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

Vol. 18 No. 1 – 2024

BUSINESS INFORMATICS

HSE Scientific Journal

CONTENTS

D.E. Palchunov, A.A. Yakobson
Development of an intelligent assistant for selection of goods in the process of dialogue with the user 7

M.E. Oshchepkov
Improving target budgeting in a corporate performance management system 22

G.L. Beklaryan
Agent-based modeling and optimization of the characteristics for research-and-production clusters 36

P.A. Koldanov, V.A. Koldanov
Nonparametric procedure for comparing the performance of divisions of a network organization 52

K.N. Hardanti, Sutrisno T., E. Saraswati, A. Prastiwi
Determinants of an auditor's continuance intention with respect to use of the Audit Tools and Linked Archives System (ATLAS): A model of extended expectation confirmation... 65

T.C. Luong , O.T. Tran
Product information recognition in the retail domain as an MRC problem 79

Publisher:
HSE University

The journal is published quarterly

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

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

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

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

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

ABOUT THE JOURNAL

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

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

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

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

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

The journal is distributed both in printed and electronic forms.

EDITORIAL BOARD

ABOUT HSE UNIVERSITY

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

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

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

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

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

ABOUT HSE GRADUATE SCHOOL OF BUSINESS

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

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

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

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

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

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

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

DOI: 10.17323/2587-814X.2024.1.7.21

Development of an intelligent assistant for selection of goods in the process of dialogue with the user

Dmitry E. Palchunov [a]
E-mail: palch@math.nsc.ru

Alexander A. Yakobson [b]
E-mail: a.yakobson@g.nsu.ru

[a] S.L. Sobolev Institute of MathematicS, Siberian Branch of the Russian Academy of Sciences
Address: 4, Akademika Koptyuga Ave., Novosibirsk 630090, Russia

[b] Novosibirsk National Research State University
Address: 1, Pirogova St., Novosibirsk 630090, Russia

Abstract

This article is devoted to the development of methods for creating intelligent assistants. Intelligent assistants can be used in call centers to solve customer problems, to solve technical support tasks, to help people with disabilities, to help in choosing goods, etc. We consider intelligent assistants that engage in argumentative dialogue with users, aimed at finding goods and services that maximally satisfy users' wants and needs. The development of the intelligent assistant is based on a four-level model of the subject domain and a semantic model of the user. The system under development automates the

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

process of search and decision justification through the reuse of domain cases: accumulated knowledge about previous dialogues with users. This gives the system we developed an advantage over existing analogues, which are incapable of reusing knowledge about previous dialogues. The paper develops a case-based approach to building an intelligent system capable of reasoning about its responses. For this purpose, an argumentation graph is constructed, methods for structuring domain cases are developed, and ontological homomorphisms are used to transform the available domain cases into a finished solution. A description of model-theoretical methods for constructing intelligent assistants is presented. The cases of goods, users and dialogues of an intelligent assistant with users are formally described in the form of partial models. The transformation of domain cases and similarity of cases are formalized using ontological homomorphisms of partial models. The purpose of the developed dialogue system is not only to select a solution according to the user's request, but also to find out the tasks that the user is going to solve, to analyze his argumentation, and then to justify the proposed solution to the user, to show that this particular product or service will be able to meet his needs.

Keywords: intelligent assistant, argumentative dialogue, domain case, partial model, ontological homomorphism, ontological model of the subject domain, semantic model of the user

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

Introduction

The world is currently experiencing the peak of popularity of artificial intelligence (AI) technologies. The success of ChatGPT as a universal dialogue system has shown the need of people to have an intelligent tool for solving various tasks. ChatGPT solves many tasks: from writing texts for social networks (with great success) to creating scientific papers (with not such great success).

In most cases, such systems act as extremely advanced and powerful content compilers; they search through existing data and piece by piece assemble the required result from them. At the same time, the very concept of neural networks imposes a very specific limitation on them: they rely on their own trained model, on the data entered into the training sample in advance. As a result, they cannot use recent results of their own work to improve the process of finding a solution, because retraining a neural network is a long and resource-intensive process. Also, if any fundamentally new object appears in the subject domain (for example, a new style of drawing, if we are talking about a neural network that creates images), the neural network will not be able to obtain the same result on its own, because this new data was not incorporated into it during training.

The situation is additionally complicated by the fact that neural networks are a "black box." It is practically impossible to interpret the process of their work, especially for large industrial instances. One of the possible solutions to the problems of the neural network approach is the construction of a logical system based on semantic analysis and structuring domain cases: previous sessions of work of the intelligent system both with this user and with all previous users.

1. Tasks of the intelligent assistant

Our goal is to develop methods for reasoned dialogue between an intelligent assistant and a user in order to help the user achieve goals, realize his/her intentions, satisfy needs and solve problems. In this paper, we primarily consider intelligent assistants that help users in selecting appropriate products and services.

To achieve this goal, the following tasks need to be performed in an automated manner:

- identifying the user's interests, needs, desires, goals and intentions;
- finding out how the user achieves their goals, solves their tasks, fulfils their intentions and satisfies their needs (e.g. the intention to purchase a desired product);
- identifying the user's justification, explanation, reasoning why, for example, he/she needs this particular device; finding out the specific tasks that the user is going to perform with this device (e.g. viewing and editing photos, cleaning the room or controlling a smart home);
- selecting for the user the product or service that best suits the user's tasks and needs;
- building an argumentation, justifying that a given product or service is indeed the best for the user (subject to the fulfilment of product price constraints and other non-functional requirements), or offering the user a set of products that are best suited for solving their problems;
- explanation of the differences between these products, their positive and negative qualities (in comparison with each other) in terms of solving the user's tasks and meeting his/her needs.

We apply modern argumentation theory [1–9] to develop methods for argumentative dialogue between an intelligent assistant and a user.

This article is primarily devoted to the presentation of methods and technologies of selection by the intelligent assistant of the goods most suitable for the user in the process of dialogue with the user. A more detailed description of the methods of building justification and argumentation of the fact that this product is the best for the user will be the subject of the next article.

The necessary information for the dialogue with the user of the intelligent assistant is taken from the semantic (ontological) model [10–12], the structure of which we will describe below. The semantic model is filled and replenished by extracting information from the websites of product manufacturers and online shops, as well as by analyzing customer reviews of products they purchased.

In developing the ontology model, we use a number of ontologies. These are:

- ontology of the subject domain as a whole;
- ontology of characteristics, properties, functionalities of various goods and devices;
- user ontology: user tasks, goals and intentions; what goals users achieve and in what ways.

What is extremely important in the approach we developed is that we save and analyze dialogues with the user. This is a significant difference between this approach and most existing solutions.

For example, Alice (a virtual voice assistant created by Yandex, YandexGPT 2), when having a dialogue with a user, does not "remember" even the previous line or question of the user. If the user says: “Alice, put me some song by band X”, Alice will put a song by this band. But if the user says, “Alice, I like songs by band X. Put me some song of this band.”, then Alice answers: “I have nothing to answer”.

If you ask: “Alice, what is the title of this song?” while a song is playing, she will answer. If you ask: “Alice, put the previous song on”, she will. But if you ask: “Alice, what is the name of the previous song?”, she will not be able to answer.

The intelligent assistant we are developing for selecting products for the user and for generating arguments can address:

- to the entire current dialogue with this user;
- to previous dialogues with this user;
- to previous dialogues with other users.

In this way the intelligent assistant works with cases of previously conducted dialogues. They are on the third level of the four-level ontological model, which will be described in detail below.

Currently, various organizations, such as online shops, banks, etc., use virtual assistants designed to help the user find the right product or service. However, as a rule, these systems work according to a predetermined scenario and when a new situation arises that was not foreseen in advance, they are unable to assist the client and redirect him/her to interact with a human consultant.

This situation, in which the virtual assistant is unable to find a solution or give the right recommendation, decreases the user’s motivation to work with it in the future. In addition, such systems almost always conduct a dialogue from scratch, without remembering the user and the context of the dialogue. If the user has already approached a similar problem, he or she must go through the whole process of searching for a solution again.

The system we are developing automates the solution search process by reusing previously accumulated cases (situations, domain cases). By comparing the current user’s goal and information from previous dialogues, it is possible not only to find a similar solution in the past, but also to additionally argue the proposed solution based on the coincidence of intentions. This gives the developed system an advantage over existing analogues, which are incapable of accumulating cases and arguing their solutions.

In this paper, we develop a case-based approach to building an intelligent system capable of reasoning about its answers. For this purpose, we construct an argumentation graph, develop methods for structuring domain cases and use ontological homomorphisms to transform the available cases into a ready-made solution. The goal of an intelligent dialogue system is to help a person to find an answer to a particular question. Our task is to circumvent the limitation of the neural network approach, which is the inability to take into account the results of recent user sessions. For this purpose, we implement a case approach to solution construction using ontological homomorphisms. On its basis, the construction and reasoning of the solution takes place.

2. Existing approaches and solutions

Currently, there are many dialogue systems designed for different tasks: systems that support dialogue with simple phrases, voice assistants (Alice, Siri and others), capable of more complex communication, jokes or performing simple tasks (find information on the Internet, turn on an electrical appliance), etc. The top of the development of such dialogue systems are complex language universal models (LLM) designed to solve arbitrary tasks like ChatGPT.

A separate subclass of recommender systems is worth mentioning. For them it is important not only to find and output correct information, but also to justify why the system derived a certain solution. We will consider different types of systems from the point of view of approaches to solution retrieval and its justification.

2.1. ChatGPT

A dialogue system (LLM language model) from OpenAI focused on solving arbitrary tasks [13]. It is based on a strong pre-trained InstructGPT language model used to formalize user input, while the model itself is trained using the Reinforcement Learning with Human Feedback (RLHF) approach. With the help of experts, a reward model was created that assigns a score to the correctness of the solution of the underlying model, after which automated reinforcement learning was run.

There are exceptionally few research articles on the ChatGPT architecture available at this time, as OpenAI has not disclosed such information other than what is available on the company's blog [13].

The model has a number of disadvantages.

1. When having a long dialogue, the answers become unclear and the system starts to produce incorrect answers. The reason for this is that the model is not trained on long dialogues, the focus is shifted to more detailed and elaborate answers to a small number of questions in one session.
2. The initial model does not use data from the Internet, but is limited to the data that was fed into it during training. As a result, it cannot use information from dialogues with users (e.g., a new fact about the world around us), which makes the model more dependent on the quality of the training data and creates a time lag between the emergence of new knowledge and its input into the model.
3. The model does not verify the data generated, leading to a paradoxical situation in which the system reasons in great detail about meaningless things, misleading the user (this phenomenon is metaphorically called "hallucinations").

In addition, it is important to note that ChatGPT can store the context of the current dialogue, but this information will not be used in the next session with the same user. As a result, we get that the model can generate answers irrelevant for the user that have been previously received and used in a dialogue with the same user and "forget" what the user communicated to the model earlier.

At the same time, the task of pre-training a neural network of this scale, depending on the results of dialogues, to solve the problem of dynamically updating a set of domain cases requires significant computational resources and can lead to the problem of catastrophic forgetting [14], since the incoming data can be anything. As a result, developers prefer to first type new data, process it, and run a one-time but lengthy learning process.

2.2. BlenderBot

ParlAI's dialogue-oriented BlenderBot model [15]. Due to the presence of long-term memory, the system supports long dialogues better than ChatGPT. It is able to use information previously received from a user, but the data received from one user is not used in a dialogue with other users.

The architecture of the BlenderBot system is based on the pipeline principle [16]. The system generates a response by sequentially using a series of modules, each of which performs a different task, then passes its output to the next module. The model exists in three types, depending on the number of parameters (3, 30, 175 billion parameters).

The order in which the modules [16] are called depends on the context in which the dialogue takes place. The system forms a solution depending on the context of the dialogue by accessing both its own long-term memory and by forming Internet queries. In case memory and web searches are not required, the data will be retrieved from the current dialogue.

The response is generated from this data, and the modules responsible for simulating empathy and personality in the dialogue are also involved at this stage. A complete list of modules and their detailed description can be found in the BlenderBot developers' technical report [16]. By using many modules and adjusting the order in which they are invoked, the system can, while maintaining a long dialogue, update its user data and have a dialogue on several different topics, switching between them, depending on the context of the user's last message.

Considering BlenderBot in the context of our task, we can note the successful implementation in the model of a long-term memory system and a system for deciding whether to search a database of previous dialogues or the Internet. However, data about previous dialogues are stored in memory only as a set of facts (e.g., "User 1 likes dogs", "User 2 lives in country A"), have no semantic connection with each other and are tied to a specific user. Thus, most of the context of previous dialogues is lost. Also, the system cannot reproduce its own steps in solving a particular problem.

Thus, we have considered two popular language models, one of which is intended for solving arbitrary tasks, and the other for maintaining a long and complex dialogue. It should be noted that both of these models do not show the user explicitly how the solution was obtained and do not justify or argue this solution in any way.

We next consider examples of systems designed for narrower applications, but with a more structured approach to constructing argumentative dialogue.

2.3. A system of argumentative dialogue based on argumentative structures

The system presented in [17] has been designed to lead a discussion between the user and the system on various topics. It is a text-based system (although voice interface is also supported), analyses the user's messages, extracts argumentation premises from them [17], and generates arguments based on them. Programmatically, the system is implemented as a set of modules combined using Apache ActiveMQ.

The user's phrase is converted into a "dialogue action". The authors consider four types of dialogue actions: assertion, question, concession and retreat. A logistic regression-based classifier is used to recognize these actions. It then searches for a suitable argumentation node in the argumentation graph based on the cosine similarity between sentence vectors (and the similarity is considered between averaging the value of the node and the user's phrase). The extracted argumentation is processed with respect to the subject domain and the next argument is generated.

Note an important feature of this system: it is able to evaluate the user's actions, in particular, whether he continues his thought, is about to start speaking or is about to finish. This serves as additional information when generating an answer and its justification.

The argument base is populated using the automatic argument extraction techniques developed in [18]. At the time of publication of the paper [17], the system was capable of understanding five discussion topics and supporting up to 2000 argumentation nodes for each of them. However, there is no way to dynamically add new data to the argumentation structure as the dialogue progresses, so a natural question arises: how will the system react to new information that is not in its data.

2.4. Argumentation systems based on communication discourse trees

According to speech structure theory, any coherent discourse can be described by a single discourse tree, described as a tree structure using speech act theory [30]. Each paragraph of text (or the whole text) is converted into a tree through linking sentences using speech acts (e.g., "Justification"), with the leaves of the tree containing the sentences themselves. In this way, an argumentation tree is constructed from which it is possible to determine the presence of argumentation in a paragraph/text. In addition, this approach preserves the context of the argumentation, without which even a human expert would not be able to analyze the presence of argumentation.

It is important to note that in [30] it is the fact of argumentation, not its semantic part and/or persuasiveness, that is considered. Nevertheless, such an approach can be used for semantic analysis of argumentation.

The approach based on communicative discourse trees is also discussed in [31, 32].

2.5. A framework for an argumentative dialogue on the COVID-19 vaccination

The dialogue system [19] is designed to consult the user on the topic of vaccination, with maximum justification of the system recommendations. The system is based on the construction of an argumentation graph, according to the approach of Chalagin and Hunter [20]: finding out the similarity of sentences to get an answer from the knowledge base. The method does not consider the user's previous actions and, as a result, loses the context of the dialogue. The system tries to take into account the user's arguments and construct an answer that does not contradict them and, at the same time, is consistent with the knowledge base.

The reasoning module of the system [20] consists of an argumentation graph compiled by the expert. The nodes of this graph represent either state arguments or response arguments. Associated with each node is a set of natural language sentences representing possible user arguments for that node. The search for a matching node in the graph is performed using a similarity measure of the sentences. The solution is generated on the basis of the information provided by the user, and the node found should agree with the user's data and prevent "dangers," the unacceptable points of the solution indicated by the user (in the example of the article we are talking about counter-indications to vaccination). At the same time, if the system cannot find a "safe" solution, it will still issue a response to the user, but with a request for additional information to adjust the solution. Thus, each new user argument "switches on" the corresponding node in the argumentation graph, and the links coming from this node either reinforce the corresponding solution options or switch them off from the graph.

Thus, well-known universal dialogue systems are good at many tasks, but they do not have mechanisms to explain the progress of solution construction to the user. In addition, their architecture does not allow them to quickly integrate solutions from successful user sessions into their knowledge bases.

On the other hand, specialized systems, in which reasoning is an initial requirement, mainly rely on a pre-prepared knowledge base compiled by an expert and build their solutions and arguments on its basis; going beyond this knowledge base leads to the construction of an unreliable solution. This results in the inability to work with the results of previous sessions, since they lack a mechanism for inserting such information into the knowledge base.

3. Four-level ontological model of the subject domain

As stated above, the aim of this work is to create methods for an intelligent assistant (digital assistant) to conduct a reasoned dialogue with the user [21]. In the framework of our research, the development of an intelligent assistant [22] is based on a semantic model is a four-level ontological model of the subject domain [10, 11]. Let us describe this model in more detail.

The first level of the ontological model is ontologies:

1. Ontology of the subject domain of the goods (devices) under consideration is a set of concepts describing: types of devices; structure and characteristics of devices; functionality of devices.

2. User ontology is a set of concepts describing: goals and intentions; interests, desires, needs; types (classes) of tasks to be solved.

3. Dialogue ontology is a set of concepts describing: argumentation (arguments, counterarguments, etc.); emotional evaluations of users, their satisfaction or dissatisfaction; success of a given dialogue (purchase of goods by the user, continuation of communication and other goods or termination of dialogue by the user, unwillingness to continue it further).

By ontology we mean knowledge only about the meaning of concepts, i.e., analytical statements [23–25] that do not contain information about the state of the real world.

The second level of the ontological model is general (universal) knowledge. These are synthetic statements [23, 26], knowledge about the real world:

1. Subject matter theory is properties of specific goods, their characteristics, functionality, etc.

2. Knowledge about types of users, their classification (by income level, social status, educational level), classes of tasks solved by users, hierarchy of tasks, methods of reducing tasks to subtasks and the possibility of solving the same tasks with different devices.

3. Knowledge of methods of dialogues with users — area methods of argumentation, justification of specific proposals to the user; methods of identifying the goals and needs of users, tasks solved by them; methods of determining the emotional state and emotional assessments of users.

The third level of the ontological model, the most important within the framework of this paper, is the level of domain cases. These are:

1. Product and device cases are specific devices, components, accessories, device sets, price and availability of products in shops, etc.

2. User cases are those users with whom the intelligent assistant has already had dialogues, together with their properties and characteristics; knowledge about the users, their goals, intentions, interests, needs, the tasks they solve.

3. User dialogue cases, hierarchically structured: dialogue with a single user; all dialogues with a given user; dialogues with classes of users.

The fourth level of the ontological model is evaluative and probabilistic knowledge. They are generated by analyzing the domain cases contained in the third level of the ontological model. These include:

- the likelihood that a user with certain characteristics and needs will want to purchase the device;
- the probability that a user who has (bought) device **A** will want to buy device **B**;
- evaluation of similarity of domain cases: devices, device parameters, users and dialogues with them.

Based on this four-level semantic model, we develop a precedence-based approach to construct a reasoned dialogue between an intelligent assistant and a user.

The neural network algorithms that are actively used now, due to their structure, are limited in using recent cases in their model; only when training the next version can this data be included in the training sample.

The use of the case approach solves this problem: we can add new precedents "on the fly," while the system is running. The use of the precedent approach also makes the dialogue system capable of arguing its own conclusions, justifying the choice of goods offered to the user.

The case approach relies on a set of case examples from past user sessions. For certain subject domains, it allows for building solutions from existing data by applying some transformations to it, changing the structure of the case-solution according to the user's task (such transformations, in particular, are realized by means of ontological homomorphisms of partial models, which will be described in detail below).

When implementing the case approach, a number of problems arise. First, precedents should be structured, and not as a relational table with a set of columns. This way of organization will make the search for a suitable precedent weakly related to the semantic content of the precedent. Secondly, it is necessary to evaluate the degree of similarity of precedents, both for finding a suitable "starting point" and for converting a precedent into a final decision. Thirdly, the system should be able to transform precedents according to the user's requirements.

The solution to these problems is to organize precedents into a semantic graph, where the links between precedents will reflect their similarities, showing the degree of similarity in one or another property of the precedent. This solves the problem of semantic search. We need to traverse the precedent graph following the desired semantic links. In this case, the distance of two nodes from each other will explicitly reflect the degree of similarity of the corresponding precedents. The process of precedent transformation can be considered as a transformation of partial models formally describing these precedents by means of ontological homomorphisms, extensions and contractions of partial models. In this case, the properties and parameters of the original precedent will be transformed not necessarily into the same concepts, but into ontologically similar ones. Isomorphic embedding is also possible: expansion of the original precedent, as well as contraction of the precedent, removal of unnecessary elements of the model.

Thus, we implement a software system that performs solution search based on semantic similarity of precedents with an explainable solution search mechanism.

4. A theoretical and modelling approach to the design of an intelligent assistant

It is important to note that most of the precedents we consider, both product precedents and user precedents, contain only a part of all the information about the user or device. Therefore, within the model-theoretic approach, precedents should be formally described by partial models rather than by ordinary models (algebraic systems).

Definition. Consider a signature $\sigma = \langle P_1, ..., P_m, c_1, ..., c_l \rangle$ in which $P_1, ..., P_m$ are predicate symbols and $c_1, ..., c_l$ are symbols of constants. Consider a tuple $\mathfrak{A}^p = \langle A, P_1, ..., P_m, c_1, ..., c_l \rangle$ and let for each $n \leq m$ value *n-ary* predicate P_i on $\mathfrak{A}^p$ is defined as a pair $P_i^{\mathfrak{A}^p} = (P_i^+, P_i^-)$ where $P_i^+, P_i^- \subseteq \left|\mathfrak{A}^p\right|^n$ and $P_i^+ \cap P_i^- = \varnothing$. Let us call $\mathfrak{A}^p$ a partial model in the signature σ. Let us assume that for elements $a_1, ..., a_n \in |\mathfrak{A}^p|$ if $(a_1, ..., a_n) \in P_i^+$, then it is fulfilled $\mathfrak{A}^p \vDash P_i(a_1, ..., a_n)$, if $(a_1, ..., a_n) \in P_i^-$, then it is fulfilled $\mathfrak{A}^p \vDash \neg P_i(a_1, ..., a_n)$, and if $(a_1, ..., a_n) \notin (P_i^+ \cup P_i^-)$, then the value of the predicate $P_i(a_1, ..., a_n)$ on the partial model $\mathfrak{A}^p$ is undefined.

The class of partial signature models σ denote by $K^p(\sigma)$.

We use ontological homomorphisms to transform partial models that formalize precedents. In this paper, we consider three types of ontological homomorphisms that are most important for this presentation; to illustrate, we take the example of a device like a laptop. These are ***generalization*** homomorphisms (in the partial model, the presence of a *USB A* connector in the laptop is replaced by just the presence of a *USB* connector), ***refinement homomorphisms*** (the presence of a *USB connector* in the laptop is replaced by the presence of *USB A*) and ***similarity*** homomorphisms (the presence of a *USB A connector is* replaced by the presence of a *USB C connector*).

Let us give strict definitions of ontological homomorphisms. For this purpose, let us consider ontological relations: on the set of key concepts of the ontology of the subject domain of the goods (devices) under consideration is signature predicates σ. We introduce two two-place relations: the general-private relation ***Hyp***(Q, P) and similarity relation ***Sim***(P, Q). The relation ***Hyp*** is a partial order, and the relation ***Sim*** is reflexive and symmetric (but not necessarily transitive).

Definition. Consider partial models $\mathfrak{A}^p \in K^p(\sigma_1)$ and $\mathfrak{B}^p \in K^p(\sigma_2)$ let $P^n \in \sigma_1$, $Q^n \in \sigma_2$, $\sigma_1 \setminus \{P^n\} \subseteq \sigma_2$ and it is fulfilled ***Sim***(P, Q). Mapping $h{:}|\mathfrak{A}^p| \to |\mathfrak{B}^p|$ let us call the ontological homomorphism of the **similarity of the** partial model $\mathfrak{A}^p$ into a partial model $\mathfrak{B}^p$ ($h{:}\,\mathfrak{A}^p \to \mathfrak{B}^p$) if for any $c \in \sigma_1$ and $a_1, \ldots, a_n \in |\mathfrak{A}^p|$ is fulfilled:

(a) if $\mathfrak{A}^p \vDash P(a_1, \ldots, a_n)$ then $\mathfrak{B}^p \vDash Q(h(a_1), \ldots, h(a_n))$;

(b) if $\mathfrak{A}^p \vDash \neg P(a_1, \ldots, a_n)$ then $\mathfrak{B}^p \vDash \neg Q(h(a_1), \ldots, h(a_n))$;

(c) $h(c^{\mathfrak{A}^p}) = c^{\mathfrak{B}^p}$.

The truth and falsity of the other predicates from the σ_1 are preserved.

Definition. Consider partial models $\mathfrak{A}^p \in K^p(\sigma_1)$ and $\mathfrak{B}^p \in K^p(\sigma_2)$, $\sigma_1 \setminus \{P^n\} \subseteq \sigma_2$ let $P^n \in \sigma_1$, $Q^n \in \sigma_2$, $\sigma_1 \setminus \{P^n\} \subseteq \sigma_2$ and it is fulfilled ***Hyp***(Q, P). Mapping $h{:}|\mathfrak{A}^p| \to |\mathfrak{B}^p|$ let us call the ontological homomorphism of the ***generalization of the*** partial model $\mathfrak{A}^p$ into a partial model $\mathfrak{B}^p$($h{:}\,\mathfrak{A}^p \to \mathfrak{B}^p$) if for any $c \in \sigma_1$ and $a_1, \ldots, a_n \in |\mathfrak{A}^p|$ is fulfilled:

(a) if $\mathfrak{A}^p \vDash P(a_1, \ldots, a_n)$ then $\mathfrak{B}^p \vDash Q(h(a_1), \ldots, h(a_n))$;

(b) $h(c^{\mathfrak{A}^p}) = c^{\mathfrak{B}^p}$.

