojects in multi-contour economy model. *Journal of Economic Regulation*, vol. 12, no. 4, pp. 25–39 (in Russian). https://doi.org/10.17835/2078-5429.2021.12.4.025-039

11. Bourne M., Bosch-Rekveldt M., Pesämaa O. (2023) Moving goals and governance in megaprojects. *International Journal of Project Management*, vol. 41, article 102486. https://doi.org/10.1016/j.ijproman.2023.102486

12. Jiang K., Le P., Zheng A., Zhang A., Ouyang L. (2023) Toward a systematic understanding of megaproject improvisation. *International Journal of Project Management*, vol. 41, article 102529. https://doi.org/10.1016/j.ijproman.2023.102529

13. Machiels T., Compernolle T., Coppens T. (2023) Stakeholder perceptions of uncertainty matter in megaprojects: The Flemish A102 infrastructure project. *International Journal of Project Management*, vol. 41, article 102437. https://doi.org/10.1016/j.ijproman.2023.102437

14. Martinsuo M., Ahola T. (2010) Supplier integration in complex delivery projects: Comparison between different buyer-supplier relationships. *International Journal of Project Management*, vol. 28, no. 2, pp. 107–116. https://doi.org/10.1016/j.ijproman.2009.09.004

15. Sun J., Zhang P. (2011) Owner organization design for mega industrial construction projects. *International Journal of Project Management*, vol. 29, no. 7, pp. 828–833. https://doi.org/10.1016/j.ijproman.2011.04.005

16. Zani C.M., Denicol J., Broyd T. (2024) Organization design in megaprojects: A systematic literature review and research agenda. *International Journal of Project Management*, vol. 42, no. 6, article 102634. https://doi.org/10.1016/j.ijproman.2024.102634

17. Burton R.M., Obel B. (2018) The science of organizational design: fit between structure and coordination. *Journal of Organization Design*, vol. 7, article 5. https://doi.org/10.1186/s41469-018-0029-2

18. Bakker R.M. (2010) Taking stock of temporary organizational forms: A systematic review and research agenda. *International Journal of Management Reviews*, vol. 12, no. 4, pp. 466–486. https://doi.org/10.1111/j.1468-2370.2010.00281.x

19. Denicol J., Davies A., Pryke S. (2021) The organizational architecture of megaprojects. *International Journal of Project Management*, vol. 39, no.4, pp. 339–350. https://doi.org/10.1016/j.ijproman.2021.02.002

20. Satheesh S.A., Verweij S., van Meerkerk I., Busscher T., Arts J. (2023) The impact of boundary spanning by public managers on collaboration and infrastructure project performance. *Public performance & management review*, vol. 46, no. 2, pp. 418–444. https://doi.org/10.1080/15309576.2022.2137212

21. Denicol J., Davies A. (2022) The Megaproject-based firm: Building programme management capability to deliver megaprojects. *International Journal of Project Management*, vol. 40, no. 5, pp. 505–516. https://doi.org/10.1016/j.ijproman.2022.06.002

22. Offenbeek M. A. G., Vos J. F. J. (2016) An integrative framework for managing project issues across stakeholder groups. *International Journal of Project Management*, vol. 34, no.1, pp. 44–57. https://doi.org/10.1016/j.ijproman.2015.09.006

23. Morkan B., Bertels H.M.J., Sheth A., Holahan P.J. (2023) Building megaproject resilience with stakeholders: The roles of citizenship behavior and critical transition mechanisms. I*nternational Journal of Project Management*, vol. 41, no. 5, article 102485. https://doi.org/10.1016/j.ijproman.2023.102485

24. Jayasuriya S., Zhang G., Pang R. J. (2024) Towards successful economic infrastructure partnership project delivery through effective stakeholder management. *Transportation Research Interdisciplinary Perspectives*, vol. 26, article 101173. https://doi.org/10.1016/j.trip.2024.101173

25. Mikhnenko P. (2009) Optimization of economic organization adaptation to environment changes. *Control sciences*, no. 4, pp. 32–38 (in Russian).

Appendix 1.

Key characteristics of the information system developed here

Category	Characteristics
System Architecture	Client-server architecture with cloud deployment. Web interface for interaction with end users. API gateway for integration with external systems.
API Configuration	Access methods: POST requests for sending prompts and receiving responses. Request limits: restrictions on queries per second (QPS) for system stability. Input and output data format: JSON.
Document Upload and Processing Module	Support for DOCX, PDF, TXT, CSV formats. Automatic conversion to UTF-8. Extraction of key metadata.
Data Analysis Module	Implementation of the mathematical model for change management. Application of a neural network model for generating numerical assessments. Sequential statistical processing of results.

Category	Characteristics
Results Visualization Module	Plotting the dynamics of stakeholder integration. Generating reports in HTML format.
Technical Specifications	Platform: Python 3.10. Libraries: Pandas, NumPy, Matplotlib, Seaborn, NLTK. API Qwen Cloud.
Document Upload	Support for DOCX, PDF, TXT, CSV formats. Automatic conversion to UTF-8. Extraction of key metadata.
Information Processing	Text preprocessing using the NLTK library. Forming requests to the Qwen Cloud API. Statistical processing of results using Pandas.
Functional Capabilities	Uploading and analyzing documents related to megaprojects. Generating integration matrices A, influence matrices P, and intensity matrices Q. Calculating the discrepancy vector D and control matrix C. Building predictive scenarios for event development. Evaluating the effectiveness of active TCMHs.
User Interface	Ability to select the project stage for analysis. Customization of analysis parameters (threshold values for integration, influence). Exporting results to Excel for further analysis.
System Workflow	Document upload by the user via the web interface. Text preprocessing and extraction of relevant information. Forming requests to the neural network model according to the mathematical model. Statistical processing of obtained results. Visualization and presentation of results.
Security	Data protection through AES-256 encryption. Two-factor authentication for system access. Role-based access control model.
Performance	Processing a single document up to 1 MB in size within 1-2 minutes. Analyzing a batch of 10 documents within 10-15 minutes. Maximum data volume processed in one session: 100 MB.
Scalability	Capability for horizontal scaling via Docker containers. Automatic scaling of computational resources based on load.
System Requirements	Minimum: Intel Core i5 processor, 8 GB RAM, 100 GB free disk space. Recommended: Intel Core i7 processor, 16 GB RAM, 200 GB free disk space, NVIDIA RTX 2060 GPU.
System Training	Use of a pre-trained neural network model. Capability for retraining on new data. Mechanisms for parameter calibration for specific projects.
Ergonomics	Intuitive interface. Query templates for typical tasks. Ability to save analysis settings.

Appendix 2.

Algorithm for processing textual data and generating numerical scores

Category	Components
Input Data Preparation	Conversion of documents into UTF-8 format. Extraction of metadata: creation date, authors, source. Standardization of terminology using an equivalence dictionary.
Prompt Design	Creation of a structured query template based on the variables of the mathematical model. Inclusion of checkpoints to verify the logical sequence of inferences. Use of a formalized language for describing project management processes (based on PMI PMBOK Guide).
Token Limit Constraints	Splitting documents into semantic blocks, each containing up to 32,768 tokens. Sequential loading of blocks with context preservation. Application of a sliding window to ensure information overlap between blocks.
Model Parameters	Temperature: 0.1 (increased determinism of responses). Top-p: 0.9 (ensuring diversity while maintaining quality). Max output tokens: 8,192 (comprehensive analysis of context). Repetition penalty: 1.1 (reduction of repetitiveness).
Statistical Processing of Results	Iterative refinement of numerical estimates through arithmetic mean and standard deviation calculations. A minimum of 8 document-loading iterations to ensure convergence of average values.

Appendix 3.

Prompt engineering templates

Prompt template	Comment	Code example
Chain-of-Thought Prompting	Breaking down complex tasks into a sequence of logically connected steps	1. "Step 1. Identify all stakeholder pairs with high integration." 2. "Step 2. Analyze the nature of their interaction." 3. "Step 3. Evaluate the impact on achieving project goals."
Few-Shot Learning	Providing several examples of correct answers before the main query	1. "Example 1. For the pair S1 and S4 at stage 1, the degree of integration is 0.70." 2. "Example 2. For the pair S5 and S6 at stage 3, the degree of integration is 0.60." 3. "Now perform a similar analysis for the pair S2 and S3."
Calibration Prompting	Including control questions with known answers to adjust probabilities	1. "What is the probability that S1 will have high integration with S4? (Correct answer: 0.70)"
Decomposition Prompting	Breaking down complex tasks into subtasks	1. "First, determine the integration between all stakeholder pairs." 2. "Then calculate the overall intensity of changes."

Output Parsing	Structuring responses in JSON format for ease of further processing	json
		1. {
		2. "integration_scores": {
		3. "S1_S4": 0.70,
		4. "S5_S6": 0.60
		5. },
		6. "change_intensity": 0.85
		7. }

Contextual Instructions	Including the context of result usage in the query	1. "The assessment is needed to build an integration matrix for further mathematical analysis."
Step-by-Step Feedback	Sequential refinement of model responses with feedback	1. "First version of the answer — ..." 2. "Now improve the answer, taking into account ..."

Appendix 4.

Stakeholder integration matrices by project stages

Stage 1.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1	1.00	0.85	0.90	0.70	0.10	0.20	0.20	0.30	0.60
S2	0.85	0.70	0.38	0.60	0.18	0.30	0.33	0.40	0.70
S3	0.90	0.38	1.00	0.50	0.10	0.11	0.10	0.20	0.39
S4	0.70	0.60	0.50	0.80	0.30	0.40	0.30	0.50	0.60
S5	0.10	0.18	0.10	0.30	0.60	0.55	0.40	0.30	0.22
S6	0.20	0.30	0.11	0.40	0.55	0.80	0.60	0.40	0.36
S7	0.20	0.33	0.10	0.30	0.40	0.60	0.70	0.51	0.30
S8	0.30	0.40	0.20	0.50	0.30	0.40	0.51	0.60	0.40
S9	0.60	0.70	0.39	0.60	0.22	0.36	0.30	0.40	0.70

Stage 2.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1	1.00	0.94	0.90	0.60	0.10	0.10	0.12	0.20	0.50
S2	0.94	0.80	0.33	0.50	0.10	0.24	0.20	0.30	0.60
S3	0.90	0.33	1.00	0.40	0.13	0.10	0.08	0.20	0.30
S4	0.60	0.50	0.40	0.75	0.20	0.30	0.20	0.40	0.50
S5	0.10	0.10	0.12	0.20	0.70	0.60	0.50	0.29	0.20
S6	0.10	0.24	0.10	0.30	0.60	0.90	0.70	0.40	0.18
S7	0.12	0.20	0.08	0.20	0.50	0.70	0.84	0.60	0.20
S8	0.20	0.30	0.20	0.40	0.29	0.40	0.60	0.70	0.32
S9	0.50	0.60	0.30	0.50	0.20	0.18	0.20	0.32	0.70

Stage 3.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1	0.98	0.73	0.90	0.40	0.10	0.05	0.05	0.10	0.30
S2	0.73	0.60	0.20	0.30	0.08	0.10	0.10	0.20	0.40
S3	0.90	0.20	1.00	0.30	0.10	0.03	0.00	0.13	0.22
S4	0.40	0.30	0.30	0.60	0.20	0.20	0.10	0.30	0.40
S5	0.10	0.08	0.10	0.20	0.80	0.70	0.60	0.29	0.20
S6	0.05	0.10	0.03	0.20	0.70	1.00	0.80	0.40	0.20
S7	0.05	0.10	0.00	0.10	0.60	0.80	0.90	0.50	0.18
S8	0.10	0.20	0.13	0.30	0.29	0.40	0.50	0.70	0.30
S9	0.30	0.40	0.22	0.40	0.20	0.20	0.18	0.30	0.60

Этап 4.