The truth and falsity of the other predicates from the σ_1 are preserved.

Definition. Consider partial models $\mathfrak{A}^p \in K^p(\sigma_1)$ and $\mathfrak{B}^p \in K^p(\sigma_2)$ let ${}^n \in \sigma_1$, $Q^n \in \sigma_2$, $\sigma_1 \setminus \{P^n\} \subseteq \sigma_2$ and it is fulfilled ***Hyp***(Q, P). Mapping $h{:}|\mathfrak{A}^p| \to |\mathfrak{B}^p|$ let us call the ontological homomorphism of the ***refinement of the*** partial model $\mathfrak{A}^p$ into the partial model $\mathfrak{B}^p$ ($h{:}\,\mathfrak{A}^p \to \mathfrak{B}^p$) if for any $c \in \sigma_1$ and $a_1, \ldots, a_n \in |\mathfrak{A}^p|$ is fulfilled:

(a) if $\mathfrak{A}^p \vDash \neg P(a_1, \ldots, a_n)$, then $\mathfrak{B}^p \vDash \neg Q(h(a_1), \ldots, h(a_n))$;

(b) $h(c^{\mathfrak{A}^p}) = c^{\mathfrak{B}^p}$.

The truth and falsity of the other predicates from the σ_1 are preserved.

The single or multiple use of ontological homomorphisms allows the intelligent assistant to automatically switch from descriptions of some devices to descriptions of other devices that are similar to a certain extent. For example, a user wants to buy a certain device with certain characteristics, but the required device is not available (or its price does not suit the user). Then the intelligent assistant automatically finds another device, whose partial model is ontologically homomorphic to the model of the original one, but which is available for sale, and offers this device to the user. The product (or several products) closest to the required one is automatically searched.

When suggesting devices to the user, the intelligent assistant also provides an explanation of why their difference from the user's desired one is not essential from the point of view of the tasks to be solved by the user. Such natural language explanations are either pre-defined in the semantic model, when determining the ontological similarity of concepts, or extracted from natural language texts in the process of dialogue (in particular, from product descriptions on the websites of manufacturers and shops, from customer reviews, etc.). [27, 28]).

This process continues iteratively until a device satisfactory to the user is found: the user indicates what he does not like, and the intelligent assistant selects a new variant. In this way, a reasoning graph is constructed whose vertices contain partial models corresponding to the devices, and transitions are made using ontological homomorphisms.

Here it is important to note that by precedents we mean all kinds of objects, subjects and situations with which the intelligent assistant works. The precedents are both objects of the subject domain, users themselves, and dialogues with users, the results of the previous session of the software system: artifacts of interaction between the intelligent assistant and the user. As an example, let us consider a hierarchy of structured precedents: objects of the subject domain related to the satisfaction of user needs:

1. A subject matter object, a commodity that a user needs (e.g., a computer or smartphone).
2. Subject matter object + user needs (as identified by the intelligent assistant in the dialogue process).
3. Subject matter object + user needs + class of tasks to be solved by the user. We extend the precedent by adding the tasks that the user needs to solve. It is important to note that the properties of the subject matter object (e.g., the functionality of the device) are clearly defined in the knowledge base and are independent of the user's perception. On the other hand, the tasks that the user intends to solve with a given device depend on the user's ultimate goals, desires and needs. The class of tasks to be solved is determined by the user. Having data about the user's needs and the tasks to be solved by the user, we construct a precedent as a triple: <partial model describing the device; user's needs; set (class) of tasks to be solved by the user>.

This way of representing precedents facilitates combining objective information about goods with subjective information about the user obtained by the intelligent assistant in the process of dialogue. Recall that precedents are represented at the third level of the ontological model of the subject domain.

The construction of a solution is the selection of goods required by the user and involves the comparison of both objects of the subject domain (goods, devices) and structured precedents described above. For this purpose, the apparatus of metrics is used. This will allow comparison of objects and precedents during the work of the intelligent assistant. In particular, a semantic graph of precedents with a pre-calculated (or set by an expert) measure of similarity (likeness) of precedents is used. Knowledge about similarity measures of precedents belongs to the fourth level of the ontological model of the subject domain. On the basis of this knowledge, in particular, the similarity relation is specified The similarity relation discussed above is in the definition of ontological homomorphisms.

When selecting products, user priorities are calculated based on two parameters: firstly, properties and functionality of devices and, secondly, user's needs and desires, class of tasks to be solved, that is, objective and subjective parameters. Focusing on these two types of parameters, we calculate the similarity of different precedents, including the objects of a given subject domain.

5. Software implementation of the dialogue system

The developed software system [22] is a set of five blocks (modules) that provide various stages of the system operation. The technical implementation is a MVC application on Java Spring, with REST interface.

Block 1 is responsible for performing user input and formalizing it through a speech action search mechanism [29], correcting the user model and detecting user intentions.

Block 2 is responsible for analyzing the input received, and generating system messages to request further information from the user.

Block 3 is responsible for searching for the required product, a precedent of the subject domain. It checks whether the partial model of the required product, built as a result of the dialogue with the user, is a submodel of the model of some precedent of the subject domain (i.e., the product available). The input of the block is the user's needs formulated and checked for consistency. Then, with the help of ontological homomorphisms realized on the basis of the similarity function of two partial models, the precedent most similar to the desired user is searched for.

If a precedent is found, this solution will be proposed to the user; the natural language description of the found precedent will be used as a justification (argument).

Block 4 is responsible for analyzing the user's reaction. The main function of the block is to clarify the user's requirements. If there are new data in the user's response, they are formalized using the mechanisms of blocks 1 and 2 and block 3 is started again. Thus, the solution search process is iterative.

Block 5 is responsible for the final generation of the decision and its justification (reasoning). The user is offered a product that fully meets his requirements identified in the dialogue process. A set of goods that meet the requirements but differ in price or characteristics that are not important for the user can also be presented.

Conclusion

This article develops methods of creating intelligent assistants. Intelligent assistants can be used to help users choose products as recommendation systems in call centers to solve various customer problems, to solve technical support tasks, to help people with disabilities. In this paper, first of all, we consider intelligent assistants designed to help the user to select goods.

To create intelligent assistants, we develop methods for reasoned dialogue with the user. For this purpose, we develop methods for automated construction of reasoning and argumentation. The formalization of reasoning and argumentation is done using partial models, homomorphisms and ontological homomorphisms of partial models. Ontological homomorphisms of the similarity of partial models formally describe the similarity of precedents, which serves as a mathematical basis for the construction of reasoning based on precedents.

The proposed architecture of the software system implements methods of dialogue with the use of precedents, replenishment of the database of precedents after each session of work, their organization in the form of a semantic network. Such an approach allows us to achieve transparency of the system operation, to increase flexibility of solution selection due to the analysis of semantic content of phrases entered by the user (with the help of atomic models), all of which distinguishes the system from the existing analogues.

Further development of the system is possible in the direction of improving the algorithm of precedent search in the semantic network, with the introduction of more links between precedents to increase the detail of the search, as well as the development of methods for determining the similarity and likeness of precedents. ■

Acknowledgements

The work was carried out within the framework of the state assignment of the S.L. Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences. S.L. Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, project FWNF-2022-0011.

References

1. Toulmin S.E. (2003) *The uses of argument*. Cambridge University Press.

2. Reed C., Rowe G. (2004) Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, vol. 13, no. 4, pp. 961–979.

3. Walton D., Reed C., Macagno F. (2008) *Argumentation schemes*. Cambridge University Press.

4. Macagno F., Walton D. (2015) Classifying the patterns of natural arguments. *Philosophy & Rhetoric*, vol. 48, no. 1, pp. 26–53. https://doi.org/10.5325/philrhet.48.1.0026

5. Wagemans J. (2016) Constructing a periodic table of arguments. *Argumentation, objectivity, and bias: Proceedings of the 11th international conference of the Ontario Society for the Study of Argumentation, Windsor, Ontario, Canada*, pp. 1–12. https://doi.org/10.2139/ssrn.2769833

6. Walton D., Gordon T.F. (2017) Argument invention with the Carneades argumentation system. *SCRIPTed*, vol. 14, no. 2, pp. 168–207. https://doi.org/10.2966/scrip.140217.168

7. Wachsmuth H., Potthast M., Al-Khatib K., Ajjour Y. (2017) Building an argument search engine for the Web. Proceedings of the *4th Workshop on Argument Mining*, pp. 49–59.

8. Slonim N. et al. (2021) An autonomous debating system. *Nature*, vol. 591, no. 7850, pp. 379–384. https://doi.org/10.1038/s41586-021-03215-w

9. Stab C., Miller T., Shiller B., Rai P., Gurevych I. (2018) Cross-topic argument mining from heterogeneous sources. Proceedings of the *2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3664–3674.

10. Naydanov Ch., Palchunov D., Sazonova P. (2015) Development of automated methods for the critical condition risk prevention, based on the analysis of the knowledge obtained from patient medical records. Proceedings of the *International Conference on Biomedical Engineering and Computational Technologies, Novosibirsk, 28–30 October 2015*, pp. 33–38. https://doi.org/10.1109/SIBIRCON.2015.7361845

11. Naydanov Ch., Palchunov D., Sazonova P. (2016) Development of automated methods for the prevention of risks of critical conditions, based on the analysis of the knowledge extracted from the medical histories. *The Siberian Scientific Medical Journal*, vol. 36, no. 1, pp. 105–113.

12. Naidanov Ch.A. (2019) Development of ontological model kernel, customisable for the subject domain. *Vestnik of Novosibirsk State University. Series: Information technologies*, vol. 17, no. 1, pp. 72–81.

13. OpenAI (2022) *Introducing ChatGPT*. Available at: https://openai.com/blog/chatgpt (accessed 26 January 2024).

14. McCloskey M., Cohen N.J. (1989) Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation*, vol. 24, pp. 109–165.

15. Blenderbot AI (2022) *BlenderBot 3: A Conversational AI Prototype*. Available at: https://blenderbot.ai (accessed 26 January 2024).

16. Shuster K. et al. (2022) Blenderbot 3: A deployed conversational agent that continually learns to responsibly engage. *arXiv:2208.03188*. https://doi.org/10.48550/arXiv.2208.03188

17. Higashinaka R. et al. (2017) Argumentative dialogue system based on argumentation structures. Proceedings of the *21st Workshop on the Semantics and Pragmatics of Dialogue*, pp. 154–155.

18. Lippi M., Torroni P. (2016) Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology*, vol. 16, no. 2, pp. 1–25. https://doi.org/10.1145/2850417

19. Fazzinga B., Galassi A., Torroni P. (2021) An argumentative dialogue system for COVID-19 vaccine information. *International Conference on Logic and Argumentation*. Cham: Springer International Publishing, pp. 477–485.

20. Chalaguine L.A., Hunter A. (2020) A persuasive chatbot using a crowd-sourced argument graph and concerns. *IOS Press E-books Series Frontiers in Artificial Intelligence and Applications*, vol. 326: Computational Models of Argument, pp. 9–20. https://doi.org/10.3233/FAIA200487

21. Palchunov D., Yakobson A. (2022) Automated methods for conducting an argumented dialogue with the user. *2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, pp. 880–885. https://doi.org/10.1109/SIBIRCON56155.2022.10016975

22. Palchunov D.E., Yakobson A.A. (2022) *Software system "ArgumentDialogue"*. Certificate of state registration of computer programme No. 2022664041, registered 22.07.2022.

23. Carnap R. (1956) *Meaning and necessity. A study in semantics and modal logic*. Chicago.

24. Palchunov D.E. (2006) Modelling of thinking and formalization of reflexion I: Theoretical-model formalization of ontology and reflexion. *Philosophy of Science*, vol. 4(31), pp. 86–114.

25. Palchunov D.E. (2022) Methodological aspects of the application of model theory in knowledge engineering and artificial intelligence. *2022 Ural-Siberian Conference on Computational Technologies in Cognitive Science, Genomics and Biomedicine (CSGB)*, pp. 210–215. https://doi.org/10.1109/CSGB56354.2022.9865602

26. Palchunov D.E. (2008) Modelling of thinking and formalization of reflexion. Part 2: Ontologies and formalisation of concepts. *Philosophy of Science*, vol. 2(37), pp. 62–99.

27. Palchunov D.E. (2008) Solution of the information search task on the basis of ontologies. *Business-informatics*, vol. 1(3), pp. 3–13 (in Russian).

28. Orlovsky A., Palchunov D. (2021) Development of automated methods for the domain ontology population with the help of a virtual assistant. *2021 IEEE 22nd International Conference of Young Professionals in Electron Devices and Materials (EDM)*, pp. 537–541. https://doi.org/10.1109/EDM52169.2021.9507641

29. Makhina E.D., Palchunov D.E. (2018) Software system to determine speech actions in natural language texts. *Vestnik of Novosibirsk State University. Series: Information technologies*, vol. 16, no. 4, pp. 95–106. https://doi.org/10.25205/1818-7900-2018-16-4-95-106

30. Galitsky B., Ilvovsky D., Kuznetsov S. (2018) Detecting logical argumentation in text via communicative discourse tree. *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 30, no. 5, pp. 637–663. https://doi.org/10.1080/0952813X.2018.1467492

31. Galitsky B., Ilvovsky D., Kuznetsov S.O. (2015) Text integrity assessment: Sentiment profile vs rhetoric structure. *Computational Linguistics and Intelligent Text Processing* (eds. A. Gelbukh). CICLing 2015. Lecture Notes in Computer Science, vol. 9042. Springer, Cham, pp. 126–139. https://doi.org/10.1007/978-3-319-18117-2_10

32. Galitsky B., Ilvovsky D., Kuznetsov S.O. (2015) Rhetoric map of an answer to compound queries. Proceedings of the *53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China*, vol. 2: Short Papers, pp. 681–686. https://doi.org/10.3115/v1/P15-2112

About the authors

Dmitry E. Palchunov

Dr. Sci. (Phys.-Math.), Academician of the Russian Academy of Engineering;

Leading Researcher, S.L. Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, 4, Akademika Koptyuga Ave., Novosibirsk 630090, Russia;

E-mail: palch@math.nsc.ru

ORCID: 0000-0001-9487-3256

Alexander A. Yakobson

Assistant, Department of General Informatics, Novosibirsk National Research State University, 1, Pirogova St., Novosibirsk 630090, Russia;

E-mail: a.yakobson@g.nsu.ru

DOI: 10.17323/2587-814X.2024.1.22.35

Improving target budgeting in a corporate performance management system

Maxim E. Oshchepkov
E-mail: moshchepkov@hse.ru

Graduate School of Business, HSE University
Address: 26–28, Shabolovka Str., Moscow 119049, Russia

Abstract

Currently the portrayal of the procedure for developing management actions during the planning process in the scientific and professional community does not align with the practice of systematic and consistent plan creation supported by an informational analytical system. Alternatively, the non-formalized decision-making activity of the planner, which involves a situational expert approach, becomes a dependency in the planning process. This study is aimed at developing an analytical approach for implementing the plan reconciliation procedure in the process of corporate performance planning. This will increase the utilization of capabilities of the corporate performance management system and formalize the task of generating managerial actions by adjusting targeted budgeting values using mathematical methods. For this purpose, the standard planning process is enhanced by analytical support units, including the algorithm of inverse calculations of individual key performance indicators (KPI) and an advanced module for scenario modeling. The improved model of target budgeting process presented here delivers automated formation of management actions of the budgeting department and subdivision management, guided towards accomplishing strategic goals. The application of inverse calculations provides a mathematical formulation of the task of calculating indicators of planned key values, and the Sense and Respond (SaR) system allows you to supplement the mathematical formulation with weighting coefficients of key performance indicators calculated algorithmically, relying on the manager's decisions rather than expert evaluation. The implementation of the approach we developed will improve the quality of planning by the highest priority criteria of operability, accuracy and adaptability due to the consistency and methodology of budgeting with the use of modern information technology.

Keywords: corporate performance management, budgeting, information model, inverse calculations, mathematical programming

Citation: Oshchepkov M.E. (2024) Improving target budgeting in a corporate performance management system. *Business Informatics*, vol. 18, no. 1, pp. 22–35. DOI: 10.17323/2587-814X.2024.1.22.35

Introduction

The present-day economy is distinguished by shifts in economic and social models triggered by various crisis phenomena. This has led to greater emphasis on performance measurement, not only throughout the entire economy but also within individual companies and their structural units [1]. This factor stimulated the active development and gradual implementation of corporate performance measurement and management (CPM) systems, which are grounded in a systematic approach and perspective to ensure veracity in the measurement of economic phenomena.

Along the course of CPM systems' progression and utilization various challenges are encountered which hinder achievement of sustainable development and a competitive edge. In modern realities, a company's competitiveness is reliant on the high quality of its economic activity planning, which is assessed using various criteria, primarily accuracy, operability and adaptability. Achieving high scores in these criteria depends on the tools development (systems and rules) and management technology (models and methods). Management methods at the corporate level are crucial to the success of an organization, since they determine the organization's strategy and provide for its implementation [2].

The implementation and creation of CPM systems by companies are proving to be significantly challenging in terms of building and developing effective management tools and technology in the process of target realization. Development of management actions is generally a process of non-formalized activity of the manager based on an expert approach. At the same time, the methods and procedures employed by companies do not facilitate the systematic and methodical establishment of planned key indicators corresponding to the requirements of a unified strategic goal for all of the company's structural units. One reason for these issues is insufficient comprehension that translating corporate strategy into budget planning requires a distinct mathematical foundation for the company's performance management. The issue at hand has been pertinent in recent decades, as evidenced in studies by both national and international authors: Bourne, Maryška, Doucek, Zhang, Mari, Kitova, Bruskin, Odintsov, Nikishova [1, 3–9].

This paper presents an analytical approach to improving the standard budgeting process by designing and executing the inverse calculation algorithm. This algorithm facilitates the formalization and automation of the procedure of producing management actions on partially planned indicators. Additionally, the "Actual–Forecast" module broadens the scenario modeling capabilities. Therefore, the proposed approach streamlines the budgeting process and increases the organization's overall efficiency.

The proposed algorithm of inverse calculations performs the calculation of planned key indicators considering their priority on the basis of a manager's decisions. The algorithm utilizes both the theory of inverse calculations and the SaR methodology. The method of inverse calculations provides the mathematical formulation for calculating planned key value indicators, and the SaR system supplements it with weight coefficients of key primary performance indicators. These coefficients are calculated algorithmically based on manager's decisions, rather than expert evaluation.

The scenario modeling module allows for planning and economic management in parallel with the main budgeting process to develop management actions based on actual data from financial responsibility centers (FRC) and planned key indicators calculated using the algorithm of inverse calculations and SaR, forming budgeting scenarios as a result.

The advantages of the proposed approach are in improving the quality of the planning process across all critical quality criteria, including operability, accuracy and adaptability. Operational efficiency is achieved by reducing the time and labor costs of the planning and economic department (PED) to execute the procedure of managerial actions development. Moreover, the number of adjustments of planned budget calculations by the company's subdivisions is reduced. Increased planning accuracy is provided by methodical and systematic realization of the process, which assumes interaction of the developed analytical blocks, contributing to effective coordination of strategic goals and actions for their achievement. Adaptability of planning is increased by reducing the time for data adjustments for new planning steps and increasing the speed of system adaptation to observed changes in the external environment, since the algorithm includes the condition of approximating the behavior of the decision maker.

1. The concept of corporate performance management

The concept of corporate performance management comprises information technologies and tools, management methods and processes, as well as human resources, and implies periodic measurement and analysis of key indicators to achieve specific goals [6]. The performance measurement system, according to the authors of the fundamental studies in this area, is a set of indicators used to quantitatively assess both internal (efficiency) and external performance (effectiveness) [10, 11].

The literature on corporate performance measurement describes a wide range of approaches to designing performance measurement systems that are categorized based on various factors. According to Bourne [3], two dimensions can be used: the underlying measurement procedure (needs indicators [12, 13], audit indicators [14], model indicators [15]) and the underlying approach in the perspective of the process manager's role (consultant's guide [16], facilitator's guide [11]).

Based on the concept of target management, key performance indicators can be categorized into three types [8]. Primary indicators designed to measure the costs and resource requirements necessary to achieve the set goals – they are stored in the accounting database. Integrated indicators intended to assess the efficiency of individual areas of the company's activities – they are calculated indicators. Complex indicators meant to measure the performance of the whole company and the level of achievement of strategic goals – they are derived indicators.

Target management comprises three levels: strategic, tactical, and operational. The strategic level corresponds to strategic planning carried out by managers, which determines the long-term (more than 3 years) direction of company development. This paper specifically focuses on the planning and economic department (or budgeting department), responsible for target budgeting. Target budgeting involves creating a budget aimed at achieving operational planning goals, which are defined by the values of key performance indicators from the company's strategic map. The budget serves as a tactical plan with a one-year horizon that reflects the outcomes of operational, investment and other activities carried out by the company [17].

Business performance management technology is a sequence with a closed cycle that includes at least four basic blocks, namely modeling, planning, monitoring, and analysis [18].

The modeling block operates outside of the general management cycle, since it is executed incrementally, contingent upon the pace of the organization's advancement and level of adaptation to the external environment. This block includes formulating a strategic map, building goal trees, setting key indicators and modeling business processes.

The planning block is enlarged and in the interpretations of various authors is often divided into several blocks [18]. Thus, according to Odintsov [8], the two stages that follow modeling are forecasting of key performance indicators and calculation of planned values of key performance indicators.

The planning block for key performance indicators is the most vital element of business performance management technology because it systematically provides target values of key performance indicators to the PED and other structural units for coordination of forecasts and formation of the final approved plan for the future period. This block encompasses actual–forecast analysis and calculation of forecast values of indicators during a rolling reporting period. Achievement of the operational goal specified by management for the subsequent period depends on the values of indicators obtained at this stage.

The remaining two components of the business performance management concept, namely monitoring and analysis, involve operational control of the company's financial and economic condition and comprehensive diagnostics of its indicators, correspondingly.

2. Research problem

As a basis for forming assumptions to improve the methodological basis of corporate performance management, we considered a typical solution of the closed-cycle corporate performance management system, which is common in many domestic and international companies. Its example demonstrates the bottlenecks of the algorithm, the solution of which is the focus of this study (*Fig. 1*).

The diagram illustrates the stages of the typical process of financial planning and forecasting the compa-

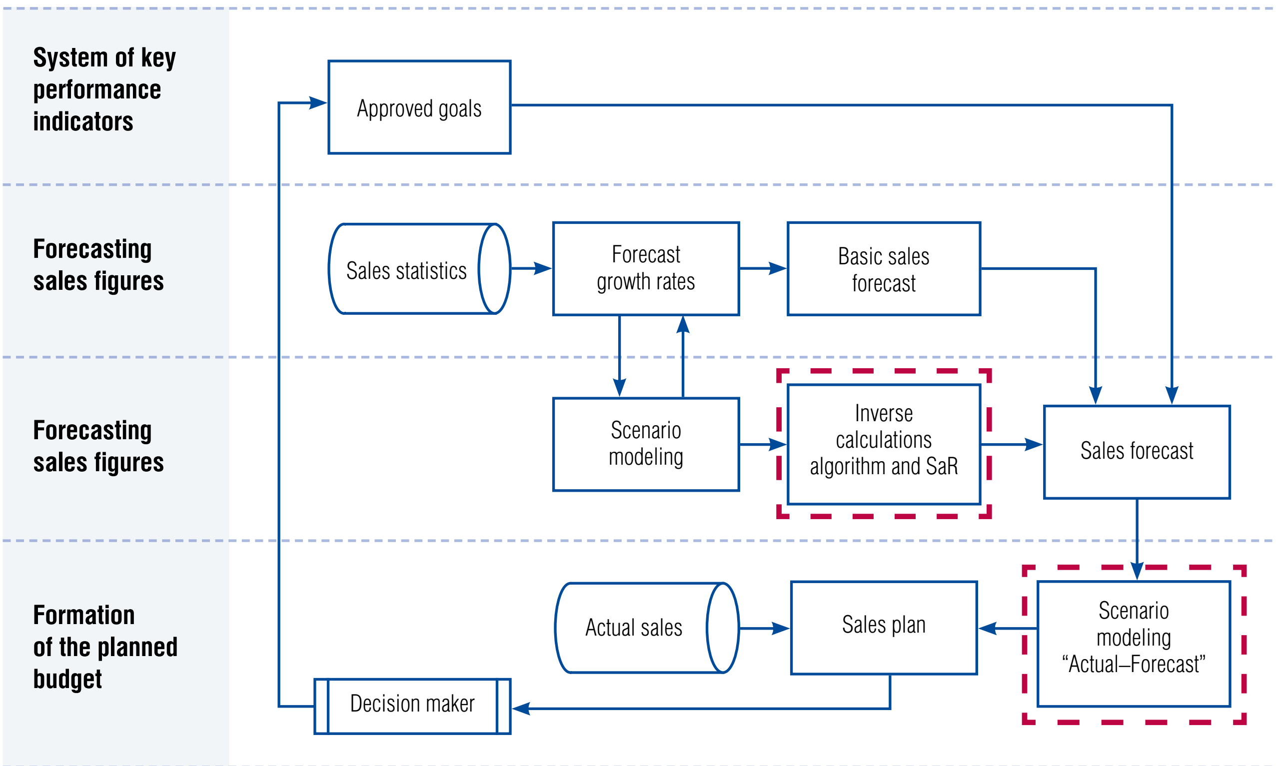

Fig. 1. Scheme of improvement of a typical planning process.

ny's sales performance. It incorporates blocks of analytical approach to determining forecast adjustments and harmonization of planning budgets (highlighted by dashed lines in the scheme).

In the depiction of the typical process of financial planning and forecasting of the company's performance presented here, several obstacles were observed. These "bottlenecks" curtail the potential systemic effects of impact on the process quality when using information system capabilities and analytical methods. Among all possible limitations of the model examined, the following have been identified as the target ones for this research:

1) use of expert assessments both at the stage of planning formation of indicators and at the stage of coordination and formation of the approved planning budget;
2) from the first point follows the lack of mathematical formalization of decision-making tasks regarding management actions on target indicators by adjusting partial indicators;
3) lack of flexibility in terms of analysis of forecasts of responsible persons at the stage of the approved planning budget formation, since scenario analysis does not provide sufficient opportunities for aggregation of various estimates, visions and forecasts for different sales directions, commodity nomenclatures.

The literature indicates that improving the quality of a typical process as a whole should be accomplished through information analytics support in each process zone, ensuring the orchestration of evolving information analytics methods while harmonizing corporate performance management with business process management [2, 19–21].

Systemic quality improvement in each zone of the process is achieved through various factors. According to Bruskin [7], the combined use of trend, information and scenario methods of predictive modeling can improve the quality of the planning process. In the planning zone – through the preparation of sales statistics based on actuals and recalculation of planned indicators by the incremental method, considering the adjustment of the rolling trend by expert evaluation. In the forecasting zone – through scenario analysis by using OLAP tools. The paper by Kitova [6] presents an attempt to overcome subjectivity and formalize the procedure of selecting multiple options for the company's strategic development according to various factors (KPIs, strategic initiatives, investment allocation) by developing a target planning model and an algorithm for selecting the optimal solution. The algorithm for selecting alternatives involves analyzing the shareholder value of the company and comparative assessment using the method of hierarchy analysis. In her work [9], Nikishova describes the use of artificial intelligence technologies to overcome information asymmetry in decision-making by the board of directors. Although the results are aimed at the strategic level of management, their study is useful for the implementation of decisions at the tactical level, because the author offers an approach of analytical support of the process of developing management actions performed in parallel with the main process, not replacing the decisive role of the board of directors. The applicability of the concept of inverse calculations in the task of corporate performance management was considered by Odintsov [8]. At the same time, the determination of the weights of partial indicators forming the integral key indicator in the author's work is carried out by means of expert methods.

The papers of some international authors are directed at improving the quality of the planning process, primarily its operability, by developing complex models and systems that combine business and technical aspects of the process. Thus, Maryška and Doucek [1] describe the REMONA model, the advantages of which are a reference model of CPM and BI design, definition of methodologies and models accompanying the system, definition of a system of indicators, measurements, implementation of the solution in the company. Prilianti and Hikmat [22] developed a platform called the Integrated performance management system (IPMS), which can translate the company's vision to a lower level of management, improving the company's performance by identifying the relationships between variables, evaluation methods and improvement priorities.

In addition, examples of analytical support for the planning process using artificial intelligence have been presented in international literature. Zhang [4] in his paper applied principal component analysis and artificial neural network analysis to build the RMT-BP model for predicting the performance indicators of companies. Various financial and non-financial indicators characterizing the performance of companies are used as input parameters, and the corresponding weights of indicators are determined by the hierarchy analysis method. Mari et al. [5] show how key indicators can be efficiently determined using linear constraints, and then how managerial actions can be computed using mathematical programming methods.

Despite the significant advancement in research on CPM technologies and their supporting techniques, the literature still has open niches to explore and improve the planning process. The cases of artificial intelligence used in the planning process discussed in the literature are based on a "black box" model, which limits their applicability to business problems that require transparency, clarity and controllability by management decision makers. The usefulness of machine learning depends on many factors that have not been sufficiently studied yet, such as the type of problem, the stage of decision-making and the technology used [23]. At the same time, the fragmented consideration of different methods contradicts the authors' theses of exploratory and review papers about the necessity to harmonize methods in their interaction and with the peculiarities of the management process. As a result of analyzing scientific studies and the practical experience of companies, it can be concluded that scientists face the problem of creating effective approaches to the practical application of modern analytical methods to improve the quality of managerial decision-making in business.

The procedure of coordinating plans to match the company's strategy with the existing opportunities, considering internal and external constraints, is a labor-intensive but necessary task. This procedure is performed through step-by-step adjustment of target values of indicators. At each stage of adjustment, the current values of strategic indicators are reviewed and analyzed by the company's management. Following the results of the analysis, management determines operational goals, which are a compromise between the actual situation of the business unit and the company's strategic plan. Operational goals serve as a starting point for calculating planned key performance indicators for future periods. It is worth noting that this calculation is performed with consideration of the constraints set by the various business units involved in the budgeting process. For example, the forecasting department sets a target profitability limit based on the forecast, while the sales department sets a different limit for the indicator, given the volume of the company's resources. Therefore, the sales budget is supplemented with the elements: target value, forecast value, minimum and maximum values [24]. These elements are used as constraints in the inverse calculations of the values of planned key indicators.

Given the exceptional importance of the planning stage, the tasks related to the development of the theoretical and methodological basis for the procedure of plan coordination require high attention of researchers and practitioners. That is why this study is aimed at developing an analytical approach to the realization of the procedure of plan coordination in the process of corporate planning of indicators which will allow one to expand the use of CPM-system capabilities and provide mathematical formalization of the task of developing managerial actions when adjusting the values of planned key indicators.

Following the results of the previous authors, this paper proposes to supplement the typical planning process with the blocks "Inverse calculations algorithm and SaR," all of which allows one to formalize and automate the procedure of generating managerial actions on partial planning indicators, and "Scenario modeling Actual–Forecast," which provides extended capabilities regarding scenario modeling, multivariate analysis and modeling of PED plans in parallel with the main budgeting model.

The advantages of the proposed approach consist in improving the quality of the planning process by all priority quality criteria: operability, accuracy, adaptability. Operational efficiency is achieved by reducing the time and labor costs of the PED to execute the procedure of generating management actions, since the manager in this case has support in the form of a calculation algorithm, advanced tools of computer modeling of plans. Furthermore, the number of subsequent adjustments of planned values by the company's departments to the budget is reduced. Improving the accuracy of planning is achieved by methodical and systematic realization of the process, which implies interaction of the developed analytical blocks contributing to the effective coordination of strategic goals and actions to achieve them. The adaptability of planning is increased by reducing the time for data adjustments for new planning steps and increasing the speed of adaptation of the system to the observed changes in the external environment, since the algorithm includes the condition of approximating the behavior of the decision maker.

3. Budgeting process model

3.1. Development of the target management toolkit

Using the methods and tools of business process modeling, diagrams reflecting the essence of the proposed approach have been constructed. The diagrams below describe the process of so-called goal realization, which involves the development of a toolkit (system of goals and indicators, procedures for forming prescriptions, rules for adjusting indicators) and technology (chain of procedures and operations) of target management.

The first diagram (*Fig. 2*) shows the key changes in the proposed modification of the budgeting process model relative to the typical solution, which consist in the application of SaR methodology and the theory of inverse calculations to formalize and automate the procedure of managerial decision-making in the budgeting process.

The stage of the process preceding the algorithm of calculations in the model is the construction of hierarchical structures of goals. Work on the formation of the "goals-indicators" system is performed on the side of business analysts. Based on planning practice, the construction of goal structures covering various aspects of the company's activities and subsequent reproduction of their indicators require the use of more general structures (network structures) that do not have the property of strict node hierarchy. Nevertheless, such structures are of limited use for formal processing, since they generate significant difficulties and ambiguity in the interpretation of the results. Therefore, in the absence of explicit restrictions, the network structures of "goals-indicators" are transformed into hierarchical structures. The procedure of formal processing of the hierarchical structure is supported by the software tools of the CPM-system, where indicator measurements are presented in the form of hierarchies.

The first step in the algorithm of formal processing of indicators is the scaling of measurement results. Differences in the measurement scales of individual partial indicators that determine the level of target indicators' achievability impose restrictions on their joint processing. Thus, the model solves the problem of converting individual indicators to a single measurement scale and forming a general integral indicator.

The scaling procedure includes preliminary normalization of initial data and calculation of individual weights of partial indicators. Initial data are transferred from the database of the information system, and the results of calculations are stored in a separate data warehouse linked to the SaR system.

3.2. The mathematical background of inverse calculations and SaR

At the initial stage of calculations, the values of jointly processed partial indicators are normalized to a unified measurement scale to obtain the target dimensionless integral indicator. The range from 0 to 1 inclusively is taken as the normalization interval to simplify data processing. For calculation of the integral indicator, an additive form of convolution is used, which is the sum of weighted values of unidirectional partial indicators. Normalization of the values of partial indicators to a unified scale is carried out by the method of linear data normalization.

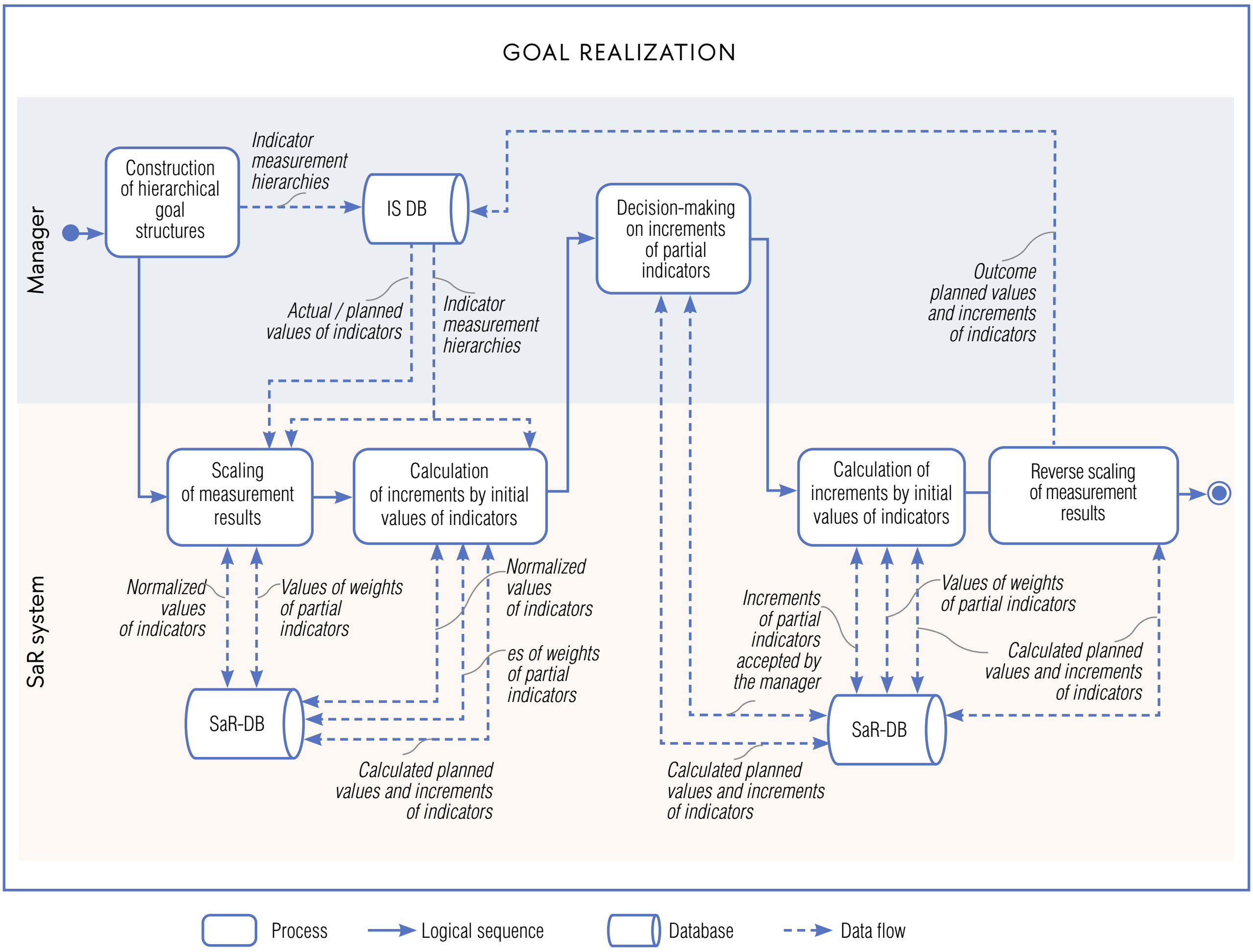

Fig. 2. Scheme of operation of the inverse calculations algorithm and SaR.

The values of partial indicators in the calculation of the integral indicator should be unidirectional: either increasing or decreasing. In addition, the values of a certain partial indicator should uniformly fill the interval defined by the empirical range of its values. If the data contain outliers, normalization is carried out using calculations of the mean and variance.

According to the approach, each partial indicator is described by a real number $h_i(x, u) \in [0, 1]$, defining the value of action u in state x. The array of numbers $h(x, u)$ can consist of q possible indicators, so that $h(x, u) = [h_1(x, u), ..., h_q(x, u)]$. The labeled array induces only a partial order on the control actions in the case of multiple indicators. The total order is associated with the partial order through the manager's actions. This is equal to associating the weights of the relative importance of the indicators.

The next step is to reevaluate the weights of partial indicators in the calculation of the integral target indicator. Since the process of calculating the weights is assumed to be cyclic and continuous, at this step the decisions made in the previous cycle are considered, and at the zero point the system uses predetermined values of weights with uniform distribution. Thus, in case of rejection of the management actions proposed by SaR, the system revises the weights of partial indicators based on the management actions actually chosen by the manager. Then, using these values, the system calculates new weights to be applied in new iteration $(t + 1)$.

Beginning at iteration t a sliding window of length k is defined such that $t - k + 1 \geqslant 0$. The values of the weights $w(t + 1)$ depend on the values of the system state variable at iteration τ, denoted as $x(\tau)$, the values of the optimal managerial actions computed by the system at iteration τ, denoted as $u(\tau)$, the values of the managerial actions actually chosen by the manager at iteration τ, after rejecting $u(\tau)$, denoted as $u'(\tau)$, for $\tau \in [t - k + 1, t] = \{t - k + 1, ..., t\}$.

For each past iteration τ, the actions actually chosen by the manager in iteration τ $(u'(\tau))$ should be preferred over the actions computed by the system $(u(\tau))$, depending on the optimization criterion. This condition ensures that the new weights are consistent with the knowledge accumulated by the system during the sliding window. This condition is formulated mathematically as follows (considered the case of maximization of the integral indicator):

$$w(t+1)(h(x(\tau), u'(\tau)) - h(x(\tau), u(\tau))) \geqslant \delta \qquad (1)$$

where the value of δ must be small, but not zero.

An additional constraint is imposed on the difference between the new values of weights $w(t + 1)$ and the values of weights at the current iteration $w(t)$. The difference between these values must be minimal to avoid fluctuations of weights that interfere with the convergence of the algorithm. This condition is formulated mathematically as follows:

$$W(t) = \sum_{\tau=t-k+1}^{t} \sum_{i=1}^{q} \left| w_i(t+1) - w_i(\tau) \right|. \qquad (2)$$

The mentioned features of the process of weighting partial indicators can be defined in terms of linear constraints in the MILP problem. The optimization problem, except for the additional conditions of a certain integral indicator, in this case will be given the form:

$$Z(\delta, w) = \delta + W(t) \to \min_{\delta,\, w(t+1)}, \qquad (3)$$

subject to:

$$\begin{cases} w(t+1)\big(h(x(\tau), u'(\tau)) - h(x(\tau), u(\tau))\big) - \delta \geq 0, \\ \delta \geq 0{,}001, \\ AbsVal\big(W(t), w_i(t+1), w_i(\tau)\big), \end{cases}$$

where δ, $w(t + 1)$ are solving variables,
$\tau \in [t - k + 1, t]$, $i \in [1, q]$.

After calculating the weights of partial indicators which form the function of the target integral indicator, there follows the phase of calculating the increments of the arguments of the function using the procedure of inverse calculations.

The target performance indicator is specified as a function $y = f(x_1(w_1), ..., x_q(w_q))$. The function uses weighting coefficients (priorities) received at the previous stage; their sum is always equal to one. The target function and its arguments may require positive increment or negative increment depending on the target settings. The sign of the increment in the target function in the system of equations depends on the pattern of change in the function y.

A solution method with the help of individual coefficients is used as a method of inverse calculations. According to this method, the manager's conditions for changing the values of partial performance indicators are expressed by specifying additional coefficients (k_q) when determining the required arguments of the target function:

$$x_q + \Delta x_q = k_q x_q. \qquad (4)$$

The sign of the increment of the arguments determines the values of the individual coefficients, which are calculated for each argument of the function: if the increment is positive, the individual coefficient is multiplied by its argument $(k_q x_q)$; if it is negative, the argument is divided by the individual coefficient $\left(\frac{x_q}{k_q}\right)$.

The number of additional equations in the system depends on the number of function arguments: the number of function arguments q will correspond to the number of additional equations $q - 1$. Constraints on the values of individual coefficients k_q are set up based on their semantics.

Therefore, the problem of calculating increments of partial indicators from the known increment of the target indicator in the case with three arguments and positive increment of both the target function and the arguments is described by a system of equations:

$$\begin{cases} y+\Delta y = f\left(k_1x_1, k_2x_2, k_3x_3\right), \\ \dfrac{k_1x_1 - x_1}{k_2x_2 - x_2 + k_3x_3 - x_3} = \dfrac{w_1}{w_2 + w_3}, \\ \dfrac{k_2x_2 - x_2}{k_1x_1 - x_1 + k_3x_3 - x_3} = \dfrac{w_2}{w_1 + w_3}. \end{cases} \quad (5)$$

Solving this system of equations regarding k_1, k_2 and k_3, it is possible to obtain the values of the incre-

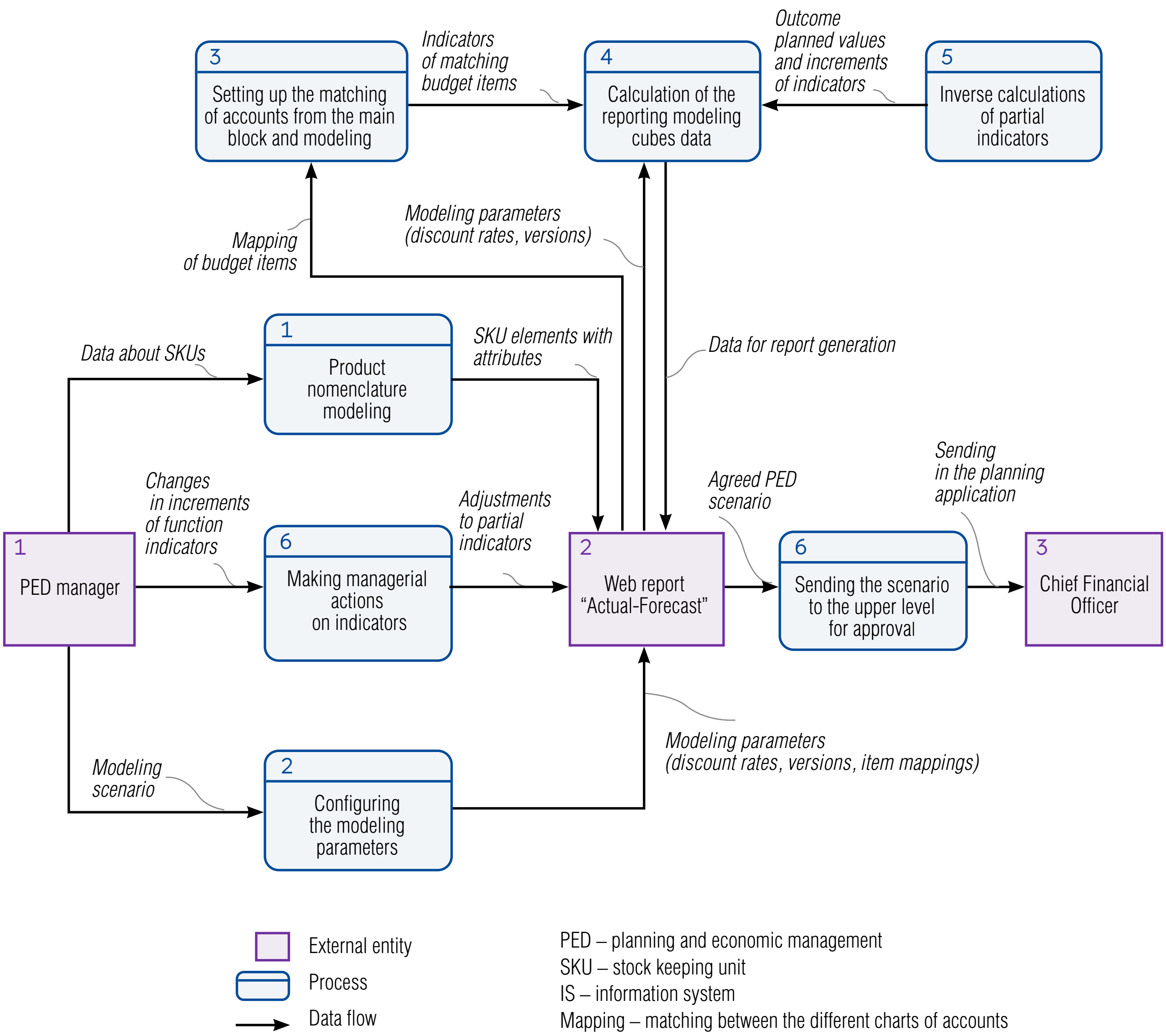

Fig. 3. Data flow diagram of the scenario modeling block "Actual–Forecast"

ments of the target function arguments x_1, x_2 and x_3. The results of the solution must satisfy the condition: $k_1, k_2, k_3 \geqslant 1$.

At the next steps of the target realization algorithm, the manager receives information about the calculated increments of partial indicators and decides either to initiate the action proposed by the system or to initiate another action. The decision made by the manager corresponds to the optimal solution at the current iteration and will be used by the system in a new iteration of calculations when updating the weights. Since data normalization was performed before the calculations, the next step of the algorithm involves a reverse scaling procedure using the inverse scale converter algorithm.

Indicators of the company's activity are the factors which form the basis for the formation of strategic indicators of the company's activity determined by the management. The growth of some indicators can lead to the growth of strategic indicators, and others to their reduction. Thus, using dynamic programming it is possible to obtain a balanced system of strategic indicators of the company's development, satisfying the requirements of management by changing the values of indicators of the company's activity. As a result, based on the strategic indicators of the company's activity, it is possible to determine its corporate performance as a convolution of these indicators.

3.3. Development of target management technology

The second part of the target realization process, the development of target management technology, is described in the context of information flow in the corporate performance management system. For this purpose, the data flow diagram is used, which demonstrates data flows between individual operations of the specialized block of scenario modeling "Actual–Forecast" included in the basic model of the company's activity (*Fig. 3*). This block allows the PED to develop management actions based on real data from the FRCs in parallel with the main budgeting process, forming budget scenarios as a result.

As the starting point of the budget modeling process for the next planning period, the PED manager's actions are considered to analyze and form the budget scenario in the interactive reporting form "Actual–Forecast" after entering actual and planned data for separate FRCs into the system. The outlined actions of the manager include: modeling of the nomenclature and relevant attributes of the products to be considered in the scenario; setting of modeling parameters – specifying the appropriate values of the discount rate, selecting the current version, specifying the matching between different charts of accounts to be analyzed and reported.

Both as the data in the reporting form is updated and at the manager's direct request (function launch), the rules in the corresponding multidimensional cubes of the system perform calculations of indicator values and output to the reporting form. The procedure of inverse calculations to update the calculation rules for launching requires a decision from the manager. The process of generating managerial actions on indicators means choosing between the values of adjustments suggested by the system and manual input of adjustments.

The output of the process produces a plan approved by the PED manager and is passed by the system procedures to the top level of approval – the Chief Financial Officer.

Conclusion

This paper presents an improved model of the target budgeting process, which enables automated generation of management actions of the budgeting department and subdivisions' management towards achieving strategic goals. In contrast to the findings of previous research, in addition to the parallel assessment of the company's state, the generation of management actions using artificial intelligence and rolling planning, the budgeting process model in this study includes a modeling block and aims to reduce the frequency of adjustments to operational goals that are set at the beginning of the planning period.

The second aspect of the theoretical contribution of the paper is the development of a methodological framework of inverse calculations in support of SaR methodology in the budgeting process of corporate performance management. The application of inverse calculations provides a mathematical statement of the problem of calculating indicators of planned key values, and the SaR system allows us to supplement the mathematical statement with weighting coefficients of primary performance indicators calculated algorithmically based on the manager's decisions instead of expert evaluation.

Applying the proposed approach in practice allows us to significantly reduce the value of transaction costs emerging at the stage of numerous adjustments of versions formed by the FRCs and scenarios formed by the PED from planning imprecisions and labor costs to identify and correct them. In addition, the approach we developed ensures improvement of the real quality of operational planning by the highest priority criteria of operability, accuracy and adaptability of planning due to systematic and methodical budgeting with the involvement of modern information technologies.

The directions of further research involve the integration of the developed information model of budgeting in the process of determining a comprehensive indicator characterizing the corporate performance of the company using the method of dynamic programming to calculate the strategic performance of the company based on the convolution of indicators.