	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1	0.95	0.18	0.90	0.10	0.00	0.05	0.00	0.10	0.10
S2	0.18	0.30	0.10	0.10	0.00	0.10	0.10	0.10	0.20
S3	0.90	0.10	1.00	0.02	0.00	0.03	0.00	0.00	0.08
S4	0.10	0.10	0.02	0.70	0.30	0.70	0.40	0.40	0.44
S5	0.00	0.00	0.00	0.30	0.90	0.84	0.72	0.40	0.30
S6	0.05	0.10	0.04	0.70	0.84	1.00	0.70	0.52	0.40
S7	0.00	0.10	0.00	0.40	0.72	0.70	0.80	0.60	0.30
S8	0.10	0.10	0.00	0.40	0.40	0.52	0.60	0.80	0.40
S9	0.10	0.20	0.08	0.44	0.30	0.40	0.30	0.40	0.70

Appendix 5.

Matrices of the stakeholders' impact on project objectives by project stages

Stage 1.

	G1.1	G1.2	G1.3	G1.4
S1	1.00	0.80	0.88	0.80
S2	0.70	0.40	0.50	0.90
S3	0.90	0.72	0.80	0.50
S4	0.60	0.30	0.40	0.70
S5	-0.72	-0.90	-0.48	-0.40
S6	-0.40	-0.30	-0.40	-0.63
S7	-0.29	-0.22	-0.30	-0.30
S8	0.20	0.10	0.23	0.30
S9	0.50	0.28	0.40	0.60

Stage 2.

	G2.1	G2.2	G2.3	G2.4
S1	0.90	0.92	0.80	0.90
S2	0.68	0.80	0.92	0.70
S3	0.80	0.60	0.70	0.60
S4	0.48	0.40	0.27	0.40
S5	-0.80	-0.62	-0.50	-0.68
S6	-0.63	-0.70	-0.80	-0.60
S7	-0.50	-0.40	-0.30	-0.54
S8	0.20	0.28	0.20	0.18
S9	0.30	0.40	0.27	0.30

Stage 3.

	G1.1	G1.2	G1.3	G1.4
S1	0.98	0.59	0.40	0.50
S2	0.60	0.40	0.26	0.40
S3	0.82	0.53	0.30	0.20
S4	0.54	0.20	0.40	0.60
S5	-0.50	-0.72	-0.80	-0.54
S6	-0.58	-0.83	-0.90	-0.40
S7	-0.30	-0.37	-0.50	-0.32
S8	0.10	0.08	0.20	0.30
S9	0.18	0.20	0.30	0.40

Stage 4.

	G4.1	G4.2	G4.3	G4.4
S1	0.50	0.42	0.72	0.30
S2	0.20	0.30	0.58	0.40
S3	0.60	0.50	0.82	0.20
S4	0.10	0.20	0.30	0.60
S5	-0.20	-0.33	-0.40	0.11
S6	-0.30	-0.40	-0.54	0.70
S7	-0.10	-0.20	-0.30	0.42
S8	0.08	0.20	0.18	0.50
S9	0.20	0.30	0.40	0.63

Appendix 6.

Matrices of the intensity of project changes by project stages

Stage 1.

	G1.1	G1.2	G1.3	G1.4
S1	2.98	2.00	2.43	2.73
S2	2.13	1.33	1.71	2.10
S3	2.46	1.73	2.06	2.07
S4	1.99	1.16	1.60	1.96
S5	-0.10	-0.39	-0.09	0.02
S6	0.12	-0.25	0.05	0.20
S7	0.22	-0.11	0.14	0.30
S8	0.86	0.36	0.67	0.95
S9	1.84	1.08	1.46	1.86

Stage 2.

	G2.1	G2.2	G2.3	G2.4
S1	2.53	2.53	2.47	2.33
S2	1.80	1.92	1.77	1.72
S3	1.95	1.85	1.78	1.74
S4	1.34	1.39	1.24	1.23
S5	-0.72	-0.57	-0.56	-0.68
S6	-0.79	-0.63	-0.65	-0.76
S7	-0.68	-0.49	-0.50	-0.67
S8	0.19	0.33	0.27	0.13
S9	1.24	1.36	1.25	1.17

Stage 3.

	G1.1	G1.2	G1.3	G1.4
S1	2.32	1.36	0.97	1.26
S2	1.37	0.76	0.58	0.93
S3	1.97	1.16	0.82	0.97
S4	1.00	0.39	0.36	0.77
S5	-0.59	-1.13	-1.28	-0.52
S6	-0.85	-1.43	-1.59	-0.64
S7	-0.79	-1.26	-1.41	-0.59
S8	0.08	-0.34	-0.36	0.19
S9	0.80	0.30	0.26	0.68

Stage 4.

	G4.1	G4.2	G4.3	G4.4
S1	1.07	0.95	1.59	0.75
S2	0.23	0.26	0.43	0.54
S3	1.08	0.92	1.55	0.59
S4	-0.04	-0.03	-0.01	1.66
S5	-0.38	-0.55	-0.75	1.56
S6	-0.28	-0.38	-0.50	2.08
S7	-0.27	-0.36	-0.50	1.67
S8	-0.04	-0.03	-0.07	1.62
S9	0.14	0.20	0.31	1.47

About the author

Pavel A. Mikhnenko

Doctor of Sciences (Economics), Associate Professor;

Professor, Department of Business Informatics, Bauman Moscow State Technical University, 5, building 1, 2nd Baumanskaya St., Moscow 105005, Russia;

E-mail: pmihnenko@bmstu.ru

ORCHID: 0000-0002-1766-8029

DOI: 10.17323/2587-814X.2025.2.77.88

Parallel implementation of the simplex method in matrix form using the PyTorch library for economics and management problems

Yuriy S. Ezrokh [a]
E-mail: ezrokh@corp.nstu.ru

Alexey V. Snytnikov [a,b]
E-mail: aleksej.snytnikov@klgtu.ru

Elena Yu. Skorobogatykh [b]
E-mail: elena.skorobogatykh@klgtu.ru

[a] Novosibirsk State Technical University, Novosibirsk, Russia
[b] Kaliningrad State Technical University, Kaliningrad, Russia

Abstract

The simplex method is widely used in economic planning and forecasting tasks. However, this method is used in real economic activity to find solutions to large-scale tasks, the speed of which is not a critical factor. This significantly limits the applied value of the simplex method in the economic sphere, since currently there is a certain tendency to move to more detailed economic models, which makes it urgent to accelerate calculations based on the simplex method. In these conditions, GPU (Graphical Processor

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

Unit) computing accelerators become the most important means of accelerating calculations. The authors propose the implementation of the simplex method in matrix form for computing on GPUs using the PyTorch library. This allows you to switch to using the computing power of graphics processors in a simple and reliable way. A linear programming problem with 900 constraints is solved on a graphics accelerator 6–9 times faster than the solution on a conventional processor. This paper identifies groups of applied economic problems for which the proposed algorithms and methods can be relevant.

Keywords: modified simplex method, acceleration of calculations, graphics processors, linear programming problems in economics

Citation: Ezrokh Y.S., Snytnikov A.V., Skorobogatykh E.Yu. (2025) Parallel implementation of the simplex method in matrix form using the PyTorch library for economics and management problems. *Business Informatics*, vol. 19, no. 2, pp. 77–88. DOI: 10.17323/2587-814X.2025.2.77.88

Introduction

The simplex method [1] is widely used in economic planning and forecasting tasks [2, 3]. In a broad sense, the task of linear programming [4] is that it is necessary to maximize or minimize some linear functional on a multidimensional space under given linear constraints [5–7].

The role of linear programming, in particular the simplex method in economic analysis and planning, is described in [8]. It says that linear programming was considered as a tool for implementing neoclassical economic principles at a time when the very concept of a market economy was under attack from several directions. Linear programming has become widely used in national economic planning [9–11], especially for developing countries, and for studying individual industries, especially energy. The article [12] describes the experience of applying linear programming to the US oil refining industry. The use of linear programming combined with a more careful selection of information sources provides greater predictive power compared to the use of forecasts based on time series. The article [13] also emphasizes the need for a complete understanding of the specifics of production processes, otherwise the methods of mathematical economics [14], in particular, linear programming, give only very approximate results. Works [15, 16] characterize the use of the simplex method as follows: this procedure can be interpreted as a search for market prices that balance the demand for factors of production with their supply.

At the same time, there are practically no works that would test the use of the simplex method to solve a large number of similar tasks with multiple conditions in a short time for individual users. In other words, the simplex method is used in real economic activity to find solutions to large-scale tasks, the speed of which is not a critical factor. This significantly limits the applied value of the simplex method in the economic sphere.

Thus, based on the fact that the simplex method is actively used and there is a definite tendency to switch to more detailed models, it can be concluded that there is a need to speed up calculations based on it. First of all, we are talking about calculations performed on household/office computers rather than on supercomputers [17–19]. In this case, the Graphical Processor Unit (GPU) becomes the most important means of speeding up calculations [20, 21].

GPUs are a set of a very large number (up to 10 000) of simplified processor cores over shared memory which can run several tens of thousands of parallel processes. This means that graphics accelerators are very well suited for solving linear algebra problems (such as matrix multiplication). The computing time on GPUs is usually considered in comparison with the computing time on the Central Processor Unit (CPU).

1. Overview of parallel implementations of the simplex method

The article [22] presents the parallelization of the simplex method in a tabular version for systems with shared memory in order to solve large-scale linear programming problems with a dense matrix. The authors describe the general scheme of the method and explain the strategies adopted to parallelize each step of the standard simplex algorithm. The acceleration and parallel efficiency are analyzed in comparison with the standard simplex method when using a system with shared memory and 64 computing cores. Experiments were performed for several different tasks, up to 8 192 variables and constraints. The maximum acceleration achieved is about 19.

Work [23] considers parallelization of the simplex method based on OpenMP technology. This approach is used to improve acceleration and efficiency. The results of the parallel algorithm are compared with the usual simplex method. This work also used a feature of modern processors such as multithreading. The proposed parallelization algorithm scales easily to a different number of processor cores. In the general case, a number of computational experiments have been conducted using the example of solving the problem of distributing wagons between freight terminals of a railway station. An acceleration of 7 times is reported for 8 OpenMP streams with a data dimension of more than 102.

The article [24] discusses previous attempts to parallelize the simplex method in the context of improving the performance of sequential implementations of the simplex method and the nature of practical linear programming (LP) problems. For the main task of solving common large sparse LP problems, the author of the article did not find parallelization of the simplex method, which provides a significant improvement in performance compared to a good sequential implementation. However, some success has been achieved in the development of parallel solvers for dense LP problems or those or with special structural properties. As a result of the review, this paper determines the directions of future work aimed at developing parallel implementations of the simplex method with practical value. These directions are related to the use of parallel factorization methods and parallel sparse matrix inversion methods.

It was noted in [25] that studies devoted to the implementation of mathematical programming methods on the GPU are still in their infancy. One option is to modify existing algorithms in such a way as to achieve significant performance gains by executing on the GPU. The article [25] solves exactly this problem by presenting an effective implementation for the simplex method adapted for the GPU. The article describes how to perform the steps of the adapted simplex method in order to take full advantage of the GPU's capabilities. The experiments performed demonstrate significant acceleration compared to the sequential implementation, which highlights the enormous potential of the GPU.

The above review shows, firstly, the relevance of the work on the parallel implementation of the simplex method, and secondly, that this work should be mainly focused on computing systems with shared memory, in particular, on the GPU, which confirms the correctness of the problem statement for this article.

2. Description of the simplex method in matrix form

Due to the need to conduct a detailed analysis of the algorithm's performance, as well as to emphasize the difference from the more common simplex method in the form of tables, we describe the simplex method in matrix form (or revised simplex method), following the book [26].

Let's consider the linear programming problem in the following form: it is required to minimize $\boldsymbol{cx}$ under the condition $Ax = \boldsymbol{b}$, $x \geq 0$, where A is a matrix of size $m \times n$, having rank m. The algorithm for solving the problem is shown in *Fig. 1*.

Next, we provide a step-by-step description of the algorithm.

2.1. Initialization step

Choose an initial basic feasible solution with basis $\boldsymbol{B}$.

2.2. The main part of the algorithm

1. Solve a system of linear algebraic equations $\boldsymbol{B}x_B = \boldsymbol{b}$.

2. Solve a system of linear algebraic equations $\boldsymbol{wB} = \boldsymbol{c}_b$ with a single solution $\boldsymbol{w} = \boldsymbol{c}_b\boldsymbol{B}^{-1}$. The vector **w** is commonly called a simplex multiplier due to the fact that its components are used as multipliers for the rows of the matrix A when converting it to a canonical form. Next, you need to calculate $z_j - c_j = \boldsymbol{wa}_j - c_j$ for all nonbasic variables. Let $z_j - c_j = \max_{j \in J}\{z_j - c_j\}$, where J is the current set of indexes associated with nonbasic variables. If $z_k - c_k \leq 0$, then the current solution is the optimal solution. Otherwise, step 3 is performed with x_k.

3. Solve the system $\boldsymbol{By_k} = \boldsymbol{a}_k$. If $y_k \leq 0$, then the calculation stops, and it is concluded that the optimal solution is unlimited and lies on a straight line

$$\left\{\begin{bmatrix} \bar{b} \\ 0 \end{bmatrix} + x_k \begin{bmatrix} -y_k \\ \boldsymbol{e}_k \end{bmatrix} : x_k \geq 0\right\},$$

where $\boldsymbol{e}_k$ is a vector of length $(n - m)$, consisting of zeros, except for the component with the number k, which is 1.

If $y_k \geq 0$, then step 4 is performed.

4. Let x_k be included in the basis. Then $\boldsymbol{r}$ is the index of the blocking variable x_{B_r}, such that the basis remains unchanged as a result of the following check for the minimum ratio

$$\frac{\bar{b}_r}{y_{r,k} = \min_{1 \leq i \leq m}\left\{\dfrac{\bar{b}_r}{y_{r,k} : y_{i,k} > 0}\right\}}.$$

Next, you need to update the basis $\boldsymbol{B}$, where a_k replaces a_{B_r}, update the set of indexes J and repeat step 1.

2.3. The performance of the simplex method implementation in matrix form

The operating time of the simplex method in matrix form was measured on several model tasks on a computer with an Intel Core i7 950 processor, 3.07 GHz (*Table 1*). The model tasks were generated using an online linear programming task generator [27].

3. Software implementation of the simplex method in matrix form

Modern Python programming language tools, namely the Numpy library, facilitate extremely convenient implementation of computational algorithms based on operations with matrices. As a result, the simplex method was implemented in accordance with the procedure described in section 2.

One of the objectives of this work is to show that it is possible to use the power of high-performance computing to solve applied problems using very easy-to-use tools. In this regard, the PyTorch library, designed to work with neural networks, was used to implement the simplex method on the GPU. The PyTorch library has two important advantages from the point of view of

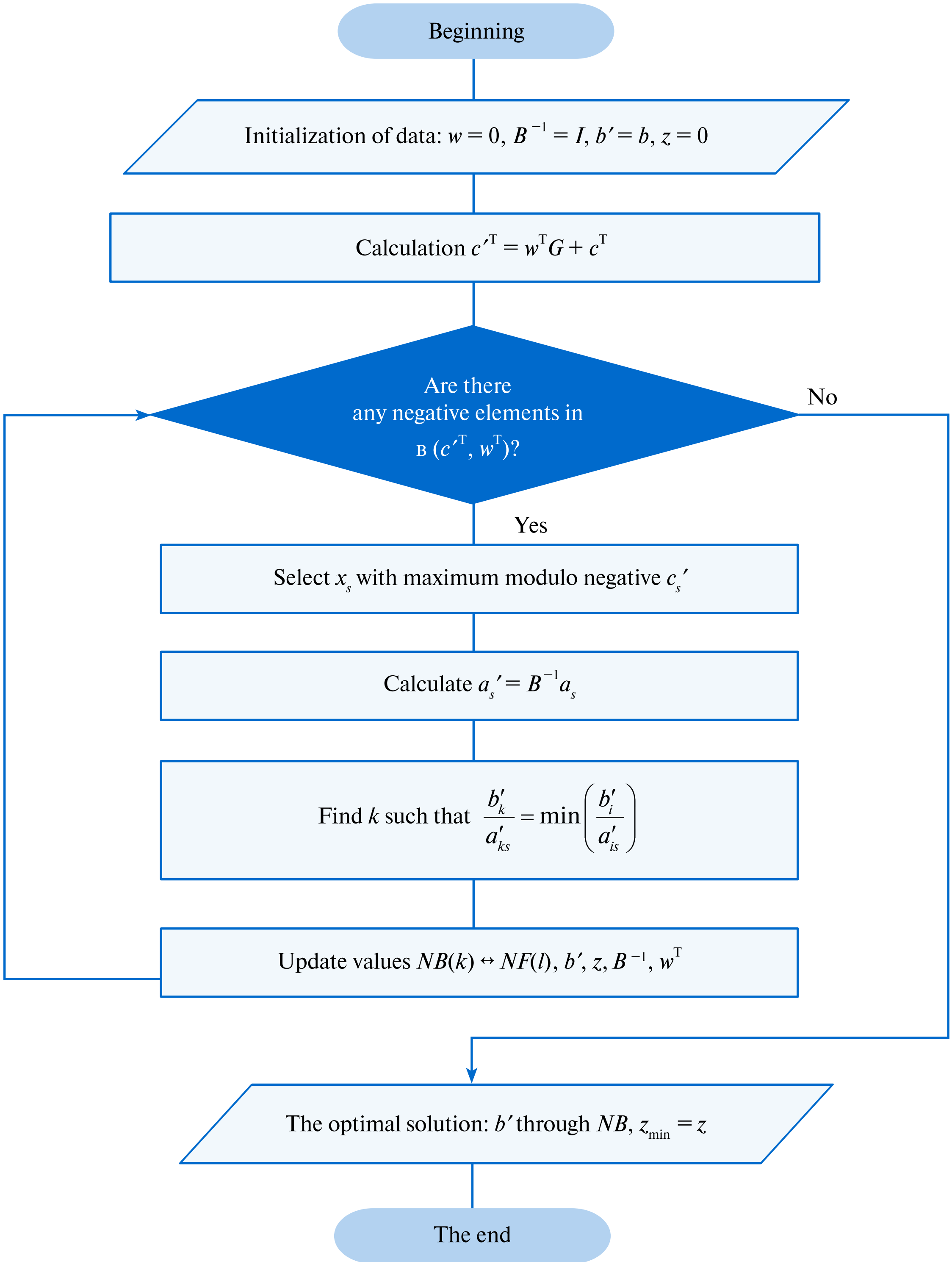


Fig. 1. Algorithm of the revised simplex method.

Table 1.

The operating time of the simplex method in matrix form

Number of restrictions	Matrix size *A*	Time, milliseconds
10	10 × 20	9.3
10	10 × 910	27.1
100	100 × 600	36.7
900	900 × 1800	222.3

the problem being solved: a large set of tools for matrix calculations and the ability to transfer calculations to the GPU without changing the program, just adding one instruction that makes sense to "transfer the matrix to the GPU." All further calculations with the specified matrix will be performed on the GPU. As can be seen from *Table 2*, when transferring the calculation to the GPU using the PyTorch library, the calculation time is noticeably reduced.

Without giving code listings, however, we can say that when using alternative implementations of the simplex method, the main of which is CUDA technology, it is necessary to perform several preparatory operations before calculating, for example, a matrix product, namely, to move the matrices from computer memory to GPU memory and convert them to a format that is optimal in terms of computing on the GPU. Then, in fact, the implementation of the matrix product is performed based on asynchronous calculations using the maximum possible number of parallel processes.

Thus, it can be seen that porting computational algorithms to the GPU using the PyTorch library is much easier than using specialized programming tools such as CUDA, OpenACC, cuBLAS, etc.

The performance of the simplex method implemented on the GPU was measured on the Nvidia TITAN X graphics accelerator, as well as on Nvidia Volta (computing cluster of the NSC NGU). The measurement results are presented in *Table 2.*

Table 2.

Performance of the simplex method implemented on the GPU

Number of restrictions	Matrix size *A*	Time on Titan, ms	Time on Volta, ms	Time on CPU, ms
10	10 × 20	1.9	1.3	9.3
10	10 × 910	4.4	4.4	27.1
100	100 × 600	5.2	5.2	36.7
900	900 × 1800	81.9	23.7	222.3

Shorter billing time for Nvidia Volta is achieved due to the so-called tensor cores, which additionally speed up operations such as matrix multiplication.

The program's performance was also analyzed on the Nvidia A100 Ampere graphics accelerator installed on the KSTU's Euler computing complex. *Figure 2* shows the dependence of the operating time of the matrix conversion procedure, as the mathematically most time-consuming operation, on the parameter N, which, from the point of view of the simplex method, represents the number of variables. Here it is important to pay attention to the fact that matrix inversion is implemented in CUDA using the PyTorch library, and the graph shown does not mean that a linear dependence of the operating time of matrix inversion on the number of variables has been obtained. To clarify this point, consider the number of CUDA threads. Due to the fact that the matrix reversal algorithm is not implemented manually on CUDA, the number of threads is selected automatically. However, the profiler makes it possible to see how many threads are used. The result is shown in *Fig. 3*.