Furthermore, as an area of future research on the problematic being studied, the design of the information infrastructure of the corporate performance management system is addressed to improve the efficiency of the development of managerial actions in the implementation of the company's strategic goals. Based on the methodology of Design Science Research, the blocks of analytical support are considered as an artifact that is described by designing information-logical schemes and evaluated through experimental testing of the approach in the work environment. ■

References

1. Maryska M., Doucek P. (2017) REMONA model for improving quality of corporate informatics performance management: How to cut cots in corporate informatics. *Quality Innovation Prosperity*, vol. 21, no. 3, pp. 15–35. https://doi.org/10.12776/QIP.V21I3.939

2. Richards G., Yeoh W., Chong A.Y.L., Popovič A. (2019) Business intelligence effectiveness and corporate performance management: An empirical analysis. *Journal of Computer Information Systems*, vol. 59, no. 2, pp. 188–196. https://doi.org/10.1080/08874417.2017.1334244

3. Bourne M., Neely A., Mills J., Platts K. (2003) Implementing performance measurement systems: A literature review. *International Journal of Business Performance Management*, vol. 5, no. 1, pp. 1–24. https://doi.org/10.1504/IJBPM.2003.002097

4. Zhang J. (2022) A neural network model for business performance management based on random matrix theory. *Mathematical Problems in Engineering*, vol. 2022, article ID 9170666. https://doi.org/10.1155/2022/9170666

5. Mari F., Massini A., Melatti I., Tronci E. (2021) A constraint optimization–based sense and response system for interactive business performance management. *Applied Artificial Intelligence*, vol. 35, no. 5, pp. 353–372. https://doi.org/10.1080/08839514.2020.1843833

6. Kitova O.V. (2012) *Concepts and information infrastructure of marketing performance management: theory and methodology*. St. Petersburg: UNECON (in Russian).

7. Bruskin S.N. (2016) *Information-analytical system of decision-making support in the field of planning sales activities of the corporation.* Moscow: MSU (in Russian).

8. Odintsov B.E. (2023) *Information systems of business performance management. Textbook and Practice for Bachelor's and Master's Degrees.* Moscow: Urait (in Russian).

9. Nikishova M.I. (2021) *Application of artificial intelligence technologies in the corporate governance system.* Moscow: Financial University (in Russian).

10. Neely A., Gregory M., Platts K. (1995) Performance measurement system design—a literature review and research agenda. *International Journal of Operations and Production Management*, vol. 15, no. 4, pp. 80–116. https://doi.org/10.1108/01443570510633639

11. Kaplan R.S., Norton D.P. (1996) *The balanced scorecard. Translating strategy into action.* Boston, MA: Harvard Business School Press.

12. Kaplan R.S., Norton D.P. (1996) Using the balanced scorecard as a strategic management system. *Harvard Business Review*, pp. 75–85.

13. Neely A.D. et al. (1996) *Getting the measure of your business*. London: Findlay.

14. Dixon J.R., Nanni A.J., Vollmann T.E. (1990) *The new performance challenge: Measuring operations for world-class competition*. Homewood, IL: Business One Irwin.

15. Krause O., Mertins K. (1999) Performance management. *Global Production Management: IFIP WG5.7 International Conference on Advances in Production Management Systems, Berlin, Germany, September 6–10, 1999* (eds. K. Mertins, O. Krause, B. Schallock), pp. 243–251.

16. Kaplan R.S., Norton D.P. (1993) Putting the balanced scorecard to work. *Harvard Business Review*, pp. 134–147. https://doi.org/10.1016/B978-0-7506-7009-8.50023-9

17. Rybalko O.A., Shalaeva L.V. (2012) Strategic planning and budgeting as basic elements of modern management system. *International Accounting*, no. 28, pp. 25–38 (in Russian).

18. Dresner H. (2007) *The performance management revolution: business results through insight and action*. Hoboken, NJ: Wiley.

19. Taticchi P., Tonelli F., Cagnazzo L. (2010) Performance measurement and management: a literature review and a research agenda. *Measuring Business Excellence*, vol. 14, no. 1, pp. 4–18. https://doi.org/10.1108/13683041011027418

20. Medeiros M.M., Maçada A.C.G., Hoppen N. (2021) The role of big data stewardship and analytics as enablers of corporate performance management. *RAM. Revista de Administração Mackenzie*, vol. 22. https://doi.org/10.1590/1678-6971/eramd210063

21. Jaklič J., Bosilj-Vukšić V., Mendling J., Štemberger M.I. (2021) The orchestration of corporate performance management and business process management and its effect on perceived organizational performance. *SAGE Open*, vol. 11, no. 3. https://doi.org/10.1177/21582440211040126

22. Prilianti E., Hikmat M.T. (2018) Proposed corporate performance management using integrated performance management system (IPMS) at PT Pos Indonesia (Persero). *International Journal of Engineering & Technology*, vol. 7, no 3, pp. 71–83. https://doi.org/10.14419/ijet.v7i3.25.17472

23. Merkert J., Mueller M., Hubl M. (2015) A Survey of the Application of Machine Learning in Decision Support Systems. *ECIS Completed Research Papers*, paper 133. https://doi.org/10.18151/7217429

24. Dugelny A.P. (2003) *Budgetary management of the enterprise*. Moscow: Delo (in Russian).

About the author

Maxim E. Oshchepkov

Doctoral Student, Department of Business Informatics, Graduate School of Business, HSE University, 26–28, Shabolovka St., Moscow 119049, Russia

E-mail: moshchepkov@hse.ru

ORCID: 0000-0003-0327-3285

DOI: 10.17323/2587-814X.2024.1.36.51

Agent-based modeling and optimization of the characteristics for research-and-production clusters

Gayane L. Beklaryan
E-mail: glbeklaryan@gmail.com

Central Economics and Mathematics Institute, Russian Academy of Sciences
Address: 47, Nakhimovsky Ave., Moscow 117418, Russia

Abstract

This paper presents a developed agent-based simulation model for the development of research-and-production clusters in Russia implemented with the use of high-tech enterprises located in four science cities (Troitsk, Obninsk, Pushchino and Protvino) as the case study. A new approach to modeling and optimization of gross metropolitan product (GMP) is proposed, taking into account the influence of the "gravity effect" on the redistribution of labor resources between developing science cities and appropriate enterprises united in single research and research-and-production clusters An important element of this approach is the formation of various scenarios for the strategic development of the research-and-production clusters being assessed and support for the possibility of choosing the most preferable scenario using an evolutionary optimization algorithm. An enlarged simulation model has been developed and implemented in AnyLogic describing the possible development trajectories of science cities with a corresponding change in the values of the most important characteristics: the number of economically active population, the number of research-and-production enterprises, the volume of products produced in high-tech sectors of the economy, GMP, etc. The designed framework is intended primarily for the management of research-and-production clusters implementing the strategy of innovative development. Such a framework uses methods of system dynamics and agent-based simulation modeling supported in the AnyLogic system, genetic optimization algorithms and GIS mapping for science cities, etc. to implement the required functionality. The approbation of the framework was completed with the use of real data published in the approved strategies of the relevant science cities development. As a result of the numerical experiments carried out, some recommendations were proposed for the development of the research-and-production clusters under study considering their mutual influence and the existing base of resources.

Keywords: research-and-production cluster, high-tech enterprise, science city, production characteristics, simulation modeling for enterprises, gross metropolitan product, agent-based modeling, system dynamics, gravity effect, production function, AnyLogic

Citation: Beklaryan G.L. (2024) Agent-based modeling and optimization of the characteristics for research-and-production clusters. *Business Informatics*, vol. 18, no. 1, pp. 36–51. DOI: 10.17323/2587-814X.2024.1.36.51

Introduction

Currently, a very pressing problem of increasing the efficiency of the public administration system is the development of research and production clusters using the existing human resource potential, including those formed on the basis of research institutes and research-and-production clusters that have experience in using own developments in the real sector of the economy.

Determining the best development scenarios for high-tech enterprises located in science cities, developing and implementing decision support systems for managing research and production clusters is the most important area of business informatics aimed at increasing the efficiency of managing innovative development processes.

As an example, high-tech enterprises located in science cities united within a single conceptual model of innovative development will be considered. Within such a model, it is possible to obtain a significant synergistic effect, due, firstly, to the possibilities of staff cooperation, as well as the use of a common high-tech resource base (e.g., supercomputers, data centers, industrial systems for 3D printing of integrated circuits, etc.). In particular, high-tech enterprises located in cities such as Troitsk, Obninsk, Pushchino and Protvino near the Moscow agglomeration and being essentially neighboring enterprises can be united into single research-and-production clusters, in particular, for the production of a microelectronic element base, precision instrumentation products, etc. At the same time, it is important to ensure the balanced development of high-tech enterprises in all small cities united in the respective research-and-production clusters in order to avoid a one-way flow of highly qualified staff and ensure the harmonious strategic development of all agent enterprises (research institutes, research-and-production clusters, etc.) as existing ones, and those formed in the corresponding science cities.

Therefore, it is necessary to develop a decision support system (software framework) for managing high-tech enterprises located in science cities and implementing an innovative development strategy. Such a complex uses methods of system dynamics and agent-based simulation modeling supported in the AnyLogic system, genetic optimization algorithms and GIS maps of research-and-production clusters, etc. to implement the required functionality.

The first fundamental work on the study of the dynamics of the development of a city and its enterprises is the book [1], in which Forrester formulated the basic principles for constructing simulation models that describe various scenarios for the evolution of the urban environment, taking into account the complex system of existing feedbacks, lag dependencies and internal urban interactions, etc.

Among the recent works in this area, it is worth highlighting [2–4], in which system dynamics methods are used to study the possibilities of achieving sustainable economic development of large cities. In particular, in [2] four scenarios of possible development for the city of Beijing were modeled using a system-dynamic approach. It is shown that the dynamics of sustainable development of the city are implemented differently during three periods (from the growth of the city's economy at the beginning to its slowdown and subsequent growth).

Among the disadvantages of using system dynamics methods, one should note the difficulties of modeling internal interactions, processes of redistribution of labor resources, difficulties with the analysis of spatial development of the city, etc. To overcome such difficulties, it is advisable to use agent-based simulation modeling methods that allow modeling interactions at the microscopic level, taking into account individual adoption systems decisions of each economic agent [5, 6].

Among the important works in this area, [7–10] should be highlighted. In particular, [7] presents an aggregated agent-based model of migration flows of the European Union countries and describes the influence of the "gravity effect" on the inter-country redistribution of human flows. This effect arises due to factors influencing the migration behavior of people, in particular, when there is a significant difference in the ratio of wages, labor market size and gross domestic product (GDP) per capita between a given country (region) and neighboring countries (regions). For instance, relatively low wages combined with a small number of available jobs tend to lead to an outflow of the population to more favorable regions. The "gravity effect" also influences internal migration between cities, when the population (and corresponding labor resources) moves from the agglomerations that are less developed to economically more developed ones. In [8] an agent-based model of population dynamics of two interacting communities consisting of migrants and natives is presented. In such a model, natives search for high-tech workplaces, and migrants search for low-tech workplaces, which is also relevant for enterprises in science cities that actively attract external labor resources. Agent-based modeling can be combined with other methods, e.g., system dynamics, discrete event modeling, etc. to create digital twins of complex socio-economic systems [9, 10]. The developed simulation models can be aggregated according to objective functions and constraints with evolutionary optimization algorithms, in particular, genetic algorithms [11, 12] to optimize the characteristics of such large-scale systems. Agent-based models of territorial development of regional and urban agglomerations can aim at solving important environmental problems, such as reducing the concentration of harmful emissions in socially significant urban areas (e.g., near kindergartens and schools) through landscaping [13], reducing emissions from industrial enterprises due to their environmental modernization [14], transformation of the urban environment [15]. To model the movement of human flows and labor resources in an urban environment, it is advisable to use the phenomenological approach previously proposed in [16], which allows us to take into account various scenarios of interaction of various agents with each other, simulate the behavior of the crowd, etc. Also noteworthy are the works [17–21] devoted to both methodological issues [17] and specific case studies carried out to study the possibilities of developing urban architecture [19], improving traffic [20] and increasing returns in certain sectors of the urban economy [21].

Thus, the combined use of various simulation modeling methods, heuristic optimization algorithms, etc., in particular, supported in the AnyLogic system, makes it possible to design a decision support system for the sustainable economic development of research-and-production clusters and corresponding science cities.

The purpose of this paper is to develop economic, mathematical and computer tools for studying the dynamics of development of individual research-and-production clusters of the Russian Federation under various scenario conditions within the framework of which the balanced development of all interacting economic agents-enterprises (research institutes, research-and-production clusters, etc.) is ensured.

1. Simulation model for the development of research-and-production clusters

This section presents the developed simulation model for the development of research-and-production clusters with implementation using the example of high-tech enterprises located in the science cities of Troitsk, Obninsk, Pushchino and Protvino. The proposed agent-based model consists of several interconnected levels (*Fig. 1*).

Top level model – the environment in which agent enterprises operate, belonging to different cities, between which flow interaction is realized, in particular, redistribution (internal migration) of the population with a corresponding change in labor resources, which can be involved in the high-tech sector of the economy (in research institutes, SPAs and etc.) in the case of the formation of new jobs (*Fig. 1*). Such flow interaction is carried out due to the influence of the "gravitational" effect [7] caused by differences in the level of wages, the size of the labor market and GDP per capita in the corresponding science cities.

The *Low level model* ensures the implementation of the logic of behavior of agent enterprises, the characteristics of which are described using system dynamics methods (*Fig. 1*). Such internal models are systems of finite-difference equations with their own resource and target characteristics [11, 22]. At the same time, to determine the values of individual parameters, integration with the upper level of the simulation model is required. In particular, the main control parameters of the model, the values of which must be the same for all agents, are set at the environment level. At the same time, individual characteristics calculated at the agent level are used in the top-level model, for example, to calculate the gross metropolitan product (GMP) generated by all enterprises.

For each enterprise agent, three possible states are specified:

- the first state, corresponding to low economic growth rates and not exceeding a given lower threshold value, is highlighted in "red";
- the second state, corresponding to the average economic growth rates, the valuesof which are in the interval between the lower and upper threshold values, is highlighted in "yellow";
- the third state, corresponding to high rates of economic growth, the values of which exceed the upper threshold value, is highlighted as "green".

The economic growth rate of agent enterprises (the dynamics of GMP) is influenced by various factors, among which the most important are investments in fixed capital, the number of jobs created, average monthly wages, etc. The production resources that enterprises already have (fixed assets, staff) are also important: the level of scientific and technological progress (STP), the ability to attract highly qualified staff, etc. Within the framework of the developed model, a centralized allocation of resources (investments, new workplaces, etc.) is proposed with their subsequent distribution among enterprises of each science city, taking into account greater support for problem organizations characterized by low rates of economic growth. This approach is relevant, first of all, for state enterprises, e.g., institutes of the Russian Academy of Sciences, research institutes, as well as commercial organizations with significant government participation.

Thus, the most important task of the system under consideration is to minimize the total number of problem ("red") enterprises with the minimum required investment and operating costs.

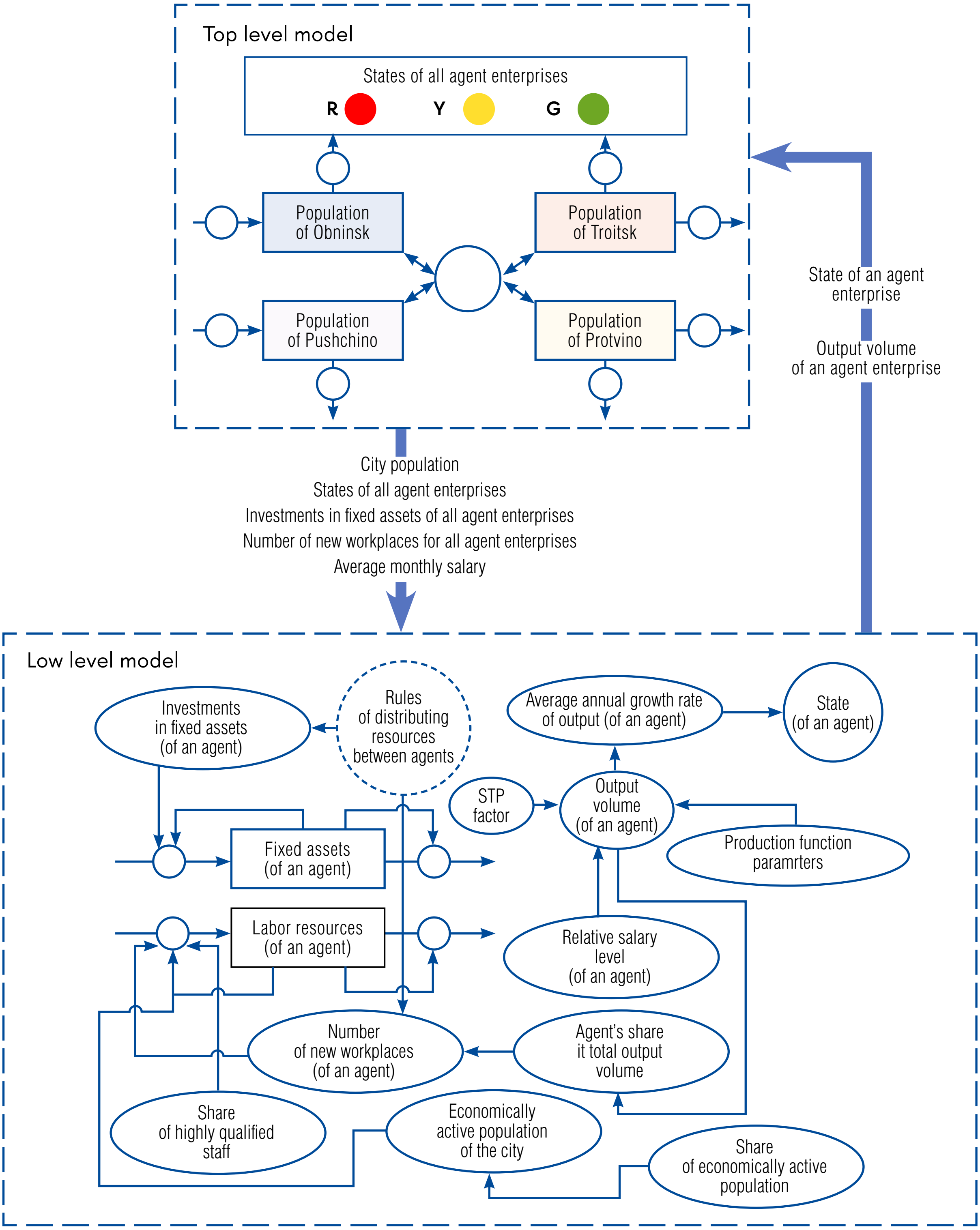

Fig.1. Aggregated architecture of a simulation model of research-and-production clusters.

1.1. Model of labor resource redistribution

The top-level model presented in *Fig. 1* ensures the implementation of the logic of redistribution of population and labor resources, taking into account the influence of the "gravity effect." A similar effect was first described in works [23, 24] devoted to the analysis of human behavior under the influence of social forces. In particular, the decision to migrate to another city is influenced by multiple factors: differences in wages, the size of the labor market, GDP per capita, climatic characteristics, level of social security, etc. To a greater extent, the "gravitational" influence of such differentiation manifests itself at the level of countries that differ significantly in their level and quality of life. However, similar factors promoting migration also apply to cities, including science cities, the features of which allow their residents to move while remaining employed in high-tech sectors of the economy.

Here,

- $T = \{t_1, t_2, ..., t_{|T|}\}$ is the set of moments of the model time (by year), $|T|$ is the strategic planning horizon (10 years);
- $I = \{i_1, i_2, ..., i_{|I|}\}$ is the set of indices of science cities, $|I|$ – total number of science cities;
- $\{\nu_i, \eta_i\}$, $i \in I$ are the fertility and mortality rates in the i-th science city;
- $\{GMP_i(t_k), P_i(t_k)\}$, $i \in I$ are the gross metropolitan product and population of the i-th science city at moment t_k, $(t_k \in T)$;
- $\{W_i(t_k), L_i(t_k)\}$, $i \in I$ are the level of wage and amount of labor resources of the i-th science city employed in high-tech sectors of the economy at moment t_k, $(t_k \in T)$;
- $\{\tilde{w}_1, \tilde{w}_2, \tilde{w}_3\} \in [0, 1]$ are the weighting coefficients that determine the level of influence of individual factors on migration flows, where $\tilde{w}_1 + \tilde{w}_2 + \tilde{w}_3 = 1$;
- $\varkappa_i \in [0, 1]$ is the coefficient that determines the level of influence of the "gravity effect" on emigration to the i-th science city.

An emigration rate from the i-th science city $(i \in I)$ to the neighboring ξ-th science city $(\xi \in I, \xi \neq i)$ at moment t_k, $(t_k \in T)$ is equal to

$$E_{i\xi}(t_k) = \begin{cases} \varkappa_i \left(\dfrac{1}{g_{i\xi}(t_k)} - 1 \right) P_i(t_{k-1}), \text{ if } g_{i\xi}(t_k) < 1, \\ 0, \text{ if } g_{i\xi}(t_k) \geq 1, \end{cases} \tag{1}$$

where

$$\ln\left(g_{i\xi}(t_k)\right) = \tilde{w}_1 \ln\left(\frac{W_i(t_{k-1})}{W_\xi(t_{k-1})}\right) + \tilde{w}_2 \ln\left(\frac{L_i(t_{k-1})}{L_\xi(t_{k-1})}\right) + \tilde{w}_3 \ln\left(\frac{\dfrac{GMP_i(t_{k-1})}{P_i(t_{k-1})}}{\dfrac{GMP_\xi(t_{k-1})}{P_\xi(t_{k-1})}}\right). \tag{2}$$

The equations (1)–(2) describe the influence of the "gravity effect" on migration flows. Expression (2) reflects the influence of such key factors as wages $W_i(t_{k-1})$, number of employed labor resources $L_i(t_{k-1})$ and GDP per capita

$$\frac{GMP_i(t_{k-1})}{P_i(t_{k-1})}$$

on the migration behavior of population (i.e., migration rates) in the i-th science city $(i \in I)$. The choice of such factors is determined, first of all, by the observed difference in their values in relation to the science cities under consideration (i.e., Troitsk, Obninsk, etc.). The influence of other potentially important characteristics, e.g., differences in climate, transport accessibility, ecology, etc., is not significant from the geographic proximity and spatial homogeneity of the studied agglomerations. The use of relative and scaled (logarithmized) values of influencing factors allows them to be aggregated using (2) for subsequent calculation of migration rates to neighboring science cities.

From (1)–(2) it follows that if a given science city is less attractive than a neighboring one which has higher wages, more jobs and more GDP per capita, then a positive emigration flow to this science city is provided. The opposite scenario is also possible, when a given science city is more attractive to residents of neighboring cities. In this case, the formation of positive immigration flows will be ensured.

The population of the i-th science city ($i \in I$) considering the emigration to the neighboring ξ-th science cities ($\xi \in I, \xi \neq i$), the immigration from the neighboring ξ-ths science cities ($\xi \in I, \xi \neq i$), as well as the fertility and mortality at moment t_k, ($t_k \in T$) is equal to

$$P_i(t_k) = P_i(t_{k-1}) - \sum_{\xi=1}^{|I|} E_{i\xi}(t_k) + \sum_{\xi=1}^{|I|} E_{\xi i}(t_k) + \\ + v_i P_i(t_{k-1}) - \eta_i P_i(t_{k-1}). \quad (3)$$

Together with the redistribution of the population between neighboring science cities, the redistribution of highly qualified labor is being implemented accordingly.

The amount of labor resources that can be employed in high-tech enterprises of the i-th science city ($i \in I$) at moment t_k, ($t_k \in T$) is equal to

$$L_i^*(t_k) = \mu_i \omega_i P_i(t_k), \quad (4)$$

where

$\mu_i \in [0, 1]$ is the share of economically active population in the i-th science city ($i \in I$);

$\omega_i \in [0, 1]$ is the share of highly qualified labor resources from the economically active population in the i-th science city ($i \in I$).

1.2. Model of behavior of enterprise agents

The lower-level model presented in *Fig. 1* is used to calculate the dynamics of the average annual growth rate of output of agent enterprises with a corresponding assessment of the states of such agents. The most important characteristic of such enterprises is the production function, which describes the influence of fixed assets, labor resources, the level of scientific and technological development and other factors on the dynamics of output volume. A detailed description of the principles for constructing production functions, including the Cobb-Douglas type used in this work is presented in [25, 26]. There are also examples of constructing multiplicative production functions that take into account the influence of wages on output, as a factor that significantly affects the productivity of labor resources (e.g., [27]). A feature of the production function proposed in this article is that it takes into account the relative (in relation to all neighboring science cities) level of wages. The higher this level, the more qualified labor resources can be employed at enterprises in a given city.

Let be,

- $J_i = \{j_{i1}, j_{i2}, \ldots, j_{i|J_i|}\}, i \in I$ is the set of indices of agent enterprises located in the i-th science city, where $|J_i|$ is the total number of enterprises;
- $s_{j_i}(t_k) \in \{1, 2, 3\}$ are the states of the j_i-th agent enterprise of the i-th science city at momen t_k, ($t_k \in T$): $s_{j_i}(t_k) = 1$ is the first state corresponding to low economic growth rates ("red"), $s_{j_i}(t_k) = 2$ is the second state, corresponding to the average rate of economic growth ("yellow"), $s_{j_i}(t_k) = 3$ is the third state, corresponding to high rates of economic growth ("green");
- $\{\tilde{K}_{j_i}(t_k), \tilde{L}_{j_i}(t_k)\}, i \in I, j_i \in J_i$ are the fixed assets and labor resources of the j_i-th agent enterprise of the i-th science city at moment t_k, ($t_k \in T$);
- $\{\delta\tilde{K}_{j_i}(t_k), \delta\tilde{L}_{j_i}(t_k)\}, i \in I, j_i \in J_i$ are the rate of new fixed assets commissioning and the labor resources inflow of the j_i-th agent enterprise of the i-th science city at moment t_k, ($t_k \in T$);
- $\{Q_i(t_k), Y_i(t_k)\}, i \in I$ are investments in fixed assets and the number of new workplaces allocated by the i-th science city at moment t_k, ($t_k \in T$);
- $\{\alpha_{j_i}, \beta_{j_i}\}$ (0, 1), $i \in I, j_i \in J_i$ are the coefficients of elasticity of output volume in relation to fixed assets and labor resources of the j_i-th agent enterprise of the i-th science city, where $\alpha_{j_i} + \beta_{j_i} = 1$;
- $A_{j_i}, i \in I, j_i \in J_i$ is the factor (coefficient) reflecting the influence of scientific and technological progress (STP) on the dynamics of output volume;
- $\{\mu, \iota\}$ are the known retirement rates of fixed assets and labor resources, respectively.