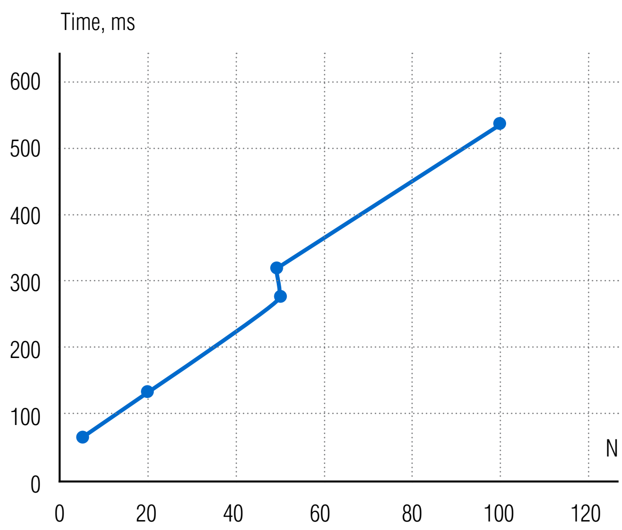


Fig. 2. Dependence of the matrix reversal time on the graphics accelerator Nvidia A100 from the number of simplex method variables.

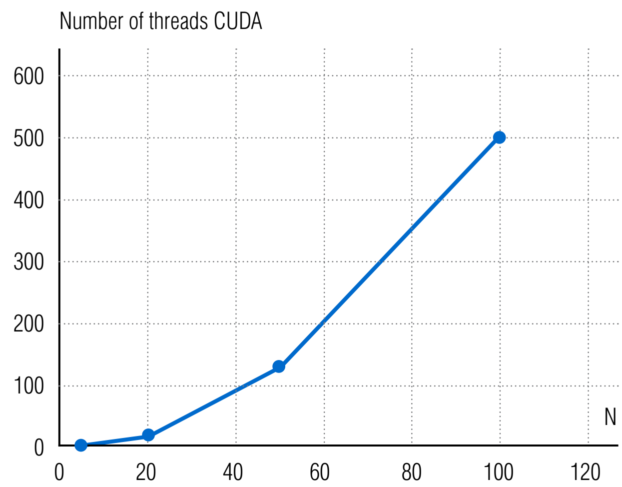


Fig. 3. The number of CUDA threads used when performing matrix inversion in the implementation of the simplex method on the Nvidia A100 graphics accelerator.

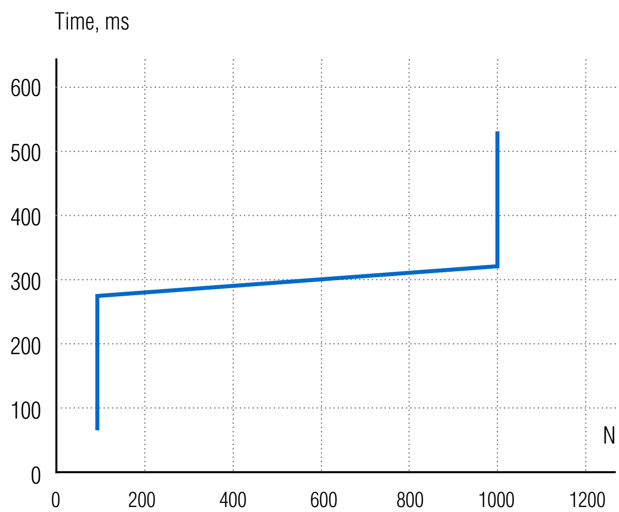


Fig. 4. Dependence of the matrix reversal time on the graphics accelerator Nvidia A100, depending on the number of limitations of the simplex method.

In addition, the dependence of the matrix reversal time on the number of constraints of the simplex method M looks interesting (*Fig. 4*).

4. Directions of applied economic application of the software implementation of the simplex method in matrix form

The algorithms and approaches described above can be used to solve three main groups of applied economic problems, including in the field of marketing, management (in the framework of customer experience management, as well as the implementation of assortment and product policies), plus finance.

The first group is commercial tasks in the field of marketing promotion of goods/services that require an instant solution in the form of an online response in order to obtain a positive super-effect of spontaneous purchases. An example of this is the application of grocery hypermarkets, which generates the final consumer basket of the current order in the presale mode. Given the increasing level of detail of information about the customer (his personal data, as well as the purchase history), the number of parameters acting as limitations within the simplex method is becoming large enough to justify its use. In such a situation, the client is not ready to wait for the time required to calculate the CPU (taking into account the queue of clients), as well as to transfer information. Given the speed of modern life, the client is focused on obtaining exceptionally instant results achieved with the use of graphics accelerators, which is accompanied by an appropriate explanation and justification.

The second group of economic tasks is the intellectualization of the work of high-performance chatbots using fuzzy logic, that is, generating non-standard answers to pre-programmed questions. This will significantly increase the ability to perform the function of a personal assistant, including: (a) expand the planning horizon of the schedule, which will be able to take into account the limitations of family members, work schedules, climatic seasonality, etc.; (b) prepare sound recommendations in the field of a healthy lifestyle, including the composition and formulation of combined products that take into account gerontological, anthropometric, personal data, taste preferences, etc. It is obvious that the increasing scale of the use of assistant bots will allow for more targeted interaction with consumers, as well as the use of recommendation system algorithms for sellers (tasks at the interface with those described above in the first group). For example, based on the characteristics of food products (from the simplest composition of proteins, fats and carbohydrates, to the characteristics of the complementarity of individual products or the impossibility of their joint consumption), the problem of forming the most optimal portfolio of orders in hypermarkets delivering their products to customers can be solved. In such a situation, the linear constraint matrix may significantly exceed the limits defined in the article, which emphasizes the importance of practical application. At the same time, it is worth emphasizing that the customer flow of large e-commerce entities tends to grow sharply, a trend that may continue at least in the medium term. At the same time, even a decrease in the growth rate of Internet business in 3–5 years will most likely lead not to a recession, but to the release of relevant economic indicators to a plateau, while maintaining a significant amount of remote interaction between companies and customers in the long term.

The third group combines financial and economic tasks, where the speed of decision-making is one of the key factors for successful algorithmic trading on the exchange and over-the-counter markets. It is worth noting that this group of tasks is currently, on the one hand, one of the largest in terms of the number of market participants, since the vast majority of traders are focused on using various mathematical models when organizing their activities. On the other hand, the competition of available algorithmization methods and the high "cost of error" of trading, which is often carried out at the expense of customers or using margin trading schemes, make it difficult to test hypotheses, which inevitably occurs as part of the application of new machine learning methods.

It is worth emphasizing that the above groups do not exhaust the tasks where the simplex method can be effectively applied in matrix form on the GPU. These include traditional economic issues in the field of optimization, taking into account limited resources, as well as other, for example, creative tasks (Basadur Simplex).

Conclusion

The research we conducted has shown that transferring the implementation of the simplex method in matrix form to the GPU using the PyTorch library makes solving the problem much easier than using specialized programming tools such as CUDA, OpenACC, cuBLAS, etc. The question arises as to how good the results of this simpler method are. For the test calculations performed, when transferring the calculation to the GPU using the PyTorch library, the calculation time is reduced by 3–5 times, depending on the size of the task.

On the other hand, it is important to be able to speed up calculations using the simplex method for relatively small tasks and for ordinary workstations used in real economic calculations, rather than for super-large problem statements that can only be solved on a supercomputer, which corresponds to real practice, since even calculations carried out within the framework of economics do not provide examples of very large tasks for the simplex method. This paper presents three main groups of applied economic problems that can be effectively solved using the tools described in the article.

It should also be noted that CUDA technology allows us in almost all cases to achieve the highest efficiency of implementation on the GPU compared to PyTorch. However, CUDA is so complex that its use to solve some real problem is usually a separate big question. This work uses a less effective but much simpler tool. ■

References

1. Kanorovich L.V. (1939) *Mathematical methods of organization and planning of production*. Leningrad: Leningrad State University Press (in Russian).

2. Makarov V.L., Bakhtizin A.R., Sushko E.D., Sushko G.B. (2022) Creating a supercomputer simulation of society with active agents of different types and its testing. *Herald of the Russian Academy of Sciences*, vol. 92, no. 5, pp. 458–466 (in Russian).

3. Kolev M., Georgiadou S. (2022) On simulation and modeling in economics. *Asian-European Journal of Mathematics*, vol. 15, no. 10, article 2250239. https://doi.org/10.1142/s1793557122502394

4. Davydov I., Kochetov Yu., Tolstykh D., et al. (2023) Hybrid variable neighborhood search for automated warehouse scheduling. *Optimization Letters*, vol. 17, no. 9, pp. 2185–2199. https://doi.org/10.1007/s11590-022-01921-6

5. Chistyakova T.B., Shashikhina O.E. (2022) Intelligent software complex for modeling the process of planning multi-assortment industrial productions. *Applied Informatics*, vol. 17, no. 5(101), pp. 41–50 (in Russian). https://doi.org/10.37791/2687-0649-2022-17-5-41-50

6. Beklaryan L.A., Beklaryan G.L., Akopov A.S., Khachatryan N.K. (2024) Dynamic and agent-based models of intelligent transport systems. *Economics and Mathematical Methods*, vol. 60, no. 2, pp. 105–122 (in Russian). https://doi.org/10.31857/S0424738824020091

7. Mochurad L., Boyko N., Sheketa V. (2020) Parallelization of the method of simulated annealing when solving multicriteria optimization problems. Proceedings of the *CEUR Workshop, Lviv, May 21, 2020*, pp. 12–24.

8. Arrow K.J. (2008) George Dantzig in the development of economic analysis. *Discrete Optimization*, vol. 5, no. 2, pp. 159–167. https://doi.org/10.1016/j.disopt.2006.11.007

9. Ratushnyi A., Kochetov Y. (2021) A column generation based heuristic for a temporal bin packing problem. Proceedings of the *20th International Conference Mathematical Optimization Theory and Operations Research (MOTOR 2021), Irkutsk, Russia, July 5–10, 2021* (eds. P. Pardalos, M. Khachay, A. Kazakov). Lecture Notes in Computer Science, vol. 12755, pp. 96–110. Springer, Cham. https://doi.org/10.1007/978-3-030-77876-7_7

10. Kochetov Yu.A., Shamray N.B. (2021) Optimization of placement and relocation of emergency medical teams. *Discrete Analysis and Operation Research*, vol. 28, no. 2(148), pp. 5–34 (in Russian). https://doi.org/10.33048/daio.2021.28.702

11. Kovesnikov V.A., Mehtiev A.Ya. (2020) Study of the accumulative-sorting method for solving parametric optimization problems. *Control Issues*, no. 2, pp. 28–35 (in Russian). https://doi.org/10.25728/pu.2020.2.3

12. Manne A.S. (1958) A linear programming model of the U.S. petroleum refining industry. *Econometrica*, no. 26, pp. 67–196.

13. Chenery H.B. (1949) Engineering production functions. *The Quarterly Journal of Economics*, no. 63, pp. 507–531.

14. Makarov V.L., Bakhtizin A.R., Beklaryan G.L., Akopov A.S. (2021) Digital plant: methods of discrete-event modeling and optimization of production characteristics. *Business Informatics*, vol. 15, no. 2, pp. 7–20. https://doi.org/10.17323/2587-814X.2021.2.7.20

15. Scarf H.E. (1990) Mathematical programming and economic theory. *Operations Research*, vol. 38, no. 3, pp. 377–385.

16. Apalkova T.G., Kosorukov O.A., Mishchenko A.V., Tsourkov V.I. (2024) Mathematical models for managing the production and financial activities of an enterprise. *Herald of the Russian Academy of Sciences. Theory and Systems of Control*, no. 2, pp. 107–129. https://doi.org/10.31857/S0002338824020109

17. Gergel V., Grishagin V., Liniov A., Shumikhin S. (2021) Parallel computations in integrated environment of engineering modeling and global optimization. Proceedings of the *16th International Conference (PaCT 2021), Kaliningrad, Russia, September 13–18, 2021* (ed. V. Malyshkin), Lecture Notes in Computer Science, vol. 12942, pp. 413–419. Springer, Cham. https://doi.org/10.1007/978-3-030-86359-3_31

18. Gergel V., Kozinov E. (2021) Parallel computations for solving multicriteria mixed-integer optimization problems. *Communications in Computer and Information Science*, vol. 1437, pp. 92–107. https://doi.org/10.1007/978-3-030-81691-9_7

19. Shichkina Y., Kupriyanov M., Awadh AM.M.H. (2020) Application of methods for optimizing parallel algorithms for solving problems of distributed computing systems. Proceedings of the *International Conference on Cyber-Physical Systems and Control (CPS&C'2019), St. Petersburg, Russia, June 10–12, 2019* (eds. D. Arseniev, L. Overmeyer, H. Kälviäinen, B. Katalinić), Lecture Notes in Networks and Systems, vol. 95, pp. 212–224. Springer, Cham. https://doi.org/10.1007/978-3-030-34983-7_21

20. Makarov V.L., Bakhtizin A.R., Beklaryan G.L., Akopov A.S., Strelkovskii N.V. (2022) Simulation of migration and demographic processes using FLAME GPU. *Business Informatics*, vol. 16, no. 1, pp. 7–21. https://doi.org/10.17323/2587-814X.2022.1.7.21

21. Beklaryan A.L., Akopov A.S., Beklaryan L.A. (2021) Implementation of the Deffuant model within the FLAME GPU framework. *Advances in Systems Science and Applications*, vol. 21, no. 4, pp. 87–99. https://doi.org/10.25728/assa.2021.21.4.1161

22. Coutinho D.A.M., Lins e Silva F.O., Aloise D., Xavier-de-Souza S. (2019) A scalable shared-memory parallel simplex for large-scale linear programming. *arXiv:1804.04737v2*. https://doi.org/10.48550/arXiv.1804.04737

23. Mochurad L. (2020) Parallelization of the Simplex method based on the OpenMP technology. Proceedings of the *4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020), Lviv, Ukraine, April 23–24, 2020*, pp. 952–963.

24. Hall J.A.J. (2010) Towards a practical parallelization of the Simplex method. *Computational Management Science*, vol. 7, pp. 139–170. https://doi.org/10.1007/s10287-008-0080-5

25. Bieling J. (2010) An efficient GPU implementation of the revised simplex method. Proceedings of the *IEEE International Symposium on Parallel Distributed Processing, Workshops and PhD Forum (IPDPSW), Atlanta, GA, USA*, pp. 1–8. https://doi.org/10.1109/IPDPSW.2010.5470831

26. Bazaraa M.S., Jarvis J.J., Sherali H.D. (2010) *Linear Programming and Network Flows*. Hoboken, NJ: John Wiley & Sons, Inc.

27. Instituto Superior Técnico, Universidade de Lisboa (2025) *LP random problem generator*. Available at: https://web.tecnico.ulisboa.pt/~mcasquilho/compute/or/Fx-LP-generator.php (accessed 20 April 2025).

About the authors

Yuriy S. Ezrokh

Doctor of Sciences (Economics), Associate Professor;

Head of Department, Department of Economic Informatics, Novosibirsk State Technical University, 6, Karl Marx St., Novosibirsk 630073, Russia;

E-mail: ezrokh@corp.nstu.ru

ORCID: 0000-0002-8367-1840

Alexey V. Snytnikov

Doctor of Sciences (Technology);

Professor, Department of Applied Informatics, Kaliningrad State Technical University, 1, Sovetskiy Ave., Kaliningrad 236022, Russia;

E-mail: aleksej.snytnikov@klgtu.ru

ORCID: 0000-0003-4111-308X

Elena Yu. Skorobogatykh

Candidate of Sciences (Pedagogy), Associate Professor;

Associate Professor, Department of Applied Informatics, Kaliningrad State Technical University, 1, Sovetskiy Ave., Kaliningrad 236022, Russia;

E-mail: elena.skorobogatykh@klgtu.ru

ORCID: 0000-0001-6050-4831

DOI: 10.17323/2587-814X.2025.2.89.101

Ecosystem approach to strategic management on the example of agriculture

Vladimir I. Budzko
E-mail: vbudzko@ipiran.ru

Viktor I. Medennikov
E-mail: dommed@mail.ru

Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russia

Abstract

This paper considers the transformation of methods and models of strategic management based on the ecosystem approach within the framework of the formation of a unified digital platform for management. The ecosystem approach to the socio-economic development of society is gaining popularity as a result of global social requirements for environmental protection and a careful attitude to use of limited natural resources. Environmental problems in the Russian agricultural sector are increasing, in particular due to the process of formation of agro-industrial associations, mainly in the form of agricultural holdings. In this case, there is a problem of a systematic approach to using of technologies for the integration of all types of resources involved in production, taking into account the growing number and importance of environmental factors. Mathematical modeling is proposed as the main method of strategic management research. Unlike most of the existing models, which are often iconographic, the model proposed allows us to consider a larger number of factors. This makes it possible to assess various options of the modeling objects development using a simulation approach. As a result of the research, a mathematical model of strategic management of agroholdings for sustainable development was developed. It is shown that the development strategy should be implemented considering an appropriate automated management information system. This will lead to a radical change in the whole system of management and production. It will allow the enterprise to apply strategic goal-setting, focusing first of all on quality, controllability and other components of competitiveness. The mathematical model proposed provides justification of unified methods of long-term digitalization applicable for large agricultural associations, as well as for small and medium-sized farms, which will be able to cooperate with agricultural holdings relying on the principles of outsourcing.

Keywords: strategic management, ecosystem, agriculture, digital management platform

Citation: Budzko V.I., Medennikov V.I. (2025) Ecosystem approach to strategic management on the example of agriculture. *Business Informatics*, vol. 19, no. 2, pp. 89–101. DOI: 10.17323/2587-814X.2025.2.89.101

Introduction

The ecosystem as a scientific concept stems from biology, where a system is understood as a symbiosis of physical objects and biological organisms in some habitat in which they are interdependent and interact effectively [1–3]. The ecosystem approach is a concept based on new digital solutions aimed at improving the efficiency of control actions in the field of environmental protection. The ultimate goal is to transfer to the descendants as many natural resources as possible, while reducing the environmental burden on mankind and the natural environment.

A natural ecosystem can be defined as a system consisting of living and non-living elements interacting with each other in certain spatial areas [4]. When transferring the concept of ecosystem to socio-economic systems, the first part of this term (“eco”) means the environment of some organizational structure, and the second part (“system”) describes (from the point of view of the systems theory) the interrelation of elements functioning according to certain rules. Thus, in socio-economic research an ecosystem is defined through exogenously specified elements of the environment, as well as agents endogenously acting together according to the rules of the system, to the benefit of the participants of such interaction [5, 6].

Nowadays, research in the field of using the ecosystem approach regarding socio-economic development of mankind is becoming more and more important due to the rapidly increased global social demand for environmental protection, as well as for more efficient and environmentally friendly use of dwindling natural resources. This was confirmed by the establishment by the UN General Assembly (in 1987) of the International Commission on Environment and Development, which developed the concept of sustainable environmentally friendly development of the entire world community. In addition to the social demand mentioned above, the concept considers other complementary factors, including demographic, financial and social ones.

Environmental problems in Russia are increasing. In agriculture, the control by regulatory bodies over compliance with soil fertility preservation norms has decreased. The rules of crop rotation are not observed; chemical fertilizers and substances against diseases and pests are widely used, with the aim to increase the volume of production and its profitability. All this negatively affects not only harmful organisms but also beneficial ones interacting with them. Such external impact on ES and non-observance of some rules of the state of soil equilibrium, taking place over a quite long period of time, turns the soil into a warehouse of toxic chemicals with their further transmission through crops and animals to humans themselves. Thus, in [7] it is reported that in Russia about 2 million hectares of land are degraded per year, which leads to production losses of 3.9 million tons. Consequently, the fertility of chernozems in central part of Russia has almost halved over the last ten years.

Conducting crop rotation with ensuring the necessary level of an agricultural producer's profitability requires the development of long-term and current plans, as well as their subsequent regular adjustment and management, taking into account changing conditions. For each specific enterprise, it is impossible to develop an individual digital tool for solving this task, because of the increased number of internal and external environmental factors. Therefore, some analog of the computer aided design system is essential. An integrated management information system (MIS) should provide information support for such a process, and an ecosystem approach is fully applicable for its construction. At the same time, since market relations require seeking the most rational solutions while determining areas of activity, building the management information system is related with the best methods and approaches developed in the field of strategic management. The concept of sustainable environmentally friendly development, along with the political and economic instability worldwide during the last decade, has led to active research in the field of strategic management. Therefore, integrated crop rotation management requires creation of business ecosystems with elements of strategic management [8].

1. Models and methods of strategic management

The theory of management leads to formation of strategic management as an independent integral scientific field [9, 10]. In these works, planning had a short-term annual horizon, the reason for which was the general economic stability in developed countries, in particular, in the relationship with the external world (environment) and provision of various resources. The extension of the planning horizon from one to several years was a consequence of the growing problems in the influence of various factors of the external environment on the work of businesses.

Accelerated development of science, technology and engineering, as well as the globalization of the economy (the appearance of production facilities with globally distributed assembly sites for individual product components) force most firms to engage in long-term planning and appropriate management. The development of the concept of strategic management became the result of such trends and research has surged in this area. Among the most famous foreign scientists who have worked in this field, we can mention Strickland and Thompson [11], as well as Mintzberg, Albert, Porter, Bowman, King, Drucker, Ansoff, Mescon, Khedouri, Cleland, Ahlstrand, Lampell, Chandler. It is also possible to distinguish such Russian specialists as Vikhansky [12], Shekhovtseva [13], Terekhova-Pushnaya [14] and some others.

For some time, the models of strategic management were subjected to various modifications [13]. As a result of the evolution, a large number of models based on graphics and flowcharts appeared. Quite often the difference between models was only in the number and content of consecutive stages of strategic management description. At the same time, all models have some common descriptions of the stages of SM, which include:

- goal setting of the firm's activity;
- allocation of a number of tasks aiming to achieve strategic goals;
- selection of tasks (from a certain set of options) that are the most important (optimal) from the point of view of strategy implementation;

♦ implementation of the tasks selected;

♦ monitoring, analysis and evaluation of the implementation of the approved trajectories focused on strategic goals.