Then, the output volume of the j_i-th agent enterprise ($j_i \in J_i$) of the i-th science city ($i \in I$) can be specified with the use of the Cobb-Douglas type production function at momentt_k, ($t_k \in T$):

$$\tilde{V}_{j_i}(t_k) = A_{j_i} S_{j_i} \left(\tilde{K}_{j_i}(t_k)\right)^{\alpha_{j_i}} \left(\tilde{L}_{j_i}(t_k)\right)^{\beta_{j_i}}, \tag{5}$$

where

$$S_{j_i} = \frac{W_i}{\frac{1}{|I|}\sum_{i=1}^{|I|} W_i}, \tag{6}$$

$$\tilde{K}_{j_i}(t_k) = \tilde{K}_{j_i}(t_{k-1}) + \delta\tilde{K}_{j_i}(t_k) - \mu\tilde{K}_{j_i}(t_{k-1}), \tag{7}$$

$$\tilde{L}_{j_i}(t_k) = \tilde{L}_{j_i}(t_{k-1}) + \delta\tilde{L}_{j_i}(t_k) - \iota\tilde{L}_{j_i}(t_{k-1}), \tag{8}$$

$$\delta\tilde{K}_{j_i}(t_k) = \begin{cases} Q_i(t_k)\dfrac{\tilde{V}_{j_i}(t_{k-1})}{\sum_{j_i=1}^{|J_i|}\tilde{V}_{j_i}(t_{k-1})}\left(1+\gamma\dfrac{\sum_{j_i=1}^{|J_i|} r_{j_i}(t_{k-1})}{|J_i|}\right), \\ \text{if } s_{j_i}(t_{k-1}) = 1, \\ Q_i(t_k)\dfrac{\tilde{V}_{j_i}(t_{k-1})}{\sum_{j_i=1}^{|J_i|}\tilde{V}_{j_i}(t_{k-1})}\left(1-\gamma\dfrac{\sum_{j_i=1}^{|J_i|} r_{j_i}(t_{k-1})}{|J_i|}\right), \\ \text{if } s_{j_i}(t_{k-1}) = 3, \end{cases} \tag{9}$$

$$\delta\tilde{L}_{j_i}(t_k) = \begin{cases} Y_i(t_k)\dfrac{\tilde{V}_{j_i}(t_{k-1})}{\sum_{j_i=1}^{|J_i|}\tilde{V}_{j_i}(t_{k-1})}\left(1+\gamma\dfrac{\sum_{j_i=1}^{|J_i|} r_{j_i}(t_k)}{|J_i|}\right), \\ \text{if } s_{j_i}(t_{k-1}) = 1, \\ Y_i(t_k)\dfrac{\tilde{V}_{j_i}(t_{k-1})}{\sum_{j_i=1}^{|J_i|}\tilde{V}_{j_i}(t_{k-1})}\left(1-\gamma\dfrac{\sum_{j_i=1}^{|J_i|} r_{j_i}(t_k)}{|J_i|}\right), \\ \text{if } s_{j_i}(t_{k-1}) = 3, \end{cases} \tag{10}$$

$$r_{j_i}(t_k) = \begin{cases} 1, \text{ if } s_{j_i}(t_k) = 1, \\ 0, \text{ if } s_{j_i}(t_k) \neq 1. \end{cases} \tag{11}$$

Figure 2 shows the state-flow chart of the j_i-th agent enterprise ($j_i \in J_i$) of the i-th science city ($i \in I$) implementing the transition rules to new states $s_{j_i}(t_k) \in \{1, 2, 3\}$.

Transitions to new states (indicated by the sign ? in *Fig. 2*) are carried out using the following rule:

$$s_{j_i}(t_k) = \begin{cases} 1, \text{ if } \dfrac{1}{t_k}\sum_{\tilde{t}_k=1}^{t_k}\dfrac{\tilde{V}_{j_i}(\tilde{t}_k)}{\tilde{V}_{j_i}(\tilde{t}_{k-1})} < \varphi_1, \\ 2, \text{ if } \varphi_1 \le \dfrac{1}{t_k}\sum_{\tilde{t}_k=1}^{t_k}\dfrac{\tilde{V}_{j_i}(\tilde{t}_k)}{\tilde{V}_{j_i}(\tilde{t}_{k-1})} < \varphi_2, \\ 3, \text{ if } \dfrac{1}{t_k}\sum_{\tilde{t}_k=1}^{t_k}\dfrac{\tilde{V}_{j_i}(\tilde{t}_k)}{\tilde{V}_{j_i}(\tilde{t}_{k-1})} \ge \varphi_2. \end{cases}$$

Here, $\{\varphi_1, \varphi_2\}$ are the specified threshold values that determine the conditions for compliance of agent enterprises with states of low, medium and high rates of economic growth.

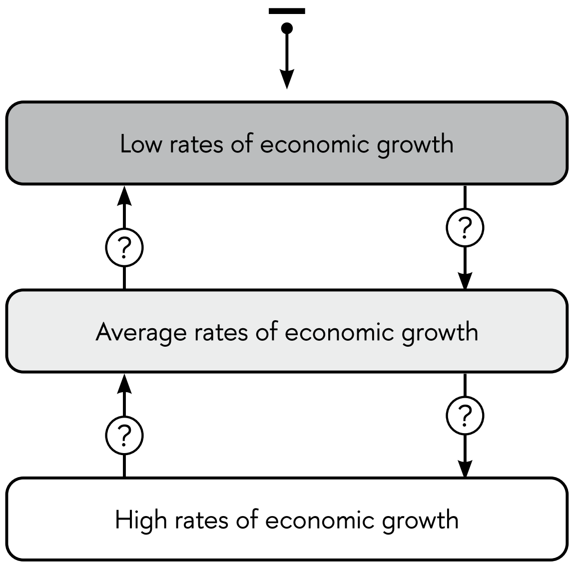

Fig.2. State-flow chart of an agent enterprise.

The proposed model takes into account the influence of the relative wage level (6) on output (5). The use of relations (9)–(11) allows us to provide priority support to agent enterprises that are in the first state of low economic growth rates. Here, $\gamma \in [0, 1]$ is the coefficient determining the overall level of such support. Thus, investments in fixed assets and new jobs allocated for each science city are distributed among "problem" agent enterprises in proportion to their contribution to the total output.

1.3. Optimization problem statement

A system is considered in which it is necessary to ensure the minimization of the total number of enterprises characterized by low rates of economic growth for all selected science cities. For the sustainable development of research-and-production clusters, it is desirable to ensure the transition of the majority of problem enterprises to states of medium and high rates of economic growth and the preservation of such favorable states for the longest possible time interval. However, in the current conditions of a shortage of investment capital, as well as limited opportunities to create new high-tech jobs, it is important to find such states and modes of operation of the system in which minimizing the number of problem organizations is achieved with minimal costs at least by some given point in model time. As a result, a qualitative improvement in the characteristics of the corresponding enterprises will be ensured, practically guaranteeing the preservation of the states they have achieved over a sufficiently long time interval due to the inertia of the values of production factors. In the absence of statistical information about the existing budget and investment restrictions, it is justified to reduce such an optimization problem to a single-criteria one with the following objective functional, estimated at the final moment of model time $t_{|T|}$, ($t_{|T|} \in T$):

$$N(t_{|T|}) = n_1 \sum_{i=1}^{|I|} \sum_{j_i=1}^{|J_i|} r_{j_i}\left(Q_i(t_k),\, Y_i(t_k),\, W_i(t_k)\right) + \\ + n_2 \sum_{i=1}^{|I|} Q_i(t_k) + n_3 \sum_{i=1}^{|I|} \left(Y_i(t_k) W_i(t_k)\right), \tag{13}$$

where

$$r_{j_i}\left(Q_i(t_k),\, Y_i(t_k),\, W_i(t_k)\right) = \begin{cases} 1, \text{if } s_{j_i}(t_k) = 1, \\ 0, \text{if } s_{j_i}(t_k) \neq 1. \end{cases} \tag{14}$$

Here,

- $\{Q_i(t_k),\ Y_i(t_k),\ W_i(t_k)\}$, $i \in I$ are the investments in fixed assets, number of new workplaces and wages in the i-th science city at moment t_k, ($t_k \in T$) which are control parameters of the model;
- $\{n_1, n_2, n_3\}$ are the normalizing factors that are used to reduce the values of the elements of the objective function values to the uniform scale.

The choice of values of the control parameters of the model is carried out at each moment t_k, ($t_k \in T$) and leads to changes in the production characteristics of agent enterprises, i.e., fixed assets, labor resources and output volumes in accordance with (5)–(11). As a result, the impact on the average annual rate of economic growth and the state of the corresponding enterprises is ensured, see (12)–(14).

Then the optimization problem can be formulated as follows.

Problem A. *The need to minimize the value of the total number of enterprises characterized by low rates of economic growth at the final moment of the model time* $t_{|T|}$, ($t_{|T|} \in T$), *as well as the investment and operating costs associated with the implementation of the strategy:*

$$\min_{\{Q_i(t_k),\, Y_i(t_k),\, W_i(t_k)\}_{k=1}^{|T|}} N(t_{|T|}) \tag{15}$$

s.t.

$$\underline{Q} \le Q_i(t_k) \le \bar{Q},\ \underline{Y} \le Y_i(t_k) \le \bar{Y},\ \underline{W} \le W_i(t_k) \le \bar{W}, \\ i \in I,\ k = 1, 2, \ldots, |T|.$$

Here, $\{\underline{Q}, \underline{Y}, \underline{W}\}$, $\{\bar{Q}, \bar{Y}, \bar{W}\}$ are known lower and upper limits of the simulation model control parameter values.

Problem A can be solved using a genetic optimization algorithm aggregated by objective function with the proposed simulation model. The implementation of such algorithms is based on methods of evolutionary search for the best potential solutions and is described in detail in [12, 28]. To simplify the problem under consideration, the values of control parameters in such a system can be specified for each i-th science city ($i \in I$) as constants at the initial moment of the model time.

2. Software implementation of the model

The software implementation of the developed simulation model (1)–(15) was carried out in the AnyLogic system. A feature of the proposed approach is the use of GIS maps to visualize the states of enterprise agents located in given spatial coordinates, as well as the integration of the model with a previously developed genetic optimization algorithm [28], which, in particular, provides a solution to **Problem A**.

The developed software package provides the ability to navigate through all the science cities being studied, vary the values of the control parameters of the model and visualize the states of agent enterprises on the map.

Figure 3 shows the implementation of the model for the redistribution of population and labor resources between the science cities under consideration (top-level models) in AnyLogic.

As seen from *Fig. 3* the system dynamics methods are used to construct this model (e.g., [11, 22]). The population of Obninsk, Troitsk, Protvino and Pushchino is modeled using four system levels (“reservoirs”) interacting through flows at rates whose values depend on the difference in wage levels, the size of the labor market, etc. Thus, the influence of the “gravity effect” on the rate of internal migration. Fertility and mortality rates with their known values are also taken into account. The reinforcing and balancing feedbacks present in the model are marked with the letters R and B, respectively (*Fig. 3*).

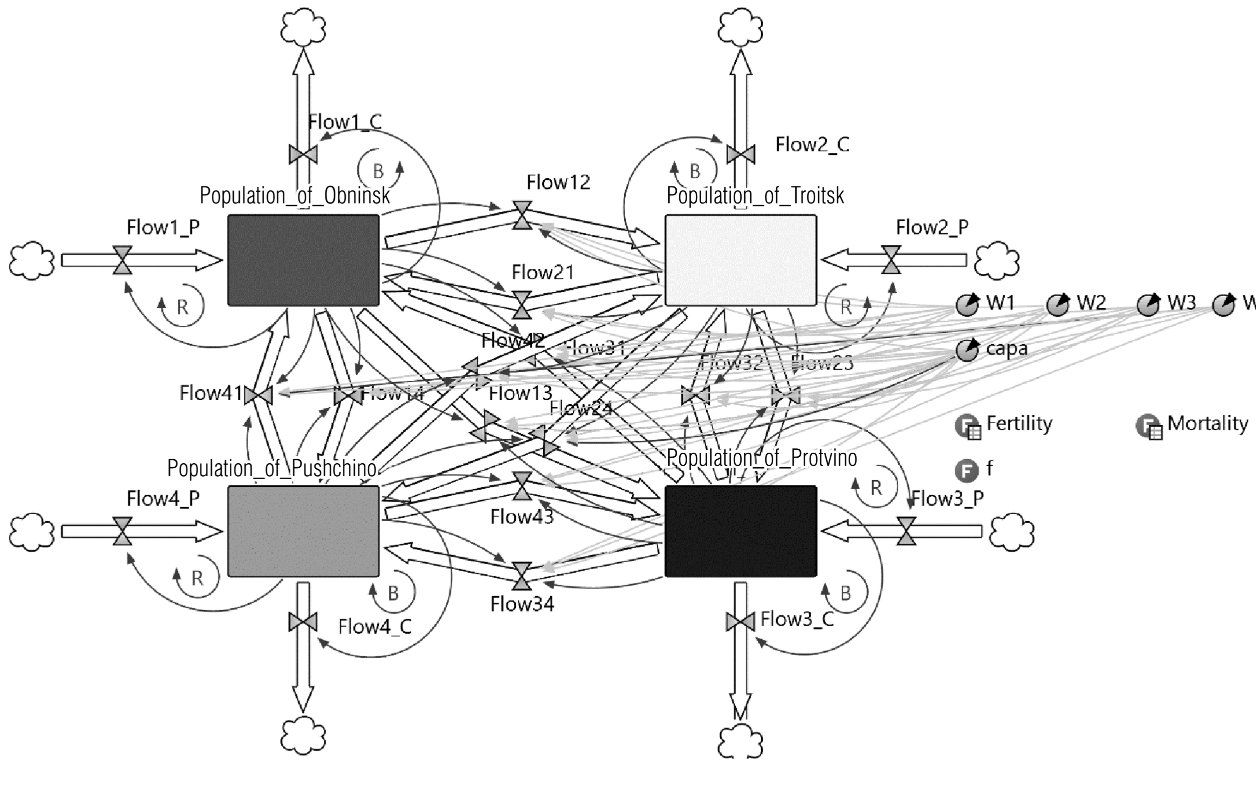

Fig. 3. Model of redistribution of labor resources in AnyLogic.

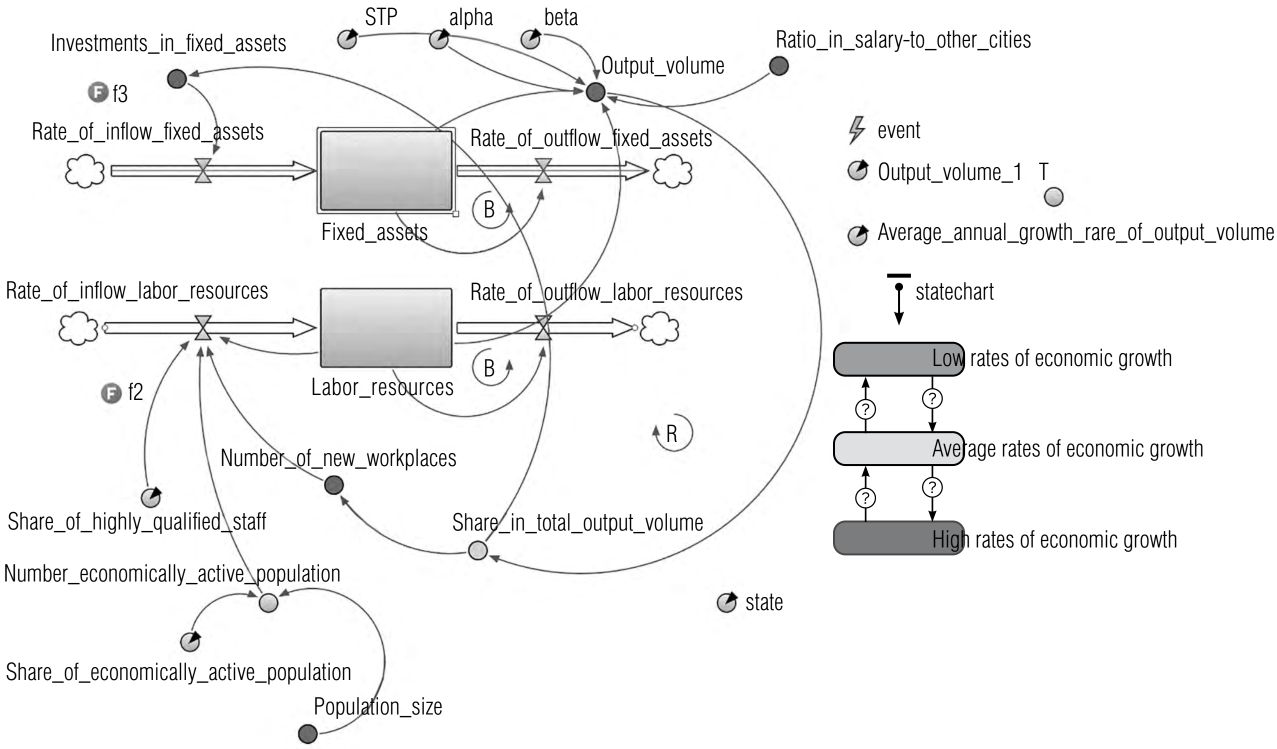

Fig.4. Model of behavior of enterprise agents in AnyLogic.

Figure 4 shows the implementation of the behavior model of enterprise agents (low-level model) in AnyLogic.

Figure 4 shows that to build this model, both the system dynamics and agent-based modeling methods are used (e.g., [5, 6, 13, 14, 16]). In particular, to model possible changes in the states of agents in accordance with the given rules, a state-flow chart is used (that is shown in *Fig. 4*). At the same time, to determine the values of the average annual growth rate of output volume, the value of which is compared with the given threshold values for the implementation of transitions between the states of each agent, the calculation uses the system dynamics model (the left part of *Fig. 4*). The model (5)–(11) uses the function of the Cobb-Douglas type to calculate the output volume, taking into account the values of control parameters. The results of simulation modeling, in particular, the state of each agent enterprise, the output volume, the amount of labor resources, etc. are automatically uploaded to the top-level model and used to simulate the "gravity effect" that determines the process of redistribution of population and labor resources (*Fig. 3*).

3. Results of optimization experiments

To improve the state of the entire ensemble of agent enterprises, in particular, to minimize the number of organizations with low rates of economic growth (less than 2% per year), the optimization experiments were carried out aimed at finding recommended values of control parameters of the model. As initial data, aggregated statistics on science cities were used, that is shared in [29], as well as the strategies for the socio-economic development of science cities published on the websites of city administrations. An example of initial data on agent enterprises for the Obninsk city is presented in *Table 1*.

Figure 5 presents a visualization of the states of agent enterprises obtained as a result of optimizing the values of control parameters of the simulation model ("Scenario based on recommended parameters") in comparison with previously approved investment and staff strategies ("Initial scenario (before optimization)".

The states of agent enterprises presented in *Fig. 5* correspond to a ten-year strategic planning horizon. A scenario for changes in such states based on recommended parameters (*Fig. 5*) was generated using a genetic optimization algorithm, aggregated by target functionality with a developed simulation model implemented in AnyLogic. *Table 2* presents the values of the control parameters of the model that correspond to the scenarios under consideration.

Table 1.

An example of initial dataset for agent enterprises of the Obninsk city (for 2016 year)

No	Company	Longitude	Latitude	Labor resources, people	Investments in fixed assets, thousand rubles	Fixed assets, thousand rubles	Issue volume, thousand rubles
1	JSC "SSC RF-IPPE"	55.089851	36.591257	2612	14782	594124	3199431
2	JSC ONPP "Technology"	55.105885	36.636379	2641	488089	6527163	5856985
3	JSC "SSC RF NIFHI"	55.063435	36.623261	528	103226	1300044	803209
4	MRNC named after A.F. Tsyba – the branch of the FSBI "NMRRC" of the Ministry of Health of the Russian Federation	55.111844	36.617534	1559	136132	2230550	2052357
5	ASP "Typhoon"	55.104454	36.609027	580	1551	80820	293792
6	VNIIRAE	55.106651	36.638687	270	3058	110706	242231
7	Public JST "Instrument Plant "Signal"	55.11013	36.59102	308	48217	996640	1061128
8	LLC "Hemofarm"	55.13496	36.6424	209	169412	2247154	1374894
9	JSC "Progress-ecology"	55.12122	36.58286	236	14758	696989	1709746
10	LLC "Nearmedic Pharma"	55.11442	36.61703	125	2810003	29994145	3506284

Obnnsk city	Initial scenario (before optimization)	Scenario based on recommended parameters
LLC "Hemofarm"	R	Y
JSC "Progress-ecology"	G	G
LLC "Nearmedic Pharma"	R	Y
MRNC named after A.F. Tsyba – the branch of the FSBI "NMRRC"	Y	Y
Public JST "Instrument Plant "Signal"	Y	G
JSC ONPP "Technology"	Y	Y
ASP "Typhoon"	G	G
JSC "SSC RF-IPPE"	G	G
VNIIRAE	G	G
JSC "SSC RF NIFHI"	R	Y

R – First stage (low growth rate)
Y – Second stage (average growth rate)
G – Third stage (high growth rate)

Fig.5. States of agent enterprise before and after optimization.

From *Table 2* it follows that in order to improve the conditions of agent enterprises in the science cities under study, it is necessary to significantly increase investments in fixed assets (mainly in Troitsk, Protvino and Pushchino), as well as to multiply the number of workplaces that should be created. At the same time, it is recommended to increase the level of wages in Obninsk and Pushchino in order to prevent the outflow of staff to neighboring cities (in particular, Troitsk). As a result, a fundamental improvement in the conditions of agent enterprises will be ensured (see *Fig. 5*), most of which will move to the third state of high economic growth rates.

Conclusion

This paper presents a new agent-based simulation model for the development of individual research-and-production clusters in Russia with implementation in the AnyLogic system. The model

Table 2.

Values of model control parameters

Control parameters	Science city	Basic scenario (before optimization)	Scenario based on recommended settings
Investments in fixed assets (million rubles/year)	Obninsk	3320	4341
	Troitsk	453	3108
	Protvino	246	2252
	Pushchino	374	2100
Number of new jobs (units/year)	Obninsk	1000	3020
	Troitsk	500	2550
	Protvino	300	1551
	Pushchino	200	1121
Average monthly salary (rubles/month)	Obninsk	62 000	72 000
	Troitsk	66 000	66 000
	Protvino	60 000	60 000
	Pushchino	45 000	55 000

developed consists of two levels (*Fig. 1*). The top-level model implements the environment in which agent enterprises belonging to various science cities operate, between which flow interaction is realized with a corresponding redistribution of the population and labor resources. The lower-level model provides the implementation of the logic of behavior of enterprise agents, the characteristics of which are described using system dynamics methods. An important optimization problem has been formulated to minimize the total number of enterprises characterized by low rates of economic growth, taking into account investment and operating expenses. The proposed two-level simulation model is implemented in the form of a software package (using the example of research-and-production clusters of Obninsk, Troitsk, Protvino and Pushchino), intended primarily for municipal and regional authorities implementing the strategy of innovative development. The use of such a system made it possible to formulate recommended values of control parameters (such as investments in fixed assets, the number of new jobs, etc.), providing a significant improvement in the conditions of agent enterprises located in the science cities under study over a ten-year strategic planning horizon (*Fig. 5* and *Table 2*). The economic-mathematical and computer toolkit developed is intended primarily for managing research-and-production clusters implementing the strategy of innovative development.