In all the models, the most important function of strategic management (which is essential for our models) is analyzing internal resources, as well as features of the external environment that are significant for companies, including economic, social, political, environmental, technological and some other factors. Recent events show that the external environment is becoming more and more dynamic, forcing us to seek answers to a range of questions:

♦ what is the portrait of a prospective buyer of the company's products;

♦ to what extent can the company's products be substituted by products of other firms, considering the financial condition of buyers in the future;

♦ what is the price distribution of goods depending on regional wages, delivery volumes and the level of logistics development;

♦ what are the level and trends of product development in terms of public, business and regulatory requirements for quality, health safety, design and prospective characteristics;

♦ what are the conditions for obtaining loans and their impact on the cost of goods, the level of competition, its dynamics and expected forms;

♦ what is the degree and quality of adoption of digital technologies by expected competitors;

♦ what is the forecast of trends in the rulemaking of the state institutions affecting the company's interests.

As a result of attempts of different organizations to take into account most of the factors of the external environment listed above, the theory of business ecosystems was developed [8].

2. Methods of strategic management based on a single digital platform

James F. Moore, the founder of the theory of business ecosystems, gives the following definition of a business ecosystem [8]. By analogy with natural ecosystems, a business ecosystem is an economic community that includes (according to certain criteria, as is customary in system analysis) a group of interacting actors (organizations and individuals) from the business world. This means that an ecosystem of any organization includes both its own divisions and related entities, including customers, suppliers, intermediaries (participants of value added chains), as well as all kinds of departments, bodies and institutions which provide monitoring of compliance with standards, regulations and laws. Moreover, competitors that have or may have a significant impact on the ecosystem, as usual, are also included in the ecosystem. The influence of all these elements on the entire ecosystem should be considered at all management levels, relying on appropriate exchange of information.

In [7, 15, 16] a unified digital management platform (DMP) is studied for developing a typical MIS for multi-sector agricultural enterprises. In the light of the above mentioned arguments, this platform claims the role of a digital tool for implementing the concept of business ecosystems in practice. This solution also represents a basis for developing a strategic management system. The DMP consists of three main sub-platforms that could be adopted as digital cloud standards:

♦ collection, transmission and storage of primary accounting information (*Fig. 1*);

♦ database of technological information (*Fig. 2*);

♦ knowledge base reflecting the algorithmic support of management functions (in crop production, 240 typical algorithms are ontologically identified).

An example of enlarged modular information model of crop production in agricultural farms is given in *Fig. 2*.

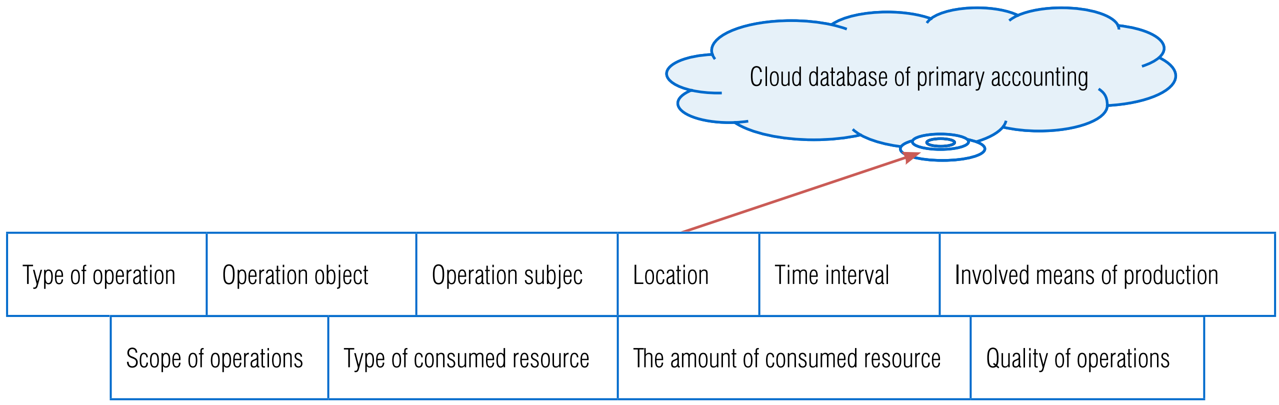


Fig. 1. Modules of the sub-platform for collecting, transmitting and storing primary accounting information.

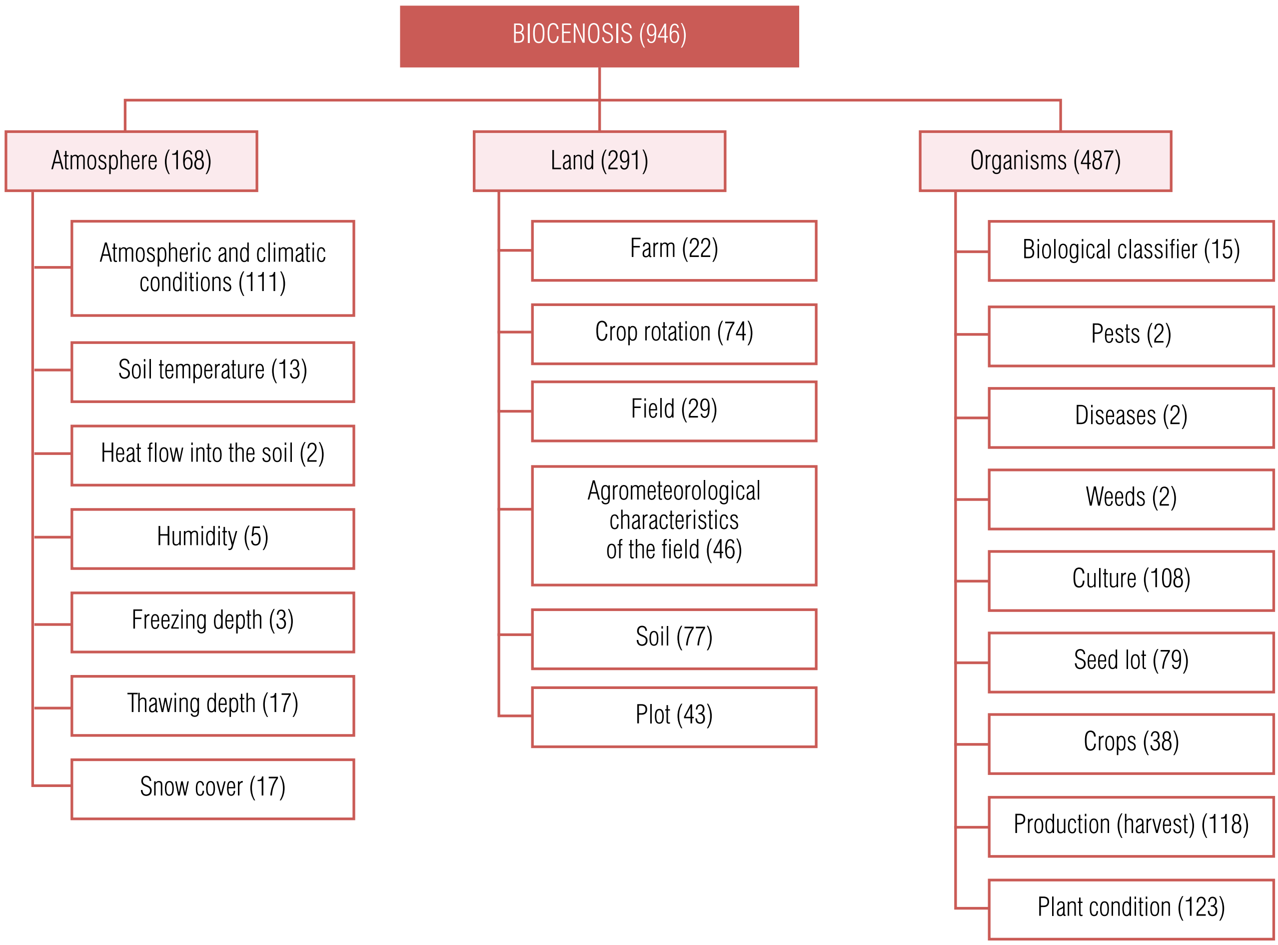


Fig. 2. Large-scale modular information model of crop production.

The sub-platform for collecting, transmitting and storing primary accounting information has an inter-industry universal nature for the main production sectors. Two other sub-platforms represent digital standards of a specific industry. DMP also contains all the master data and classifiers that are necessary for management algorithms. It should be noticed that the universality of the first sub-platform, implemented as an appropriate toolkit, is related with storing all the technological operations performed and all the information collected using the Internet of Things. This may explain the argument advanced by academician Glushkov in favor of the OGAS project ("to overtake without catching up") since conceptually the DMP is far ahead of the overseas approach to the digitalization of the economy [17, 18]. This is related with the statement of J'son & Partners Consulting that at present in the US agricultural sector there are only two types of specific sub-platforms: aggregators collecting and accumulating primary data and sub-platforms of applied tasks [19].

Let us consider the expected results of the application of the MIS implemented on a single DMP regarding the development of various strategic management systems.

As noted above, the business ecosystem of any enterprise includes many economic agents in contact with it, as well as government and regulatory institutions that provide compliance regarding laws and standards. Growing integration trends of the digital economy significantly strengthen this relationship. It is possible to note the emerging requirements of market agents and private buyers to receive online information about various characteristics of goods (for example, in Russia the Honest Sign system is actively developing). Such requirements based on digital economics integration technologies have been realized in the concept of traceability of goods and services. A digital tool implementing the traceability concept allows us to provide suitable information about the manufacturer, composition of the goods (with quality, price, cost indicators), production date and expiration date, and other data, not only to participants of production and logistics chains, but also to all stakeholders, including regulatory authorities. As a result of the spread and improvement of this tool, customer behavior patterns in digitally developed countries is changing towards choosing higher quality and safer goods, medicines and food products. As follows from the considered characteristics and capabilities of the DMP, an ideal tool for providing traceability of goods and services is MIS developed on the DMP basis [20], which also includes a digital platform of logistics management [21] (*Fig. 3*).

Moreover, it can be concluded that DMP, in combination with a number of other digital technologies, such as artificial intelligence allows us to make reliable and objective calculations of the individual contribution of each participant in the value chain to the cost of production. This allows us to draw the following conclusion: the business behavior of all participants in the chain will undergo a radical transformation caused by full awareness of the real contribution of each of them when transferring their products to the chain [20]. We forecast that in 15 to 20 years, when DMP will be recognized due to its practical efficiency, it can be predicted that this platform will become an instrument for changing the service model of socio-economic relations to a product model of behavior for the whole society [22].

3. Methods of strategic management of an agro-industrial association based on a single digital platform

Scientific interest in strategic management in Russian agriculture sector is explained (behind tasks of a general nature) by growing environmental problems and declining soil fertility. This interest is also facilitated by ongoing structural reforms, which leads to rapid increase in agricultural holdings bringing together enterprises from different industries and regions with different levels of development. The

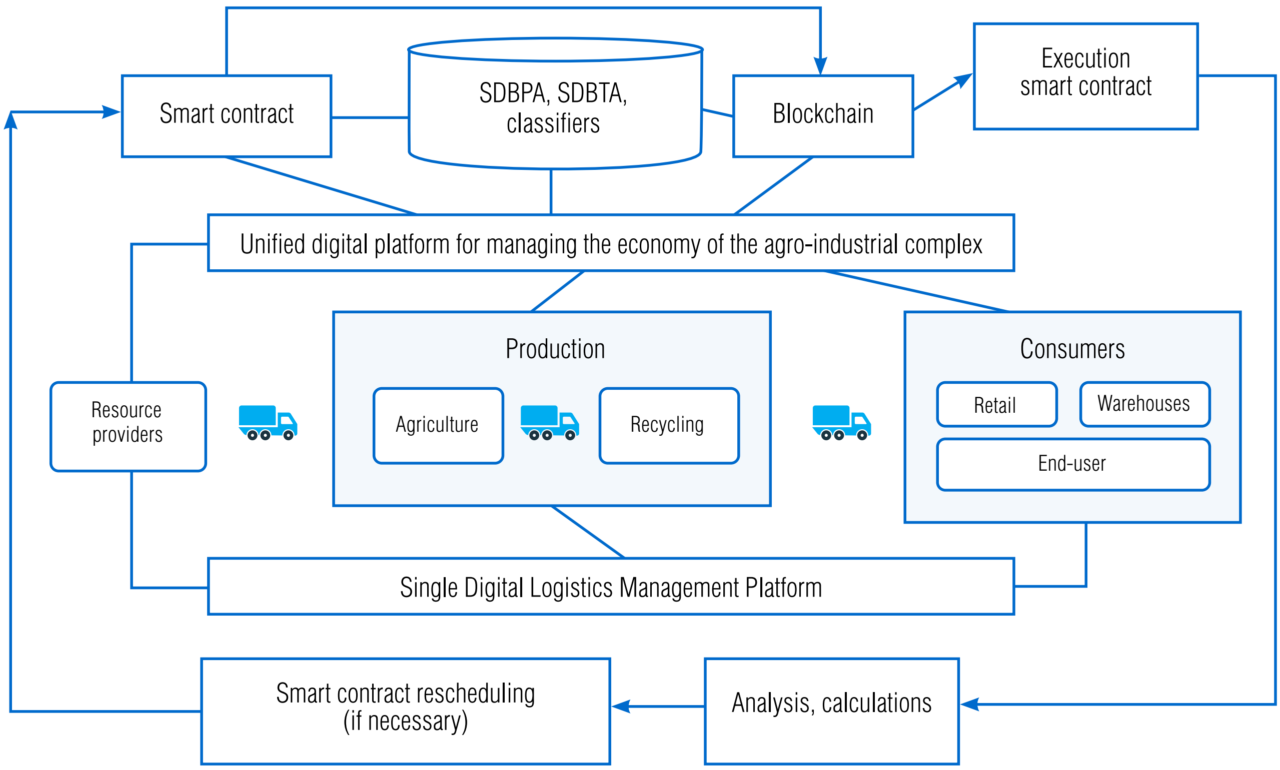


Fig. 3. Scheme for implementing product traceability within a single automated control system based on the DMP.

main dominant motive for this is the need to integrate agricultural organizations with more profitable livestock processing enterprises. For these enterprises, such relationships guarantee the access to relatively cheaper (comparing with market process) raw materials and essential technological stability. Often these enterprises act as initiators of such integration. This leads to reduction of market intermediaries' participation in supply chains and, in the case of organizing own retail chains – to implementing the principle "from the field to the counter" (this principle is similar to appropriate trends in foreign countries).

The creation of holdings gives rise to a number of problems associated with the alignment of production facilities, as well as providing effective integration of all the resources (material, labor, financial and digital resources essential in the era of digital economy) while forming optimal logistics chains.

To resolve the problems listed above, let us discuss the strategic management model of an agricultural holding, considering its integration trends.

As in has been mentioned, holdings include the crop, livestock and processing industries of the agro-industrial complex. To shorten the model, we do not consider secondary industries that carry out auxiliary functions, and we take into account their impact on total costs within the main industries. However, in the future (considering that the holding has completely switched to the DMP) it is mandatory to include in the model the logistics block, as an important element of the integration processes. Further, we assume that only crop and livestock products serve as raw materials for processing enterprises, and only crop products serve as raw materials for the livestock industry (although, in practice, some processing waste is included in the feed ration of some animals). Due to quite strict crop rotation requirements, to produce such a set of plant crops only for the needs of the livestock and processing industries is not possible. Therefore, additional products should be supplied either to the market or to processing. The scheme of factors taken into account when developing a strategic management system for an agroholding is shown in *Fig. 4*.

Considering the assumptions described, there is a problem of defining an optimal prospective structure for the production of the holding's final products. This makes it possible to maximize profit, using effective integration technologies regarding all the resources used in production. Moreover, it is essential to take into account investments in the economy, production technologies, as well as mechanisms for achieving the required level of product competitiveness (*Fig. 5*). The next stage of strategic management will allow us to introduce a program for transition from the initial state of the agroholding to the final state obtained as a result of strategic planning of the first stage [23].

It is known that the main processes of agricultural production are determined by the structure of scientifically justified crop rotation [24]. The choice of crop rotation determines technologies of land processing and caring for plantings, the scope of measures

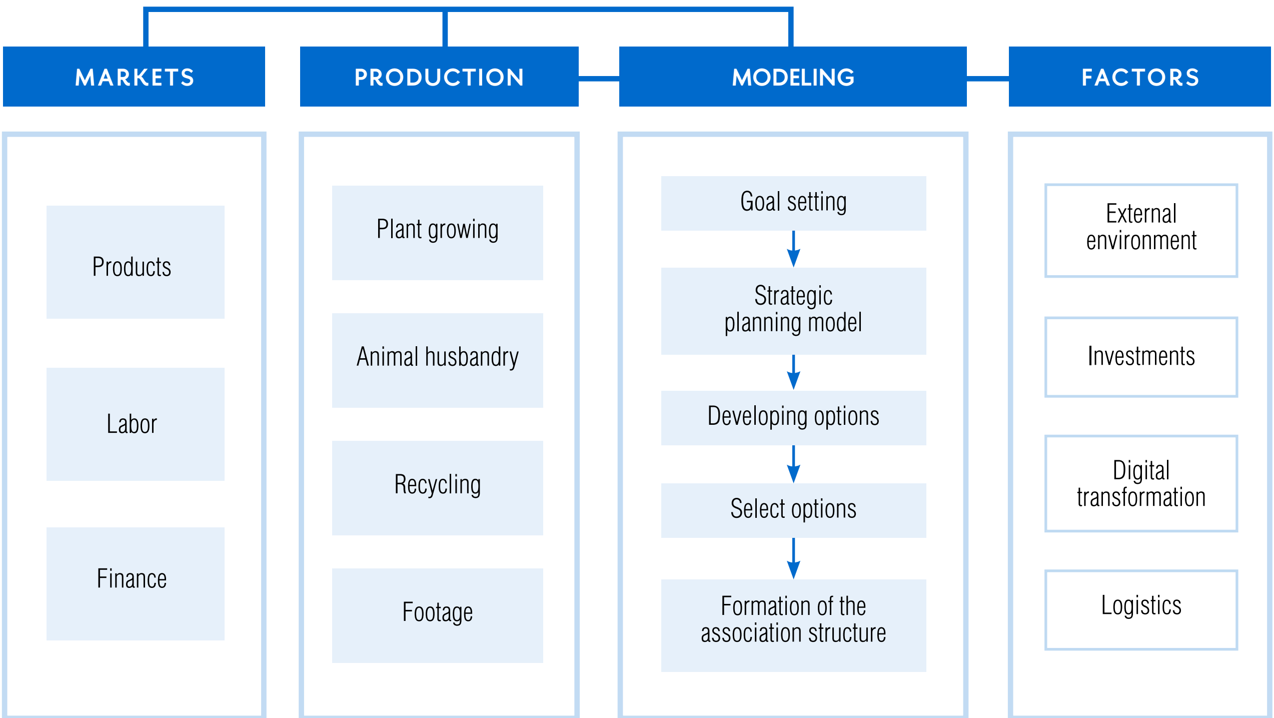


Fig. 4. Factors taken into account in the strategic management of an agricultural holding.

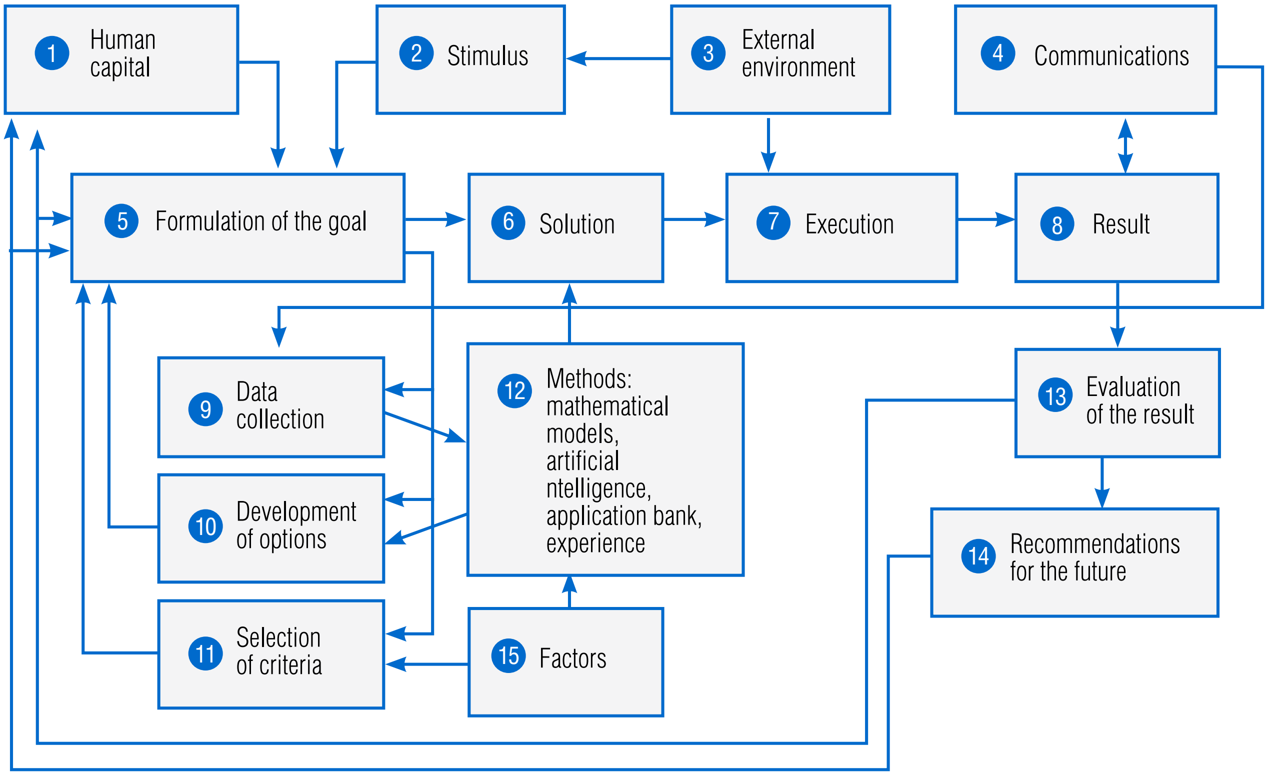


Fig. 5. Block diagram of strategic management of the holding.

for cultivated plants protection, the list of chemicals planned for use, the brands and nomenclature of technical means, the structure and territorial distribution of warehouses and production buildings, etc. It can be said that the final result of agricultural production as whole depends on the selected structure of crop rotation, in which land is the main resource. Crop rotation provides rational and efficient use of other resources applied for growing crops. Therefore, the formation of strategic management systems of agroholdings should be preceded by the development of scientifically approved crop rotation, the mathematical model of which is presented in [25].