Further research will be aimed at studying the problems associated with overcoming the shortage of highly qualified labor, as well as the creation of detailed agent-based models for the development of research-and-production clusters. ■

References

1. Forrester J.W. (1969) *Urban dynamics*. M.I.T. Press, Cambridge.
2. Li G., Kou C., Wang Y., Yang H. (2020) System dynamics modelling for improving urban resilience in Beijing, China. *Resources, Conservation and Recycling*, vol. 161, 104954. https://doi.org/10.1016/j.resconrec.2020.104954
3. Diemer A., Nedelciu C.E. (2020) System dynamics for sustainable urban planning. *Sustainable Cities and Communities. Encyclopedia of the UN Sustainable Development Goals* (eds. W. Leal Filho, A. Marisa Azul, L. Brandli, P. Gökçin Özuyar, T. Wall). Springer, Cham, pp. 760–773. https://doi.org/10.1007/978-3-319-95717-3_115
4. Armenia S., Barnabè F., Pompei A., Scolozzi R. (2021) System dynamics modelling for urban sustainability. *Urban Sustainability. Springer Texts in Business and Economics* (eds. J. Papathanasiou, G. Tsaples, A. Blouchoutzi). Springer, Cham, pp. 131–173. https://doi.org/10.1007/978-3-030-67016-0_4
5. Makarov V., Bakhtizin A., Epstein J. (2022) Agent-based modeling for a complex world. Part 1. *Economics and the Mathematical Methods*, vol. 58, no. 1, pp. 5–26 (in Russian). https://doi.org/10.31857/S042473880018970-6
6. Makarov V., Bakhtizin A., Epstein J., (2022) Agent-based modeling for a complex world. Part 2. *Economics and the Mathematical Methods*, vol. 58, no. 2, pp. 7–21 (in Russian). https://doi.org/10.31857/S042473880020009-8
7. Makarov V., Bakhtizin A., Beklaryan G., Akopov A., Rovenskaya E., Strelkovskiy N. (2019) Aggregated agent-based simulation model of migration flows of the European Union countries. *Economics and the Mathematical Methods*, vol. 55, no. 1, pp. 3–15 (in Russian). https://doi.org/10.31857/S042473880004044-7
8. Makarov V., Bakhtizin A., Beklaryan G., Akopov A. (2020) Agent-based modelling of population dynamics of two interacting social communities: migrants and natives. *Economics and the Mathematical Methods*, vol. 56, no. 2, pp. 5–19 (in Russian). https://doi.org/10.31857/S042473880009217-7

9. 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

10. Makarov V.L., Bakhtizin A.R., Beklaryan G.L., Akopov A.S. (2019) Development of software framework for large-scale agent-based modeling of complex social systems. *Programmnaya Ingeneria (Software Engineering)*, vol. 10, no. 4, pp. 167–177 (in Russian). https://doi.org/10.17587/prin.10.167-177

11. Akopov A.S. (2012) Designing of integrated system-dynamics models for an oil company. *International Journal of Computer Applications in Technology,* vol. 45, no. 4, pp. 220–230. https://doi.org/10.1504/IJCAT.2012.051122

12. Akopov A.S. (2014) Parallel genetic algorithm with fading selection. *International Journal of Computer Applications in Technology*, vol. 49, no. 3–4, pp. 325–331. https://doi.org/10.1504/IJCAT.2014.062368

13. Akopov A.S., Beklaryan L.A., Saghatelyan A.K. (2019) Agent-based modelling of interactions between air pollutants and greenery using a case study of Yerevan, Armenia. *Environmental Modelling & Software*, vol. 116, pp. 7–25. https://doi.org/10.1016/j.envsoft.2019.02.003

14. Akopov A.S., Beklaryan L.A., Saghatelyan A.K. (2017) Agent-based modelling for ecological economics: A case study of the republic of Armenia. *Ecological Modelling,* vol. 346, pp. 99–118. https://doi.org/10.1016/j.ecolmodel.2016.11.012

15. Kislitsyn E.V., Gogulin V.V. (2021) Simulation of the environmental situation in a megalopolis. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve (Models, systems, networks in economics, technology, nature and society)*, vol. 1, no. 37, pp. 92–106 (in Russian). https://doi.org/10.21685/2227-8486-2021-1-8

16. Akopov A.S., Beklaryan L.A. (2015) An agent model of crowd behavior in emergencies. *Automation and Remote Control*, vol. 76, no. 10, pp. 1817–1827. https://doi.org/10.1134/S0005117915100094

17. Crooks A., Heppenstall A., Malleson N., Manley E. (2021) Agent-based modeling and the city: A gallery of applications. *Urban Informatics. The Urban Book Series* (eds. W. Shi, M.F. Goodchild, M. Batty, M.P. Kwan, A. Zhang). Springer, Singapore, pp. 885–910. https://doi.org/10.1007/978-981-15-8983-6_46

18. Chen L. (2012) Agent-based modeling in urban and architectural research: A brief literature review. *Frontiers of Architectural Research,* vol. 1, no. 2, pp. 166–177. https://doi.org/10.1016/j.foar.2012.03.003

19. Tian G., Qiao Z. (2014) Modeling urban expansion policy scenarios using an agent-based approach for Guangzhou Metropolitan Region of China. *Ecology and Society*, vol. 19, no. 3, art. 52. https://doi.org/10.5751/ES-06909-190352

20. Yun T.-S., Kim D., Moon I.-C., Bae J.W. (2022) Agent-based model for urban administration: A case study of bridge construction and its traffic dispersion effect. *Journal of Artificial Societies and Social Simulation*, vol. 25, no. 4, art. 5. https://doi.org/10.18564/jasss.4923

21. Ghandar A., Theodoropoulos G., Zhong M., Zhen B., Chen S., Gong Y., Ahmed A. (2019) An agent-based modelling framework for urban agriculture. *2019 Winter Simulation Conference (WSC)*, National Harbor, MD, USA, pp. 1767–1778. https://doi.org/10.1109/WSC40007.2019.9004854

22. Akopov A.S. (2012) System dynamics modeling of banking group strategy. *Business Informatics*, vol. 2, no. 20, pp. 10–19 (in Russian).

23. Stewart Q.J. (1950) The development of social physics. *American Journal of Physics*, vol. 18, pp. 239–253. https://doi.org/10.1119/1.1932559

24. Yap Y.L. (1977) The attraction of cities: A review of the migration literature. *Journal of Development Economics,* vol. 4, no. 3, pp. 239–264. https://doi.org/10.1016/0304-3878(77)90030-x

25. Kleiner G.B. (1986) *Production functions. Theory, methods, application*. Moscow: Finance and Statistics (in Russian).

26. Suvorov N.V., Akhunov R.R., Gubarev R.V., Dzyuba E.I., Fayzullin F.S. (2020) Applying the Cobb–Douglas production function for analysing the region's industry. *Economy of Regions*, vol. 16, no. 1. pp. 187–200 (in Russian). https://doi.org/10.17059/2020-1-14

27. Hellerstein J.K., Neumark D. (2007) Production function and wage equation estimation with heterogeneous labor: Evidence from a new matched employer-employee data set. *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches* (eds. Ernst R. Berndt and Charles R. Hulten). Chicago: University of Chicago Press, pp. 31–72. Available at: https://www.nber.org/system/files/chapters/c0873/c0873.pdf (accessed 21 February 2024).

28. Beklaryan G.L., Akopov A.S., Khachatryan N.K. (2019) Optimisation of system dynamics models using a real-coded genetic algorithm with fuzzy control. *Cybernetics and Information Technologies,* vol. 19, no. 2, pp. 87–103. https://doi.org/10.2478/cait-2019-0017

29. *Guide on the results of the analysis of the compliance of the indicators of scientific and industrial complexes of science cities of the Russian Federation with the requirements established by paragraph 8 of Article 2.1 of Federal Law No. 70-FZ "On the status of the science city of the Russian Federation", and the achievement of the results provided for by the action plans for the implementation of the strategies of socio-economic development of science cities of the Russian Federation in 2022* (in Russian). Available at: https://minobrnauki.gov.ru/upload/iblock/77c/cemgzf9g61hhktvme7dfmm9feddbfzvv.pdf (accessed 21 February 2024).

About the author

Gayane L. Beklaryan

Cand. Sci. (Econ.);

Senior Researcher, Department of Theoretical Economics and Mathematics Research, Laboratory of Computer Modeling of Social and Economic Processes, Central Economics and Mathematics Institute, Russian Academy of Sciences, 47, Nakhimovsky Ave., Moscow 117418, Russia;

E-mail: glbeklaryan@gmail.com

ORCID: 0000-0002-1286-0345

DOI: 10.17323/2587-814X.2024.1.52.64

Nonparametric procedure for comparing the performance of divisions of a network organization

Petr A. Koldanov [a]
E-mail: pkoldanov@hse.ru

Vladimir A. Koldanov [b]
E-mail: vlad.kold@gmail.com

[a] Laboratory of Algorithms and Technologies for Network Analysis, HSE University
Address: 136, Rodionova St., Nizhny Novgorod 603093, Russia

[b] Nizhny Novgorod State University of Engineering and Economics
Address: 22A, building 2, Oktyabrskaya St., Nizhny Novgorod Region, Knyaginino 606340, Russia

Abstract

To solve the problem of comparative efficiency analysis of branch operations for a small volume of randomly observed data, a non-parametric approach is relevant, since it does not require a probabilistic model of observations. Comparing the results of the non-parametric approach with the results obtained within the traditionally used Gaussian model is also relevant. Additionally, obtaining a consistent comparison of a group (of no less than three) branches is important. Currently, the non-parametric approach and the corresponding comparison with the known results of solving the problem considered in this work obtained within the framework of the normal model are absent. In addition, insufficient attention is paid to the search for methods of obtaining consistent solutions. This work to some extent fills these gaps. This work uses non-parametric statistical methods and theory of simultaneous hypothesis testing to address these problems. This paper proposes a procedure for comparative analysis of the efficiency of several units within a network organization with a small volume of observations based on the Mann–Whitney tests. We carry out a comparison of the results obtained from the proposed non-parametric procedure with results based on extensions of Student's *t*-tests. We propose a method for reducing the number of compatibility problems based on the search for an appropriate significance level. We provide an example of a fully consistent comparison of the efficiency of branch operations.

Keywords: network organization, efficiency of branch operations, Mann–Whitney tests, incompatibility problem

Citation: Koldanov P.A., Koldanov V.A. (2024) Nonparametric procedure for comparing the performance of divisions of a network organization. *Business Informatics*, vol. 18, no. 1, pp. 52–64.
DOI: 10.17323/2587-814X.2024.1.52.64

Introduction

Various aspects of comparing the effectiveness of organizations are discussed in numerous scientific papers, for example [1–3]. As a rule, comparisons are based on many indicators. Obviously, the success of such comparisons depends on how adequately and qualitatively comparisons on individual indicators can be made, especially if these indicators are of a random nature. In this latter case, the methods of mathematical statistics are typically used. Such methods are divided into:

- parametric, relying on a specific probabilistic model of the analyzed indicators. In this case, the normal distribution is most often used as a probabilistic model [4, 5],
- non-parametric, free from a detailed probabilistic model, and sometimes from the assumption of the random nature of the analyzed data as well [3, 6–9].

Many tasks, including the one this paper is aimed at, can be considered within the framework of both approaches. In this case, it becomes necessary to compare the conclusions based on parametric model procedures to the conclusions of nonparametric procedures. Such a comparison is one of the goals of this work.

There are many known results comparing parametric and nonparametric tests of hypothesis checking against an alternative. With a finite number of observations, such a comparison is made on the basis of the analysis of the test power function determined by the probability of rejecting the hypothesis. With an unlimited number of observations, the comparison of hypothesis checking tests against an alternative is based on the calculation of asymptotic efficiency indicators [10]. The specificity of the problem considered in this work lies in the need to select one of many solutions based on a small number of observations.

This paper proposes a procedure for comparative analysis of the performance of divisions of a network organization. The results of applying such a procedure can be used to make informed management decisions by the managers of a network organization. In this case, the efficiency of a division is understood as the ratio of the number of sales of a certain product (for example, the number of cars) to the number of potential buyers. A network organization is understood as a set of units operating according to a common scheme. Examples of such organizations are a network of branches of a large automobile company or a chain of Pyaterochka stores, etc.

In this paper, as an illustrative example, we consider the problem of supporting management decisions by the managers of a network of branches of a large university. Everyone is welcome to attend the preparatory courses of such branches. A natural characteristic of the efficiency of the staff of such branches is the ratio of the number of students attending preparatory courses to the number of potential applicants. Information on comparative effectiveness is the basis for making strategic decisions on the development of the branch network.

It is convenient to present a network organization in the form of a graph. The vertices of this graph correspond to divisions. The specificity of the graph we are considering is that it can have both directed and undirected edges. An undirected edge between vertices i, j is added to the graph if and only if it is decided that the i-th and j-th divisions work equally efficiently. A directed edge from vertex i to vertex j is added to the graph if and only if it is settled that the i-th unit is more efficient than the j-th unit. Note that usually either only directed edges or only undirected ones are used in graphs [11–15]. We will use them both simply because they better allow us to emphasize some of the structures of the graph under analysis, which characterize the specifics of the analyzed network organization. Among such structures there will be cliques [16] (a set of vertices, any two of which are connected by a non-directed edge), which characterize a set of divisions working equally efficiently. In what follows, we will call such cliques undifferentiated classes. Another example of a structure is a complete subgraph with only directed edges. Such a structure will be called a structure of ordering or dominance.

In many problems, in particular in the example we are analyzing, it is natural to consider the number of sales as a random variable. At the same time, analysis of real data, especially when there is little of it, can lead to contradictory conclusions. In this case, the corresponding graph contains some logically contradictory structures, for example, subgraphs of three vertices, two edges between which are undirected, and one is directed. This type of problem arose in [17] when discussing the problem of testing hypotheses of homogeneity of at least three populations and was called the incompatibility problem. Applied tasks in which the incompatibility problem arises (the problem of inconsistently combining the results of comparing effectiveness of the pairs of departments) were considered in [4, 5] within the framework of the normal model. The solution to the incompatibility problem was based on the introduction of an additional parameter Δ and the transition to tasks of comparing the effectiveness of two divisions with accuracy Δ. Moreover, if the efficiencies of divisions i, j differed by less than Δ, then it was decided that their effectiveness was the same (with an accuracy of Δ). This technique allows us to solve the incompatibility problem, but leads to an additional problem of choosing Δ. Another goal of this work is to find ways to solve the incompatibility problem without introducing an auxiliary parameter Δ.

In this work, unlike [4, 5], the assumption of a normal distribution of the number of sales is not used. Pairwise comparison of the effectiveness of two units is based on the use of Mann–Whitney tests. The procedure for comparative analysis of divisions by efficiency is based on a combination of nonparametric tests of pairwise comparison of two divisions. This uses a graphical representation, which is convenient for visualizing emerging incompatibility problems. The proposed nonparametric procedure is applied to the analysis of data reported in [4], and an example is given in which the inconsistency problem is overcome by analyzing p-values and appropriately selecting significance levels for pairwise comparison tests. A comparison is made with the results obtained within the normal model.

This article is organized as follows: Section 1 provides the basic notation and formulation of the problem; Section 2 describes the nonparametric procedure for comparative analysis of departments by efficiency, and its graphical representation; Section 3 provides an illustrative example, an example of solving the incompatibility problem, and compares it with the results obtained in [4].

1. Formulation of the problem

It is convenient to present data on the number of sales in the form of a matrix $\| x_{ji} \|$, where $\| x_{ji} \|$ – the ratio of the number of sales to the number of potential buyers in division j in the i-th time period, $j = 1, \ldots, N$, where N is the number of divisions of the network organization, $i = 1, \ldots, m_j$, where m_j is the number of analyzed time periods of work j-th division. We will assume that observations x_{ji} represent the values of random variables X_{ji}, which describe the ratio of the number of sales to the number of potential buyers in division j in time period i. Let us assume that all time periods are the same, and the random variables X_{ji} are independent for all $j = 1, \ldots, N$; $i = 1, \ldots, m_j$ and for a fixed j are equally distributed as X_j. Let $F_j(x)$ be the distribution function of the random variable X_j.

The problem considered in this work consists in constructing and applying for the analysis of specific data a statistical procedure for distinguishing hypotheses of the form:

$$\begin{aligned} &H_1 : F_1(x) = F_2(x) = \cdots = F_N(x), \quad \forall x \\ &H_2 : F_1(x) < F_2(x) = \cdots = F_N(x), \quad \forall x \\ &\vdots \\ &H_L : F_1(x) < F_2(x) < \cdots < F_N(x), \quad \forall x \end{aligned} \qquad (1)$$

Here, hypothesis H_1 means that the efficiency of all divisions is the same, hypothesis H_2 means that division 1 is more efficient than other divisions whose efficiency is the same, etc. Note that relations (1) do not describe all possible relations between the distribution functions $F_j(x)$, $j = 1, \ldots, N$. We limit ourselves to considering only these hypotheses, since we are only interested in the presence of a systematic shift, which can result from different performance levels of different divisions of a network organization.

Like in [4], we will use the method of constructing procedures with many solutions proposed in [17]. This method is based on reducing a multi-alternative problem to a set of appropriately selected two-alternative generating problems. In our case, to distinguish (1), it is natural to consider two-alternative hypothesis testing problems $h_{ij} : F_{ij}(x) \geq F(x), \forall x, \forall i, j = 1, \ldots, N$.

For fixed i, j combining both the tests φ_{ij}, φ_{ji} of simultaneous testing the hypotheses h_{ij} and h_{ji} with non-zero probability can lead to a logically untenable (for a given x) decision to reject both hypotheses, i.e. to the incompatibility problem. As shown in [17], to eliminate such a contradiction, it is sufficient to require that the significance levels α_{ij}, α_{ji} of the tests φ_{ij}, φ_{ji} satisfy the condition $\alpha_{ij} + \alpha_{ji} < 1$. In this case, such a combination of the tests φ_{ij}, φ_{ji} leads to a joint procedure for distinguishing three hypotheses:

$$\begin{aligned} &H_{ij}^1 : F_i(x) < F_j(x); \\ &H_{ij}^2 : F_i(x) = F_j(x); \\ &H_{ij}^3 : F_i(x) > F_j(x). \end{aligned} \qquad (2)$$

However, combining such procedures with three solutions for different i, j can lead to a contradiction, namely: with non-zero probability it can be decided that (for example):

$$F_1(x) = F_2(x) \text{ and } F_2(x) = F_3(x), \text{ but } F_1(x) \neq F_3(x).$$

To eliminate this contradiction, in [4], following the proposal of [17], a slightly modified system of generating hypotheses was considered. However, the studies were limited to the case when $F_j(x)$ is a normal distribution. In the notation of this work, the modified system of generating hypotheses has the form:

$$h'_{ij} : F_{ij}(x) + \Delta \geq F(x), \forall x, \forall i, j = 1, \ldots, N.$$

When combining tests φ'_{ij}, φ'_{jii} for simultaneous testing of hypotheses h'_{ij}, h'_{ji} we obtain a procedure for distinguishing three hypotheses:

$$\begin{aligned} &H_{ij}'^1 : F_i(x) + \Delta < F_j(x); \\ &H_{ij}'^2 : |F_i(x) - F_j(x)| < \Delta; \\ &H_{ij}'^3 : F_i(x) > F_j(x) + \Delta. \end{aligned} \qquad (3)$$

In this case, the problem of obtaining contradictory conclusions does not arise. At the same time, the introduction of the Δ parameter formally changes the original problem.

In this work, the Δ parameter is not introduced and the assumption of a normal distribution is not made.

At the same time, one of the interesting questions is to find options for consistently combining statistical rules with three solutions without introducing the Δ parameter.

Note that at present, in the intensively developing theory of simultaneous testing of many hypotheses, no emphasis is placed on the need to solve the problem of incompatibility [18–21]. Moreover, starting from [22], the problem of incompatibility is considered as too strong a requirement imposed on the procedure for simultaneous testing of many hypotheses. Within the framework of the theory of simultaneous testing of many hypotheses, approaches to the construction of procedures that control the probability of at least one error of the first type, the proportion of errors of the first type, and some others are mainly studied. In this work, on the contrary, we focus on solving the incompatibility problem, which allows for an appropriate comparison with the results obtained in [4].

2. Nonparametric comparative analysis procedure and its visualization

2.1. Procedure with three solutions

One of the most effective nonparametric procedures for distinguishing hypotheses (2) is based on the Mann–Whitney statistics [6, 9]. The Mann–Whitney statistic looks like this:

$$W_{ij}(x_i, x_j) = \sum_{r=1}^{m_j}\sum_{s=1}^{m_i} I(x_{is} < x_{jr}), \tag{4}$$

where $I(A)$ is the indicator of event A.

For fixed i, j the procedure with three solutions for distinguishing hypotheses (2) in terms of p-values can be written as

$$\varphi(i,j) = \begin{cases} d_{ij}^1 & p_{ij}^1 \le \alpha_1 \\ d_{ij}^3 & p_{ij}^3 \le \alpha_3 \\ d_{ij}^2 & p_{ij}^1 > \alpha_1,\ p_{ij}^3 > \alpha_3\,, \end{cases} \tag{5}$$

where d_{ij}^k is the decision to accept the hypothesis H_{ij}^k $(k = 1, 2, 3)$;

p_{ij}^1, p_{ij}^3 are the corresponding p-values, namely:

$$p_{ij}^1 = P_{F_i = F_j}\left(W_{ij}(X_i, X_j) \le W_{ij}(x_i, x_j)\right)$$
$$p_{ij}^3 = P_{F_i = F_j}\left(W_{ij}(X_i, X_j) > W_{ij}(x_i, x_j)\right), \tag{6}$$

where α_1, α – significance levels of tests for hypothesis testing H_{ij}^1 and H_{ij}^3, respectively.

It is assumed that among the observed values and there are no equals. The necessary adjustments in the case of equal observations can be made based on the methodology outlined in [9].

Tables of distribution of statistics (4) for small m_i, m_j are given in [9]. For large m_i, m_j, one can use the normal distribution

$$N\left(\frac{m_i \cdot m_j}{2}, \frac{m_i \cdot m_j \cdot (m_i + m_j + 1)}{12}\right),$$

which is recommended to be used when

$$\min(m_i, m_j) > 50 \text{ [9]}.$$

From (6) it is obvious that for fixed i, j, $p_{ij}^1 + p_{ij}^3 = 1$. Therefore, to apply the procedure with three solutions (5), information about the minimum p-value is sufficient:

$$p_{ij} = \min(p_{ij}^1, p_{ij}^3). \tag{7}$$

2.2. Procedure with many solutions and its graphical representation

We will obtain the procedure for distinguishing hypotheses (1) by combining procedures (5). This procedure can be written as:

$$\delta = \left(\varphi(1,2), \varphi(2,3), \ldots, \varphi(N-1, N)\right). \tag{8}$$

Let denote the efficiency of the i-th department. In the problem under consideration, two types of relationships are possible between the performance of divisions of a network organization (dominance or equivalence). The entry means that the i-th unit works more efficiently (dominates) than the j-th unit. The entry means that the i-th and j-th divisions work equally efficiently (equivalence). To visually analyze the results of applying procedure (8), we will use the technique proposed in [23–25].

For a given vector $(f_1, ..., f_N)$, we introduce matrix D with elements

$$d_{ij} = \begin{cases} 1, & f_i > f_j \\ 0, & f_i = f_j \\ -1, & f_i < f_j \end{cases}$$

and a matrix B with elements

$$b_{ij} = \begin{cases} 1, & f_i > f_j \\ 0, & \text{otherwise}. \end{cases}$$

It is easy to show [24] that the matrix B is related to the matrix D by the relation

$$D = B - B^T,$$

where B^T is the transposed matrix B.

The relationships described by the matrix B are more easily interpreted if the rows and columns of the matrix B are rearranged in such a way as to obtain an upper triangular shape (i.e., to collect, if possible, all the ones above the main diagonal of the matrix B). To obtain the upper triangular form of the matrix B, one can arrange in descending order the rows (and columns) of the matrix D by the sums of the row elements. Replacing the -1 elements of the resulting matrix with 0, we obtain matrix B, which is most consistent with the upper triangular shape. Matrix B, which is most consistent with the upper triangular form, allows us to identify the so-called "undifferentiated classes" [25]. The term "undifferentiated class" will be used to designate the largest set of units whose performance levels are not significantly different from each other. The term "largest" means that for any unit i that does not belong to a given undifferentiated class, there is at least one unit j from that class such that the performance efficiencies of units i and j are meaningfully distinguishable. The matrix B, which is most consistent with the upper triangular form, shows the undifferentiated classes as square sub-matrices, symmetric about the main diagonal, with all elements equal to 0.

Obviously, the overlapping undifferentiated classes resulting from the delta procedure mean that there is an incompatibility problem. It is convenient to visualize the matrix D in the form of a graph $G = (V, E)$, where $V = \{1, 2, ..., N\}$ is the set of vertices of the graph, $E = \{e_{ij}\}$ is the set of edges of the graph. If element d_{ij} of matrix D is equal to 1, then a directed edge from vertex i to vertex j is added to the graph G. In this case, vertex i dominates vertex j. In [11], vertex i is called the parent of vertex j, and vertex j is called the child of vertex i. If vertex i is connected by a directed path of length greater than 1 to some vertex k, then vertex i is called the ancestor of vertex k, and vertex k is called a descendant of vertex i. If element d_{ij} of matrix D is equal to 0, then an undirected edge between vertices i and j is added to graph G. If the element $d_{ij} = -1$, then, since $d_{ij} = -d_{ji}$, the graph G already contains a directed edge from vertex j to vertex i. It is obvious that all the vertices of this graph, corresponding to divisions from a certain undifferentiated class, are connected to each other by undirected edges and therefore represent cliques of the graph G. Below we will separately depict sub-graphs with only directed edges and only with undirected edges.

Note that the proposed representation will clearly reflect the existence of problems of incompatibility of the obtained conclusions, if any. Obviously, if in a representation different undifferentiated classes contain the same vertices, then the incompatibility problem occurs.

3. Illustrative example

Let us consider the task of comparative analysis of the performance of university branches, which was briefly described in the introduction. Let us denote $1f$ as the first branch of the university, $2f$ – the second branch of the university, etc. Data for analysis are borrowed from [4] and are shown i*n Table 1.*

Minimum p-values (7) of tests (5) are given in *Table 2.*

3.1 Construction of undifferentiated classes

Let us first consider the traditional significance level. Matrix $D_{0.05}$ is shown in *Table 3.*

Table 1.

Data on the number of students attending preparatory courses in various branches

$1f$	$2f$	$3f$	$4f$	$5f$	$6f$	$7f$	$8f$
103	131	187	154				
92	212	262	92	151	99	235	
122	197	376	129	164	268	338	77
48	143	283	146	141	217	239	63
86	95	231	125	140	231	187	59
89	70	203	127	173	175	123	78
147	92	276	183	141	137	139	82
134	95	258	213	187	242	185	28
Number of potential applicants in the i-th branch							
6390	7090	28900	6320	6320	11130	4660	2530

Table 2.

Minimum *p*-values

	$1f$	$2f$	$3f$	$4f$	$5f$	$6f$	$7f$
$2f$	p^1_{12} 0.4392						
$3f$	p^3_{13} 0.0023	p^3_{23} 0.0103					
$4f$	p^1_{14} 0.0364	p^1_{24} 0.0652	p^1_{34} 0.0001				
$5f$	p^1_{15} 0.0006	p^1_{25} 0.0469	p^1_{35} 0.0002	p^1_{45} 0.1678			
$6f$	p^1_{16} 0.3063	p^1_{26} 0.4775	p^1_{36} 0.0070	p^3_{46} 0.0760	p^3_{56} 0.0020		
$7f$	p^1_{17} 0.0002	p^1_{27} 0.0011	p^1_{37} 0.0002	p^1_{47} 0.0200	p^1_{57} 0.0012	p^1_{67} 0.0003	
$8f$	p^3_{18} 0.0147	p^3_{28} 0.0539	p^1_{38} 0.0007	p^1_{48} 0.1725	p^1_{58} 0.1830	p^1_{68} 0.0256	p^3_{78} 0.0175

Table 3.

Matrix $D_{0.05}$

	1*f*	2*f*	3*f*	4*f*	5*f*	6*f*	7*f*	8*f*	Σ
1*f*	–	0	1	–1	–1	0	–1	–1	–3
2*f*	0	–	1	0	–1	0	–1	0	–1
3*f*	–1	–1	–	–1	–1	–1	–1	–1	–7
4*f*	1	0	1	–	0	0	–1	0	1
5*f*	1	1	1	0	–	1	–1	0	3
6*f*	0	0	1	0	–1	–	–1	–1	–2
7*f*	1	1	1	1	1	1	–	1	7
8*f*	1	0	1	0	0	1	–1	–	2

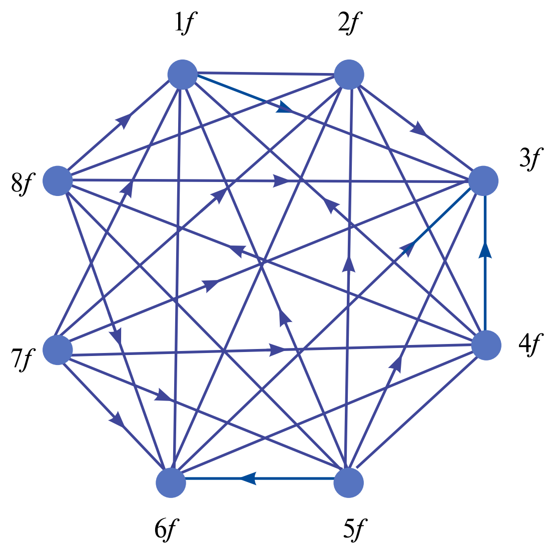

Fig. 1. Graphical representation of the matrix $D_{0.05}$.

A graphical representation of the $D_{0.05}$ matrix is shown in *Fig. 1.*

Matrix $B_{0.05}$ obtained from matrix $D_{0.05}$ reduced to upper triangular form is shown in *Table 4.*

Table 4.

Matrix $B_{0.05}$, educed to upper triangular form

	7*f*	5*f*	8*f*	4*f*	2*f*	6*f*	1*f*	3*f*
7*f*	–	1	1	1	1	1	1	1
5*f*		–	0	0	1	1	1	1
8*f*			–	0	0	1	1	1
4*f*				–	0	0	1	1
2*f*					–	0	0	1
6*f*						–	0	1
1*f*							–	1
3*f*								–

A graphical representation of the matrix $B_{0.05}$ reduced to the upper triangular form is shown in *Fig. 2.*

In *Fig. 2* it is easy to see that there are 6 cliques in this graph: {7}, {4, 5, 8}, {2, 4, 8}, {2, 4, 6}, {2, 6, 1}, {3}. Note that these cliques have common vertices, for example, vertex 4 belongs to the 3rd cliques. This indicates that there is an incompatibility problem. Thus, for $\alpha_{ij} = 0.05$, $\forall i, j = 1, \ldots, 8$ disjoint undifferentiated classes are not distinguished.

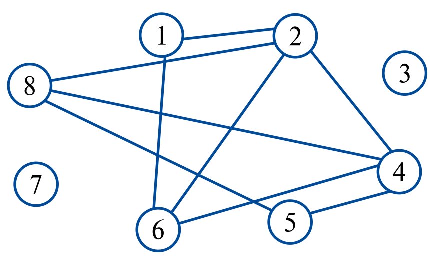

Fig. 2. Graphical representation of the matrix $B_{0.05}$ reduced to upper triangular form. The branch number is indicated by a number.

It is well known that there are no strict rules for choosing the significance level α. A possible way to reduce the number of incompatibility problems arising could be based on changing the significance level α. Analysis of the p-values given in *Table 2* suggests the advisability of constructing and studying graphical models when choosing α_{ij}, $\forall i, j = 1, ..., 8$ from the interval (p^3_{46} =0.076; p^1_{45} =0.1678). For definiteness, we choose $\alpha_{ij} = 0.1$, $\forall i, j = 1, ..., 8$. The matrix $D_{0.1}$ is given in *Table 5*.

Table 5.

Matrix $D_{0.1}$

	1*f*	2*f*	3*f*	4*f*	5*f*	6*f*	7*f*	8*f*	Σ
1*f*	–	0	1	–1	–1	0	–1	–1	–3
2*f*	0	–	1	–1	–1	0	–1	–1	–3
3*f*	–1	–1	–	–1	–1	–1	–1	–1	–7
4*f*	1	1	1	–	0	1	–1	0	3
5*f*	1	1	1	0	–	1	–1	0	3
6*f*	0	0	1	–1	–1	–	–1	–1	–3
7*f*	1	1	1	1	1	1	–	1	7
8*f*	1	1	1	0	0	1	–1	–	3

A graphical representation of the matrix $D_{0.1}$ is shown in *Fig. 3*.

The matrix $B_{0.1}$ obtained from the matrix $D_{0.1}$ reduced to upper triangular form is shown in *Table 6*.

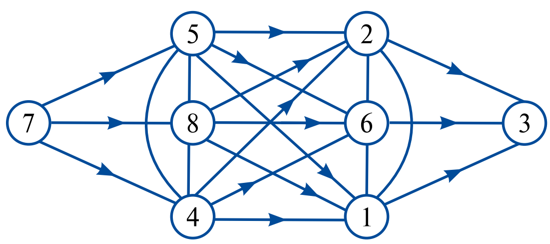

Fig. 3. Graphical representation of the matrix $D_{0.1}$.
The branch number is indicated by a number.

Table 6.

Matrix $B_{0.1}$ reduced to upper triangular form

	7*f*	5*f*	8*f*	4*f*	2*f*	6*f*	1*f*	3*f*
7*f*	–	1	1	1	1	1	1	1
5*f*		–	0	0	1	1	1	1
8*f*			–	0	1	1	1	1
4*f*				–	1	1	1	1
2*f*					–	0	0	1
6*f*						–	0	1
1*f*							–	1
3*f*								–

A graphical representation of the matrix $B_{0.1}$ reduced to the upper triangular form is shown in *Fig. 4*.

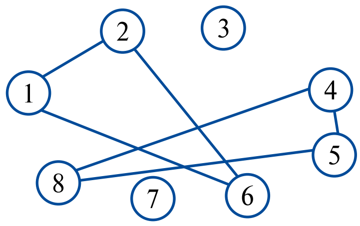

Fig. 4. Graphical representation of the matrix $B_{0.1}$ reduced to upper triangular form.
The branch number is indicated by a number.

In *Fig. 4* it is easy to see that in this graph there are 4 cliques {7}, {4, 5, 8}, {2, 6, 1}, {3} and these cliques do not have common vertices. This indicates that at the problem of incompatibility does not arise, i.e., the undifferentiated classes do not intersect.

3.2 Ordering construction of structures

A graphical representation of the matrix $D_{0.05}$ shows the structure of ordering branches according to the efficiency of their work, shown in *Fig. 5*.

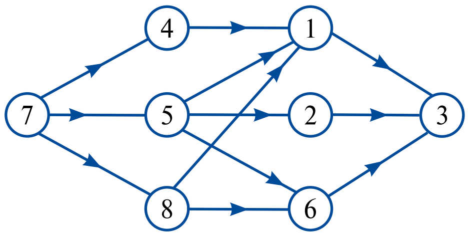

Fig. 5. Nonparametric ordering $\alpha = 0.05$.
The branch number is indicated by a number.

Only dominance relationships are shown; equivalence relationships and ancestor-descendant relationships are not shown. In particular, there is no directed edge between vertices 7 and 3, because there is a directed path $7 \to 4 \to 1 \to 3$ from the ancestor "7*f*" to the descendant "3*f*", meaning strict ordering: 7*f* is more effective than 4*f*, 4*f* is more effective than 1*f*, 1*f* is more effective than 3*f*. Note that there are no directed paths in this graph $7 \to 4 \to 2 \to 3$; $7 \to 4 \to 6 \to 3$; $7 \to 8 \to 2 \to 3$. This indicates the absence of complete ordering between the operating efficiencies of branches at $\alpha = 0.05$. Considering that {4*f*, 5*f*, 8*f*} belong to the same undifferentiated class, the absence of such paths leads to logical contradictions. We emphasize that at $\alpha = 0.1$ complete ordering takes place (see *Fig. 6*) and logical contradictions do not arise.

3.3. Comparison

In *Fig. 7* shows graphs constructed from the results of comparing the operating efficiencies of branches both under the assumption of a normal distribution of the random variables under study [4] and in a nonparametric formulation.

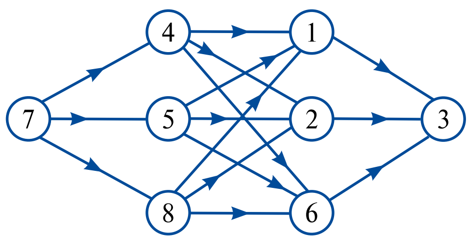

Fig. 6. Nonparametric ordering $\alpha = 0.1$.
The branch number is indicated by a number.

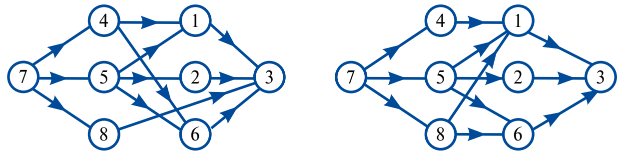

Fig. 7. Parametric (left),
nonparametric (right) ordering at α =0.05.
The branch number is indicated by a number.

The ordering graphs shown in *Fig. 7*, differ in three edges, namely: edge (4, 6) is present in parametric ordering, and absent in non-parametric ordering; edges (8, 1); (8, 6) is present with nonparametric ordering, and absent with parametric ordering.

Linear ordering, constructed according to the scheme proposed in [4], corresponding to nonparametric ordering (*Fig. 7*, right), has the form:

$$f_3 < f_1 \le f_6 \le f_2 \le f_4 \le f_8 \le f_5 < f_7 \text{ with precision } \Delta. \quad (9)$$

Obtaining such a linear ordering, formally proposed in [4] in a slightly different formulation (ordering with accuracy Δ), is based on analyzing the number of directed links leaving a specific vertex or entering a specific vertex. Namely, since vertices 4, 5, 8 (*Fig. 7*, right) are not connected by directed edges, at first glance it seems that the solution $f_4 = f_8 = f_5$ can be made with an accuracy of Δ. However, since vertex 4 dominates only one vertex 1 (vertex 4 has one directed edge going out), vertex 8 dominates two vertices 1 and 6 (vertex 8 has two directed edges going out), and vertex 5 dominates three vertices 1, 2, 6 (three directed edges emerge from vertex 5), then the solution $f_4 \le f_8 \le f_5$ is obtained with an accuracy of Δ. Writing $f_4 \le f_8$ with precision Δ means that $f_4 + \Delta < f_8$ or $|f_4 - f_8| < \Delta$. Similarly, since vertices 1, 2, 6 (*Fig. 7*, right) are not connected by directed edges, at first glance it seems that a decision $f_4 \le f_8 \le f_5$ with an accuracy of Δ can be made. However, since vertex 1 is a descendant of all vertices 4, 5, 8 (vertex 1 includes three directed edges), vertex 6 is a descendant of two vertices 5 and 8 (vertex 6 includes two directed edges), vertex 2 is a descendant of one vertex 5 (vertex 2 includes one directed edge), then the solution $f_1 \le f_6 \le f_2$ is made with accuracy Δ.

The linear ordering obtained in [4] has the form:

$$f_3 < f_1 = f_6 \le f_2 \le f_8 \le f_4 \le f_5 \le f_7 \text{ with precision } \Delta. \quad (10)$$

Orderings (9) and (10) differ very slightly. In fact, the extreme elements of the orderings coincide, the sign of $f_1 \le f_6$ has changed compared to the sign of $f_1 = f_6$, and the non-strict ordering $f_4 \le f_8$ has changed to the ordering $f_8 \le f_4$. But in both cases, no significant difference in the operating efficiency of branches f_8 and f_4 was found. This indirectly indicates the acceptability of the normal model proposed in [4].

The undifferentiated classes presented in *Fig. 8* are distinguished by three edges, namely: the edge (4, 6) is present in the nonparametric construction of undifferentiated classes, and is absent in the parametric construction; edges (8, 1); (8, 6) is present in the parametric construction of undifferentiated classes, and is absent in the nonparametric construction. This is quite consistent with the comparison of ordering structures (see *Fig. 7*).

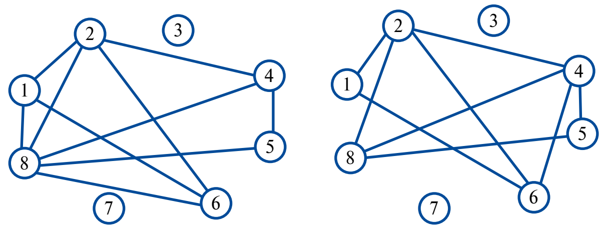

Fig. 8. Parametric (left) and nonparametric (right) undifferentiated classes at $\alpha = 0.05$. The branch number is indicated by a number.

Conclusion

In this work, a nonparametric procedure for comparative analysis of the performance of several divisions of a network organization based on a small volume of observations has been constructed. We give an example of the application of the proposed approach to a comparative analysis of the performance of university branches. The results of the comparative analysis obtained by the proposed nonparametric procedure are compared with the results obtained within the framework of the normal model [4]. It is shown that the results of nonparametric ordering without introducing an additional uncertainty parameter Δ and the ordering results obtained within the normal model with the introduction of Δ are quite close. An example of a completely consistent comparison of the performance of several divisions of a network organization is provided. ■

Acknowledgments

The results of sections 1 and 2 were prepared within the framework of the Fundamental Research Program of the National Research University "Higher School of Economics" (HSE University). The results of section 3 were prepared with the support of a grant from the Russian Science Foundation (project 22-11-00073).

References

1. Aleskerov F.T., Belousova V. (2007) Effective development of the branch network of a commercial bank. *Upravleniye v kreditnoy organizatsii (Management in a credit institution)*, no. 6, pp. 23–34 (in Russian).
2. Kryukov A.M. (2010) Efficiency analysis of branch and division activities is a prerequisite for business stability. *Strategic decisions and risk management*, vol. 4, pp. 84–87 (in Russian). https://doi.org/10.17747/2078-8886-2010-4-84-87
3. Myznikova M.A. (2022) Quality of strategic management under ambiguity: Assessment within the framework of sustainable development. *Business Informatics*, vol. 16, no. 3, pp. 36–52. https://doi.org/10.17323/2587-814X.2022.3.36.52
4. Koldanov A.P., Koldanov P.A. (2012) Multiple decision procedures for the analysis of higher school entry selection results. *Business Informatics (Biznes-informatika)*, vol. 19, no. 1, pp. 24–31 (in Russian).

5. Koldanov P.A. (2013) Efficiency analysis of branch network. *Models, Algorithms, and Technologies for Network Analysis* (eds. B.I. Goldengorin, V.A. Kalyagin, P.M. Pardalos). Springer Proceedings in Mathematics & Statistics, vol. 59, pp. 71–83. https://doi.org/10.1007/978-1-4614-8588-9_5
6. Hettmansperger T. (1987) *Statistical Inference Based on Ranks*. Moscow: Finance and Statistics (in Russian).
7. Kendall M. (1975) *Rank correlations*. Moscow: Statistics (in Russian).
8. Hollander M., Wolfe D. (1983) *Nonparametric statistical methods*. Moscow: Finance and Statistics (in Russian).
9. Lehmann E.L. (1975) *Nonparametrics: Statistical methods based on ranks*. San Francisco: Hoden-Day. https://doi.org/10.1002/zamm.19770570922
10. Kendall M., Stuart A. (1973) *Statistical Inferences and relationships*. Moscow: Science (in Russian).
11. Jordan M.I. (2004) Graphical models. *Statistical Science*, vol. 19, no. 1, pp. 140–155. https://doi.org/10.1214/088342304000000026
12. Newman M.E.J. (2010) *Networks. An introduction*. New York: Oxford University Press.
13. Opsahl T., Panzarasa P. (2009) Clustering in weighted networks. *Social Networks*, vol. 31, no. 2, pp. 155–163.
14. Horvath S. (2011) *Weighted network analysis. Applications in genomics and systems biology*. New York: Springer. https://doi.org/10.1007/978-1-4419-8819-5
15. Li S., He J., Zhuang Y. (2010) A network model of the interbank market. *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 24, pp. 5587–5593. https://doi.org/10.1016/j.physa.2010.08.057
16. Boginski V., Butenko S., Pardalos P.M. (2005) Statistical analysis of financial networks. *Computational Statistics & Data Analysis*, vol.48, no. 2, pp. 431–443. https://doi.org/10.1016/j.csda.2004.02.004
17. Lehmann E.L. (1957) A theory of some multiple decision problems, I. *The Annals of Mathematical Statistics*, vol. 28, no. 1, pp. 1–25. https://doi.org/10.1214/aoms/1177707034
18. Bretz F., Hothorn T., Westfall P. (2010) *Multiple comparisons using R*. New York: Chapman and Hall/CRC. https://doi.org/10.1201/9781420010909
19. Hochberg Y., Tamhane A.C. (1987) *Multiple comparison procedures.* New York: John Wiley & Sons. https://doi.org/10.1002/9780470316672
20. *Handbook of Multiple Comparisons* (2021) (eds. X. Cui, T. Dickhaus, Y. Ding, J.C. Hsu). New York: Chapman and Hall/CRC. https://doi.org/10.1201/9780429030888
21. Dickhaus T. (2014) *Simultaneous statistical inference. With applications in the life sciences*. Heidelberg: Springer Berlin. https://doi.org/10.1007/978-3-642-45182-9
22. Gabriel K.R. (1969) Simultaneous test procedures – some theory of multiple comparisons. *The Annals of Mathematical Statistics*, vol. 40, no. 1, pp. 224–250. https://doi.org/10.1214/aoms/1177697819
23. Cliff N. (1975) Complete orders from incomplete data: Interactive ordering and tailored testing. *Psychological Bulletin*, vol. 82, no. 2, pp. 289–302. https://doi.org/10.1037/h0076373
24. Cliff N. (1993) Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, vol. 114, no. 3, pp. 494–509. https://doi.org/10.1037/0033-2909.114.3.494
25. Donoghue J.R. (2004) Implementing Shaffer's multiple comparison procedure for a large number of groups. *Recent Developments in Multiple Comparison Procedures* (Eds. Benjamini, Y., Bretz, F. and Sarkar, S.), Institute of Mathematical Statistics, Beachwood, Ohio, pp. 1–23.

About the authors

Petr A. Koldanov

Dr. Sci. (Phys.-Math.);

Senior Research Fellow, Laboratory of Algorithms and Technologies for Networks Analysis, HSE University (campus in Nizhny Novgorod), 136, Rodionova St., Nizhny Novgorod 603093, Russia;

E-mail: pkoldanov@hse.ru

ORCID: 0000-0001-5961-0282

Vladimir A. Koldanov

Cand. Sci. (Phys.-Math.);

Associate Professor, Department of Mathematics and Computer Science, Nizhny Novgorod State University of Engineering and Economics, 22A, building 2, Oktyabrskaya St., Nizhny Novgorod Region, Knyaginino 606340, Russia;

E-mail: vlad.kold@gmail.com

ORCID: 0009-0006-1924-746X

DOI: 10.17323/2587-814X.2024.1.65.78

Determinants of an auditor's continuance intention with respect to use of the Audit Tools and Linked Archives System (ATLAS): A model of extended expectation confirmation

Kurniasari Novi Hardanti*
E-mail: kurniasarinovih@gmail.com

Sutrisno T.
E-mail: sutrisno@ub.ac.id

Erwin Saraswati
E-mail: erwin@ub.ac.id

Arum Prastiwi
E-mail: arum@ub.ac.id

Brawijaya University
Address: Veteran Street, Malang 65145, Indonesia

* Corresponding Author

Abstract

The aim of this study was to examine the determinants of the continuance intention with respect to use of the Audit Tools and Linked Archives System (ATLAS) by employing survey methods. These determinants are developed from an Expectation Confirmation Model (ECM). The sample of this study is auditors who use ATLAS in public accounting firms in Indonesia. As many as 356 data points can be processed using smartPLS. This study revealed that perceived usefulness, confirmation, information quality, top management commitment and satisfaction affected the auditor's intentions when using ATLAS. The implications of this study are (1) Public accounting firms must provide full support to auditors in using ATLAS and equip auditors through training so auditors understand that using ATLAS is very useful; (2) IAPI must pay attention to outputs that are complete, good and appropriate so that the auditor is satisfied when using ATLAS. The auditor has a tendency to continue using ATLAS if he is satisfied.

Keywords: continuance intention, audit tools, ECM, satisfaction, perceived usefulness, confirmation, information quality, top management commitment

Citation: Hardanti K.N., Sutrisno T., Saraswati E., Prastiwi A. (2024) Determinants of an auditor's continuance intention with respect to use of the Audit Tools and Linked Archives System (ATLAS): A model of extended expectation confirmation. *Business Informatics*, vol. 18, no. 1, pp. 65–78. DOI: 10.17323/2587-814X.2024.1.65.78

Introduction

The rapid development of information systems in the digital era has brought changes in various sectors. One such sector is auditing. The audit field has used information technology in the work process. The information system used by auditors in Indonesia is the Audit Tools and Linked Archives System (ATLAS).

ATLAS is one solution to improve auditor performance. Some of the benefits of ATLAS are, first, assisting the auditor in carrying out and documenting the audit process. Second, it makes it easier for the auditor to prepare audit working papers, so that the audit working papers he prepares are more systematic. Third, it reduces auditor errors when carrying out the entire audit procedure. The use of working papers processed on a computer has existed for a long time among auditors, but the ATLAS implementation is a new thing for auditors.

Auditors are not yet required to use ATLAS in preparing working papers [1]. However, the successful implementation and success of ATLAS is expected with confidence by the Ministry of Finance and Indonesian Association of Public Accountants (IAPI). In fact, not all implementations of ATLAS are in accordance with the government's expectations. Based on the results of [2], the implementation of ATLAS does not improve auditor performance and also does not improve audit quality. One reason is the low perception of the ease of using ATLAS that makes auditors reluctant to use the ATLAS system [3].

Technological and human factors are important components in the implementation of technology adoption such as ATLAS. ATLAS implementation is not considered to be in accordance with the expectations of the Indonesian Association of Public Accountants (IAPI). This is one of the causes of the human factor. Implementation failure occurs because of the reluctance or rejection of individuals within the company towards the implementation of technology [4]. Aldholay et al. [5] showed that the failure occurred due to aspects of individual behavior in the organization.

Information system implementation will be successful if the information system is accepted by its users [6].

This study will focus on the determinants of the continuance intention to use ATLAS. This model is proposed to facilitate understanding of the factors that impact ATLAS acceptance. Bhattacherjee [4] created a model, namely the Expectation Confirmation Model (ECM). ECM is different from other information system acceptance concepts. ECM focuses on the continued use of information systems, thereby providing solid explanations and long-term projections of behavior.

ECM validated in various types of studies with many sample characteristics and different countries. Several research results using the ECM model include [7–11] support for the main model of ECM, although not all of the studies that have been conducted show results consistent with previous studies. The inconsistency of the results of this study is due to differences in the characteristics of the sample and the context of the information system studied.

The ECM model has several advantages over the previous theory. However, the ECM model has not included the technological context and organizational context. The ECM model only focuses on individual contexts: confirmation constructs, satisfaction, and perceived usefulness. The researcher finds that ECM does not include the technological context and organizational context in its model.

To improve ECM, this study will incorporate both the technological context and the organizational context. The technological context is represented by system quality, information quality and service quality which had previously been formulated by [12]. In the success model, system quality measures technical success, namely the accuracy and efficiency of the system in producing information. Information quality measures semantic success, namely the success of information in conveying meaning. Service quality measures the success of the level of effectiveness, namely the influence of the information generated by the information system. The selected organizational context is top management commitment. Top management commitment is support from management (a public accountant firm) for auditors to use ATLAS.

Based on the analysis above, the aim of this study is to investigate the relationship between perceived usefulness, confirmation, information quality, design quality, service quality, top management commitment and satisfaction with the intention to continue using ATLAS among auditors in Indonesia.

This study has made its contribution in two aspects, namely the contribution to the development of theory and practice. The results of this study contribute to theory and practice related to scientific development.