For this reason, all the technological processes must be informationally and algorithmically interconnected in accordance with the integration principles. The same is relevant to logistics, the digital technologies of which must act as a link between all crop rotation elements – suppliers of materials, fuel and chemicals, temporary storage points (warehouses, silo tanks, haylage towers, grain threshing floors), all kinds of consumers of plant products (livestock farms, processing and feed mills, elevators, traders), as well as transportation entities themselves. The relevance of considering logistics as an element of crop rotation for the purpose of its optimization is confirmed by data on the huge costs for it in Russia – about 20% of GDP [21] (for comparison, in the EU these costs are 2.5 times lower).

Let us consider the mechanism for integrating the logistics block into the strategic management model, since it has a significant impact on the interregional and intra-farm life of the agricultural holding. Due to the nonlinear behavior of the logistic model, it seems reasonable to use the logistics block when designing crop rotation in the simulation mode (*Fig. 6*), one of the variants of which is described in [21].

The mathematical model of strategic management considered in [25] refers to static linear models, which, due to the existing powerful mathematical apparatus, are used in the overwhelming majority of studies in the field of strategic planning. Although agriculture is a weakly structured system with significant nonlinear dependencies, their formalization is quite complex. For example, the Liebig principle, published by the German scientist Justus von Liebig back in 1840, is still in use, and neither artificial intelligence, nor mathematical nonlinear methods can improve it. In practice, according to academician Dorodnitsyn, over 98% of mathematical models are linear. For example, back in 1970 to refute the statement of Nobel laureate Robert Solow about the absence of an economic effect from the computerization of the economy, the Economist Intelligence Unit used a simple Cobb–Douglas function, which becomes linear after taking its logarithm.

However, while using such an approach, the most serious problem in strategic management modeling tasks is finding the optimal process of transition from a known initial state of the system to the target state determined by strategic planning. Most works consider strategic management only as developing a long-term plan. A mathematical model of the second stage of the transition processes optimization is given in [25].

Conclusion

The approach to optimization of strategic management in agricultural production is discussed in this work. The basis of this approach is a management information system, which should be created in each agricultural holding on the basis of a single DMP of production. After the implementation of the DMP in the selected holdings this approach can be extended to other enterprises and farms, which could act as some additional links in the global supply chains of holdings on the principles of outsourcing. The construction of an MIS of enterprises on a single DMP will allow an entire agro-industrial complex to arrange an effective transition to organic farming with minimization of environmental risks caused by the harmful effects of natural and human factors on the products.

Successful implementation of the solutions implemented in DMP requires the availability of structured and ontologically verified information: on completed technological operations (for example, using the Internet of Things), on monitoring results, on the state of cultivated and harmful plants and soil, on the required volumes of equipment and chemicals applied, as well as other information stored in the DMP database that is essential for implementation of farming technologies. Thus, in crop production almost all cultivated plant experimental technologies for the introduction of nutrition and protection of products are being developed. All this information should be collected

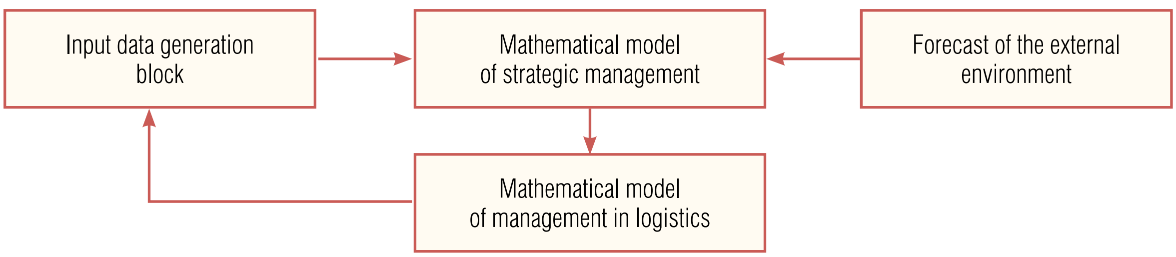


Fig. 6. Scheme of embedding the logistics block into the strategic management model.

and used, however the problem is complicated by the refusal of large enterprises to provide the necessary information even to the Ministry of Agriculture. This fact was manifested in attempts to test the model in practice, taking into account the accumulated experience within the framework of the "Electronization of Agriculture" program, in collaboration with experts from the Kuban Research Institute of Agriculture, which has large long-term experimental material on the formation of crop rotations.

The model was tested for adequacy on one of the farms in Krasnodar Region, which has over 7 thousand hectares of land (with two groups), for which recommendations for the use of crop rotation were available. The implementation, confirmed by the State Committee for Science and Technology document, showed the economic effect of increasing the yield of grain and leguminous crops by 4 c/ha without additional costs.

This experience has shown that violations of crop rotation in Russia and the lack of necessary accounting of technological operations do not allow collecting the initial information required for the model. Although the highly developed agriculture of Kuban Region significantly surpasses other regions in terms of information, it was not possible to collect all the information necessary to implement the full model, so a significantly simplified model was implemented. In this case, the necessary data had to be restored by statistical methods based on data from other farms. This is evidenced by studies showing that over 85% of the industry information is stored either on paper, or in spreadsheets. Moreover, in many farms important information is not stored at all.

This was the reason for the lack of sensitivity studies of the model. It was important to check its behavior when forming crop rotations for the choice of fields depending on the continuous or scattered method of fields placing, since this significantly affects fuel consumption, time characteristics and overall production efficiency. The continuous method is used in the case of large areas of land of one agro-ecological group located on a geographically uniform array and claiming to accommodate one of the crop rotations. Such geography and the physical and chemical properties of the soil suggest placing all fields directly next to each other. The scattered method is used in the case when fields belonging to one agro-ecological group are placed on land plots geographically isolated from each other. Simulation calculations have shown the importance of this point, so work is being carried out regarding an optimization model, which is reduced to a variation of the problem of packing in containers. ■

References

1. Budzko V.I., Medennikov V.I. (2023) Systems analysis of educational digital ecosystems in the agro-industrial complex. *High availability systems*, no. 1(19), pp. 46–58 (in Russian).

2. Medennikov V., Flerov Y. (2022) System analysis of digital ecosystem of Russian agriculture. *IEEE Xplore Digital Library. 15 International Conference Management of Large-Scale System Development (MLSD), Moscow, Russia*.

3. Ramenskaya L.A. (2020) Application of the concept of ecosystems in economic and managerial research. *Manager*, no. 4(11), pp. 16–28 (in Russian).

4. Tansley A. (1935) The use and abuse of vegetational concepts and terms. *Vegetational Concepts and Terms,* pp. 284–307.

5. Acs Z.J., Audretsch D.B., Lehmann E.E., Licht G. (2016) National systems of entrepreneurship. *Small Buzines Econimics,* vol. 46, pp. 527–535.

6. Acs Z.J., Audretsch D.B., Lehmann E.E., Licht G. (2017) National systems of innovation. *Journal of Technology Transfer*, vol. 42, no. 5, pp. 997–1008.

7. Budzko V.I., Medennikov V.I. (2022) Mathematical model for optimizing crop rotation structure based on a unified digital platform for agricultural production management. *High availability systems,* no. 4(18), pp. 5–13 (in Russian).

8. Moore J. (1996) *The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems*. Harper Business.

9. Taylor F.W. (1991) *Principles of scientific management*. Moscow: Controlling (in Russian).

10. Fayol A. (1992) *General and industrial management*. Moscow: Controlling (in Russian).

11. Thompson A., Strickland A. (1998) *Strategic Management: The Art of Developing and Implementing Strategy*. Moscow: Banks and Exchanges (in Russian).

12. Vikhansky O.S. (1998) *Strategic Management: Textbook*. 2nd ed., revised and enlarged. Moscow: Gardarika (in Russian).

13. Shekhovtseva L.S. (2006) *Strategic Management*. Kaliningrad: Publishing House of the I. Kant Russian State University (in Russian).

14. Terekhova-Pushnaya D.V. (2019) Models of strategic management and planning. *Moscow Economic Journal*, no. 9, pp. 488–497 (in Russian).

15. Alekseeva N.A., Osipov A.K., Medennikov V.I., et al. (2022) *Economic and managerial problems of land management and land use in the region*. Izhevsk: Shelest (in Russian).

16. Medennikov V.I. (2019) Mathematical model of the formation of digital platforms for managing the country's economy. *Digital Economy*, no. 1(5), pp. 25–35 (in Russian).

17. Glushkov V.M. (1975) *Macroeconomic models and principles of OGAS construction*. Moscow: Statistics (in Russian).

18. Benjamin P. (2016) *How not to network a nation: The uneasy history of the Soviet Internet*. The MIT Press.

19. J'son & Partners Consulting (2018) *Analysis of the market of cloud IoT platforms and applications for digital agriculture in the world and prospects in Russia*. Available at: https://json.tv/analytic/analiz-rynka-oblachnyh-iot-platform-i-prilojeniy-dlya-tsifrovogo-selskogo-hozyaystva-v-mire-i-perspektiv-v-rossii-20180925034341/ (accessed 18 June 2025).

20. Kulba V., Medennikov V. (2022) Product traceability digital tool powered by mathematical model for logistics digital platform. *IEEE Xplore Digital Library. 15 International Conference Management of Large-Scale System Development (MLSD)*. Moscow, Russia (in Russian).

21. Budzko V., Medennikov V., Keyer P. (2023) The logistics component of the geographic reference of the formed promising crop rotations. *BIO Web of Conferences*, vol. 66, article 03002.

22. Medennikov V.I. (2024) Evolutionary stages of the ecosystem approach to economic management in the digital era. *Management Accounting*, no. 7, pp. 310–318 (in Russian).

23. Pospelov G.S., Irikov V.A. (1976) *Program-targeted planning and management*. Moscow: Soviet radio (in Russian).

24. Evtefeev Yu.V., Kazantsev G.M. (2013) *Fundamentals of agronomy: A textbook*. Moscow: FORUM (in Russian).

25. Kulba V.V., Medennikov V.I. (2021) Taking into account regional features of strategic management of agro-industrial integrated formations in a single space of digital interaction. Proceedings of the *Fourteenth International Conference Management of the development of large-scale systems (MLSD'2021), Moscow, September 27–29, 2021.* Moscow: V.A. Trapeznikov Institute of Control Sciences, pp. 70–82 (in Russian).

About the authors

Vladimir I. Budzko

Doctor of Sciences (Technology), Academician of Cryptography Academy of Russian Federation;

Chief Researcher, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2, Vavilova St., Moscow 119333, Russia;

E-mail: vbudzko@ipiran.ru

ORCID: 0000-0002-8235-0404

Viktor I. Medennikov

Doctor of Sciences (Technology);

Leading researcher, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44-2, Vavilova St., Moscow 119333, Russia;

E-mail: dommed@mail.ru

ORCID: 0000-0002-4485-7132