- The theory contribution in this study is in terms of construct development, model development and providing empirical evidence about acceptance and ATLAS.
- The practical contribution in this study is for systems analysts and governments. The results of this study can be applied to the practice of designing and implementing ATLAS systems, so that in developing a system one can pay attention to the constructs of confirmation, satisfaction and perceived usefulness, system quality as well as top management commitment. For the government, the results of this study provide input for the Ministry of Finance and IAPI so that in making decisions to implement a system, it can pay attention to aspects of user acceptance.

1. The comprehensive theoretical basis

In this section, the concept of Expected Confirmation Theory (ECM) is introduced before the discussion regarding the relationship between related variables.

1.1. Expectation Confirmation Model

Studies on the use after adoption in the area of information systems began since [4] proposed the Expectation Confirmation Model (ECM). ECM is a model for using an information system after adoption. ECM is a development model based on Expectation Confirmation Theory (ECT) developed by [13]. The concept of ECT is integrated in the technology acceptance model (TAM) in information systems and with further refinement to address its theoretical weaknesses, [4] added a perceived usefulness variable. Perceived usefulness is felt after using an information system, not before using the system.

ECM is a theoretical model of information system sustainability hypothesizing that expectations followed by initial acceptance lead to confirmation by comparing anticipated consequences. When actual performance is confirmed, users of the information system will be satisfied and potentially this leads to continued use. In contrast to other consumer acceptance models that focus on the first use of a new information system, ECM focuses on continuing use after acceptance of the information system, and provides solid explanations and long-term scale projections of user behavior [4].

Continuance is a form of behavior after adoption [14]. Information system continuance is basically the same as repurchasing behavior, namely following the initial decision and being influenced by initial use and potentially stopping use [4]). The use of advanced information systems can be determined by the intention resulting from certain reasons. The user's interest to continue using a technology can be referred to as continuance intention.

1.2. Research model and hypothesis

Figure 1 shows the research model developed from the previously described literature. This study is different from previous studies. This study carries out construct development in the ECM proposed by [4] which includes the technological context and organizational context. The constructs in the context of this technology are system quality, information quality and service quality. System quality and information quality will determine the attitude of system users [15]. System quality, information quality and service quality will also determine individual satisfaction in using information systems [16]. Several researchers, including [12, 17–19] in an information system success model with different indicators and measurements, have used the technological context. Organizational context is also needed because it is closely related to individual attitudes in using the system. The selected organizational context is top management commitment. Top management commitment will greatly affect employee satisfaction because with this role, management can monitor the quality of the system that is being implemented within the company [20]. This model is expected to be able to prove what factors can influence satisfaction and interest in continuing to use ATLAS.

1.2.1. The hypothesis of perceived usefulness

Perceived usefulness is defined as a person's belief that he will use the system if the system has utility value [21]. The construct of perceived usefulness was originally described through TAM by [21]. TAM relates the perceived usefulness construct to the behavioral interest construct. TAM found that perceptions of usability and ease of use are prominent beliefs that influence information systems' acceptance behavior in the field of technology. ECM modifies the relationship between perceived usefulness and satisfaction. Perceived usefulness is expected to influence user satisfaction after working with the information system. The results of [4] study state that perceived usefulness has a positive effect on satisfaction. The more useful an information system, the stronger the individual satisfaction in using the information system.

The following studies [22–24] also strengthen ECM that the positive influence between perceived usefulness constructs on satisfaction in various types of applications. Based on the research results of [22], perceived usefulness has a positive impact towards satisfaction. Oghuma et al. [24] found perceived usefulness is the construct that has the greatest influence on their model. From the previous explanation, the formulation of the hypothesis is as follows.

H1: Perceived usefulness has a positive effect on satisfaction for using the ATLAS system.

1.2.2. The hypothesis of confirmation

Confirmation is defined as the perception between the conformity of expectations with reality after someone uses the system [4]. The research results of [4] show that confirmation has a positive effect toward satisfaction. Confirmation is the strongest predictor of satisfaction. The more in line with the performance of the information system with user expectations, the higher the individual satisfaction in using the information system. The positive relationship between confirmation and satisfaction formulated and proven by [4] in ECM, has been supported by many studies. From the previous explanation, the formulation of the hypothesis is as follows.

H2: Confirmation has a positive effect toward satisfaction for using the ATLAS system.

1.2.3. The hypothesis of information quality

Information quality is characteristics system output when used by user [12]. This model shows that the quality of information has a positive impact toward satisfaction. This positive relationship formulated and proven by [12] in the DeLone and McLean Information Systems (IS) success model has been supported by many studies [25–29]. From the previous explanation, the formulation of the hypothesis is as follows.

H3: Information quality has a positive effect on satisfaction for users of the ATLAS system.

1.2.4. The hypothesis of system quality

System quality has a positive impact toward satisfaction because when a system has good performance in providing the information needed by users, users will be satisfied when working with the system [12]. This work shows that system quality has a positive impact on satisfaction. The positive relationship has been supported [25–28]. From the previous explanation, the formulation of the hypothesis is as follows.

H4: System quality has a positive effect on satisfaction for using the ATLAS system.

1.2.5. The hypothesis of service quality

Service quality is defined as the quality of support by users when using the system directly. There is a positive effect between service quality and satisfaction because if users get good enough support when using a system, for example when there are difficulties using the system (errors), then there are IT staff who are ready to help quickly. In this case, user satisfaction will increase in using the system [12]. This success model shows that service quality has a positive effect toward satisfaction. The positive relationship formulated by [12] in the DeLone and McLean IS success model has been supported by many studies [26–28]. From the previous explanation, the formulation of the hypothesis is as follows.

H5: Service quality has a positive effect on satisfaction for users of the ATLAS system.

1.2.6. The hypothesis of top management commitment

Sinha et al. [30] define top management commitment as top management's active involvement in establishing and monitoring policies, communicating and encouraging employees to achieve their goals. Management plays an important role in the success of an organization. Top management commitment is believed to be able to increase the level of confidence and willingness of employees to complete tasks properly. Addis et al. [20] proved that top management commitment has a positive effect on satisfaction in quality management practices of manufacturing organizations in Ethiopia. From the previous explanation, the formulation of the hypothesis is as follows.

H6: Top management commitment has a positive effect on satisfaction when using the ATLAS system.

1.2.7. The hypothesis of satisfaction on continuance intention

Bhattacherjee [4] presented satisfaction as an emotion related to previous experience by users of information systems. The satisfaction construct is explained through ECM. ECM relates the satisfaction construct to the repeated use intention construct with the logic of thinking that the user's repeated use intention will be determined mainly from the user' satisfaction when using the given system. If the user feels bad emotions when using an information system, then this could be a reason for the user to stop using the information system. ECM proved that satisfaction has a positive effect on intention to use again.

The following studies also strengthen the theory formulated by [4] about a positive influence of the satisfaction construct on intention to use repeatedly in various types

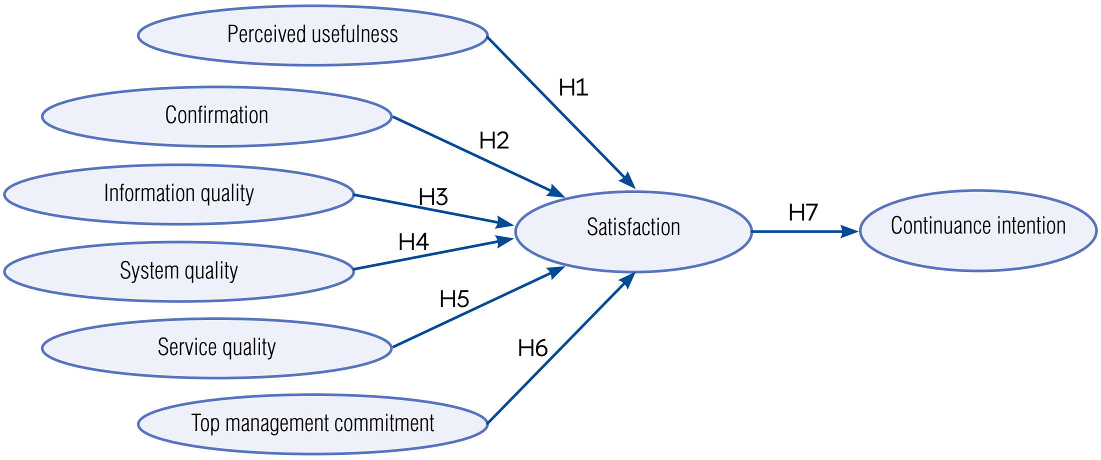

Fig. 1. Model research.

of applications by [22, 31, 32]. From the previous explanation, the formulation of the hypothesis is as follows.

H7: Satisfaction has a positive effect on the intention to continue using the ATLAS system.

2. Research method

The population in this study was auditors who use the ATLAS system in their work processes in all public accounting firms (PAF) in Indonesia that were registered with the Indonesian Public Accountants Association (IAPI) in 2022.

The unit of analysis is an auditor who works in PAF and has used the ATLAS system in his work. The sampling technique in this study is non-probability sampling. Non-probability sampling is a technique in which all members of the population do not have the same opportunity to be sampled. This study uses judgment sampling in the sampling procedure, namely the sample is taken based on existing criteria on auditors who have used ATLAS in their work process.

The data collection technique used is a survey data collection technique. The survey method of this study is a mail survey. The Mail Survey conducted in this study was by distributing questionnaires via WhatsApp, Instagram, Email and Telegram. Mail surveys have a potential problem anticipated by researchers, namely low return rates. To ensure that researchers get a high rate of return on questionnaires, certain steps are taken.

The statistical method used is Partial Least Square (PLS). By using SmartPLS ver. 4.0 M3. PLS is a structural equation modeling that can solve multiple regression problems with small samples and multicollinearity.

$$S = \beta_1 PU + \beta_2 C + \beta_3 IQ + \beta_4 SYQ + \beta_5 SEQ + \beta_6 TMC + e$$
$$CI = \beta_7 S + e,$$

where:
S – satisfaction;
PU – perceived usefulness;
C – confirmation;
IQ – information quality;
SYQ – system quality;
SEQ – services quality;
TMC – top management commitment;
CI – continuance intention;
β_i – coefficient construct;
e – error.

The constructs in this study are information quality, system quality, service quality, top management commitment, confirmation, satisfaction, perceived usefulness and intention to continue using it. The instruments used to measure constructs in this study are instruments that have been used in previous studies [4, 10, 20, 27, 33], making it possible to increase the validity and reliability of the measurements. The measurement uses a Likert scale from 1 to 7 which has the following meanings: (1) Strongly Disagree, (2) Disagree, (3) Somewhat Disagree, (4) Neutral, (5) Somewhat Agree, (6) Agree, and (7) Strongly Agree.

Before the actual distribution of the questionnaire, the researcher conducted a pilot test with the aim of convincing himself that the questionnaire items were sufficient, correct and understandable to the respondents. Researchers conducted a pre-test as follows:

1. Translate the original English instrument into Indonesian.
2. Ask for help from individuals who are experts in English to translate it back into English. Next, the results of the translation into English will be compared by the researcher with the original instrument so that from this step it is hoped that there will be no differences in the meaning or significance of the translated instrument.
3. Asking negative questions on several items in the questionnaire so that the questionnaire is not biased and there is a form of control over the questionnaire.
4. Discuss the meaning of each indicator with several ATLAS users.

Then, confirmation of the public accountant's ability to accept research and confirmation that the public accounting firm is indeed using ATLAS is carried out first by phone. After the survey was approved, the researcher brought and took the questionnaire directly to the research location for the reason of increasing the response rate from those returning the questionnaire.

The pilot test was conducted on 34 students of PPAK and Master of Accounting as well as apprentice students at PAF who had used ATLAS in the audit process. After testing, the results show that all items in the questionnaire are valid and reliable. Once it is known that the items in the questionnaire are valid and reliable, the researcher distributes the questionnaires to the real respondents in the field.

3. Results

All questionnaires that can be used and processed are 356 questionnaires. The majority of respondents, 39.89%, were junior auditors, 51.40% were senior auditors and 8.71% were supervisors. The number of female respondents was 33.99% and 66.01% were male. The majority of respondents were aged 31–40 years and the majority of respondents' highest education was the bachelor's degree. The majority of respondents had more than 5–10 years of work experience and more than 10 years of experience using computers.

3.1. Validity and reliability testing

PLS model evaluation is done by evaluating the outer model and inner model. Evaluation of the outer model is carried out by conducting convergent validity tests, discriminant validity tests and reliability tests. In order to fulfil convergent validity, all constructs must have AVE value of more than 0.5. All constructs in *Table 1* have an AVE value more than 0.5. It can be concluded that convergent validity has been fulfilled.

In order to qualify for discriminant validity, all indicators must have a factor loading value of more than 0.7. The test results show that all indicators in *Table 2* have a value of more than 0.7; it means that discriminant validity is fulfilled.

Table 1.

Output quality criteria overview model with SmartPLS

	AVE	Composite reliability	R-squared	Cronbach's Alpha
Perceived usefulness	0.837	0.954		0.935
Confirmation	0.828	0.935		0.896
Information quality	0.835	0.953		0.934
System quality	0.809	0.944		0.921
Services quality	0.798	0.941		0.916
Top management commitment	0.830	0.936		0.898
Satisfaction	0.840	0.940	0.969	0.905
Continuance intention	0.807	0.926	0.577	0.881

Table 2.

Output cross loading with SmartPLS

Indicator	C	CI	IQ	PU	S	SEQ	SYQ	TMC
C1	0.899							
C2	0.910							
C3	0.921							
CI1		0.893						
CI2		0.890						
CI3		0.912						
IQ1			0.915					
IQ2			0.906					
IQ3			0.920					
IQ4			0.914					
PU1				0.920				
PU2				0.900				
PU3				0.921				
PU4				0.917				
S1					0.913			
S2					0.908			
S3					0.928			
SEQ1						0.889		
SEQ2						0.886		
SEQ3						0.902		
SEQ4						0.898		
SYQ1							0.891	
SYQ2							0.895	
SYQ3							0.912	
SYQ4							0.899	
TMC1								0.905
TMC2								0.910
TMC3								0.918

All constructs must have a Cronbach's alpha value composite reliability must be greater than 0.7. Tests show that all constructs in *Table 2* have Cronbach's alpha values and composite reliability of more than 0.7. This means that all constructs are reliable.

Evaluation of the inner model or structural model is a stage for evaluating the relationship between constructs. The Inner Model is evaluated with R2 and statistical testing. The results of the structural model testing using the R2 value and the significance test through the path coefficient values for each path. This is the results of the significance test for the path coefficient values or *t*-values for each path.

Through *Fig. 2*, it can be concluded that the R2 values for satisfaction and interest in continuing to use the system are 0.97 and 0.58. This explains that the constructs of perceptions of usefulness, confirmation, information quality, system quality, service quality and top management commitment are able to explain the satisfaction construct by 97%; the remaining 3% is explained through other variables outside the proposed model. Continuing intention to use ATLAS can explain the satisfaction construct by as much as 58%.

The significance test is obtained through the results of the *p*-value. The hypothesis is supported if the *p*-value has a value less than 0.05. Based on *Table 3* it can be concluded that H1a, H1b, H1c, H1f, and H2 are supported because they have *p*-values less than 0.05, while H1d and H1e are rejected because they have *p*-values greater than 0.05.

This study succeeded in proving that perceptions of usefulness, confirmation, information quality and top management commitment have a positive impact on satisfaction when using the ATLAS system. Meanwhile, the satisfaction construct has a positive effect on the intention to continue using ATLAS. This study failed to prove the effect of system quality and service quality on satisfaction. The following is a discussion for each construct.

4. Discussion and conclusion

This study examines the relationship between perceived usefulness, confirmation, information quality, system quality, service quality, top management commitment, satisfaction and intention to continue using ATLAS based on the development of the ECM model [4] and [12]. The focus of this research was to investigate the auditor's continuance intention for use of ATLAS in

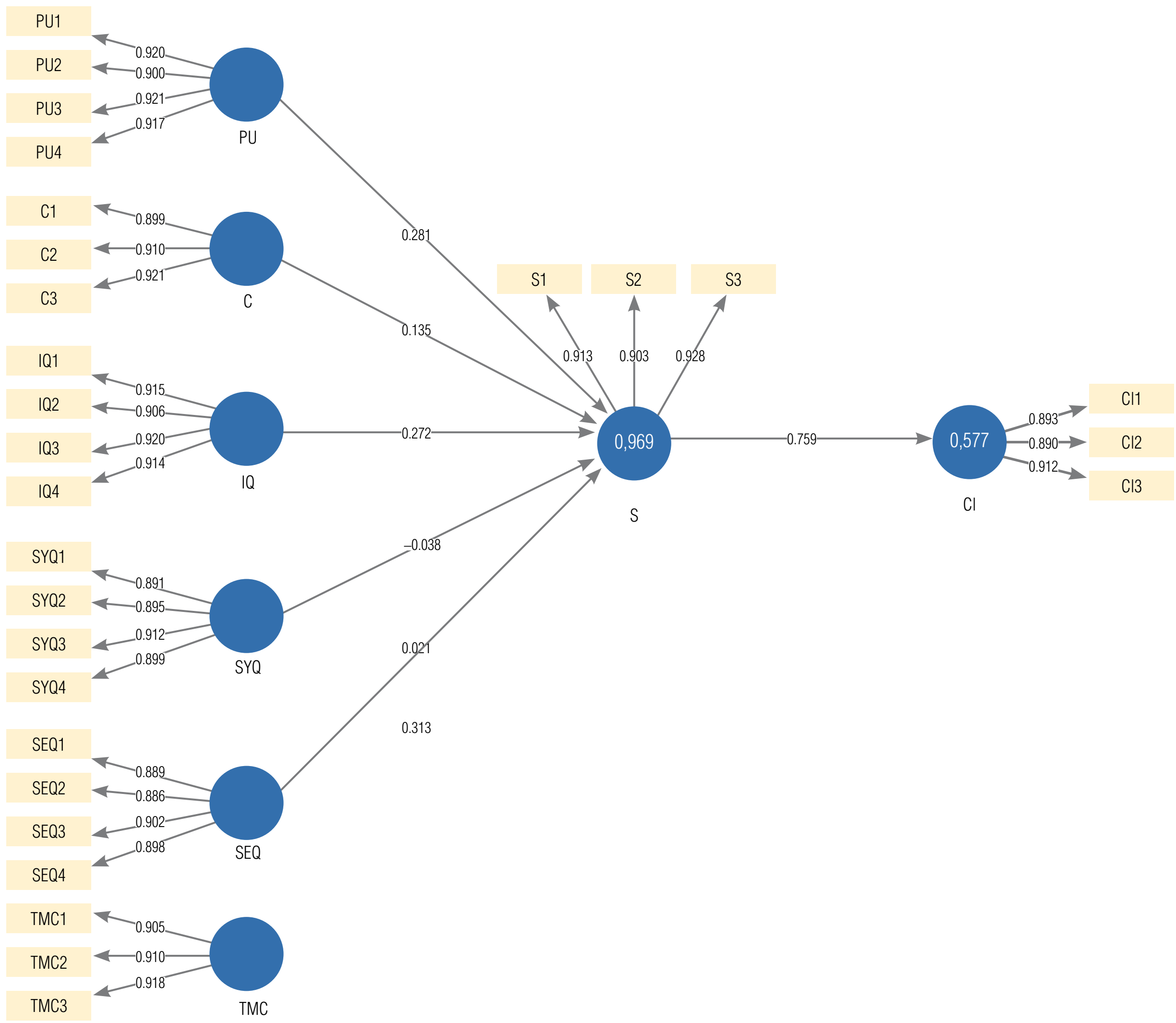

Fig. 2. Measurement model output.

the context of information systems. This study confirms that (a) the more useful a system is, the higher the auditor's satisfaction with its use; (b) the more expectations are confirmed in ATLAS, the higher the perceived satisfaction created; (c) the more complete the information available on the ATLAS, the higher the auditor's satisfaction; (d) the more useful the information provided in ATLAS, the higher the auditor's satisfaction in using it; (e) the more satisfied the auditor is in using ATLAS, the higher the interest in continuing its use in the future. The findings support the following implications in this study.

First, the greater the perception of usability, the higher the satisfaction when using ATLAS. The outcomes from this research are consistent with [22, 34–36]. This empirical evidence is consistent with ECM that perceived usefulness has an effect on satisfaction. These results imply that perceived usefulness can be a predictor of perceived satisfaction in using the system for auditors. This empirical evidence has implications that the auditor is satisfied using ATLAS if the auditor feels ATLAS is useful in supporting his performance. Thus, PAF must equip auditors to provide more training so that auditors understand that

Table 3.

Results of hypothesis testing with SmartPLS

Hypotesis	Path	Original sample	t-statistic	p-value	Conclusion
H1	$PU \rightarrow S$	0.281	3.629	0.000	Supported
H2	$C \rightarrow S$	0.135	2.024	0.022	Supported
H3	$IQ \rightarrow S$	0.272	3.606	0.000	Supported
H4	$SYQ \rightarrow S$	–0.038	0.678	0.249	Rejected
H5	$SEQ \rightarrow S$	0.021	0.374	0.354	Rejected
H6	$TMC \rightarrow S$	0.313	4.436	0.000	Supported
H7	$S \rightarrow CI$	0.759	23.145	0.000	Supported

the use of ATLAS is very useful and can support performance. When auditors feel they have the ability and control after receiving sufficient provision from the training process, they will tend to use the ATLAS system well and have high confidence.

Second, auditor expectations that have been confirmed in ATLAS positively influence auditor satisfaction as supported by previous research [22–24]. This empirical evidence is consistent with ECM that confirmation has an impact on satisfaction. These results imply that expectations can be a predictor of perceived satisfaction when auditors use the system. Good experience and maximum service for auditors must be provided to meet the expectations of auditors when using ATLAS. Thus, IAPI as the manufacturer of ATLAS must provide the best performance from ATLAS so that ATLAS can be used continuously by auditors.

Third, satisfaction increases when the quality of information is fulfilled. The same findings were found in previous studies [25–29]. This empirical evidence is consistent with the information systems achievement model, that the quality of information affects satisfaction. These results indicate that the quality of information can play an important role in understanding satisfaction in using a technology. Thus, ATLAS must be designed to be useful, understandable, attractive and believable. High quality information must be maintained and improved to provide even more satisfying results.

Fourth, top management commitment affects auditor satisfaction when using ATLAS. In other words, when management provides full support to the auditors to use ATLAS, the auditors will be satisfied with the new system. This finding is in accordance with the findings of the previous study [20]. Top management commitment must be continuously improved. This empirical evidence has implications that auditors have a high level of satisfaction when auditors in PAF get full support from top management. Top management commitment is believed to be able to increase the level of confidence and willingness of employees to complete tasks properly. Thus, in designing a system, PAF does not only think about the greatness of the implemented system, but also must provide full support to the auditors in the application of ATLAS.

In addition, perceived satisfaction positively influences intention to continue using ATLAS. The outcomes of this study are consistent with studies accompanied by [22, 32]. This empirical evidence is consistent with the ECM and information system success models, that satisfaction influences the intention to continue using the system. The findings in this study indicate that satisfaction is a strong factor influencing intention to continue using ATLAS. That is, even though auditors initially have a positive perception of ATLAS, the auditors will not continue to use ATLAS if they are dissatisfied with the system. Moreover, the use of the AT-

LAS system is voluntary, so the auditor has the power to continue using ATLAS or not. Thus, PAF management needs to think about motivating its employees so that they are more motivated to use the ATLAS system properly. When auditors are motivated to use the ATLAS system, they will tend to continue using the ATLAS system.

On the other hand, system quality and service quality do not affect satisfaction when using ATLAS. This finding is inconsistent with previous research [25–28]. This empirical evidence is supported by the study of [37]. This empirical evidence is suspect because the satisfaction felt by the auditor is subjective so that the perception of each auditor can be different. This study proves that the relationship between information quality and satisfaction in the DeLone and McLean model is not sufficient to explain the phenomenon of someone accepting or rejecting use of the ATLAS system.

With the development of technology, ATLAS was introduced as an information system that combines the entire audit process from the initial engagement to the independent auditor's report. The significance of this study is to identify the factors that influence the intention to continue using ATLAS. From the results, this study suggests a strategy to increase interest in continuing the use of ATLAS by verifying the causal relationship between related factors, and the use of ATLAS for auditors is voluntary.

The limitation of this study is that it only uses one construct related to the organizational context, namely Third, satisfaction increases when the quality of information is fulfilled. The same findings were found in previous studies [25–29]. This empirical evidence is consistent with the information systems achievement model, that the quality of information affects satisfaction. These results indicate that the quality of information can play an important role in understanding satisfaction in using a technology. Thus, ATLAS must be designed to be useful, understandable, attractive and believable. High quality information must be maintained and improved to provide even more satisfying results.

Fourth, top management commitment affects auditor satisfaction when using ATLAS. In other words, when management provides full support to the auditors to use ATLAS, the auditors will be satisfied with the new system. This finding is in accordance with the findings of the previous study [20]. Top management commitment must be continuously improved. This empirical evidence has implications that auditors have a high level of satisfaction when auditors in PAF get full support from top management. Top management commitment is believed to be able to increase the level of confidence and willingness of employees to complete tasks properly. Thus, in designing a system, PAF does not only think about the greatness of the implemented system, but also must provide full support to the auditors in the application of ATLAS.

In addition, perceived satisfaction positively influences intention to continue using ATLAS. The outcomes of this study are consistent with studies accompanied by [22, 32]. This empirical evidence is consistent with the ECM and information system success models, that satisfaction influences the intention to continue using the system. The findings in this study indicate that satisfaction is a strong factor influencing intention to continue using ATLAS. That is, even though auditors initially have a positive perception of ATLAS, the auditors will not continue to use ATLAS if they are dissatisfied with the system. Moreover, the use of the ATLAS system is voluntary, so the auditor has the power to continue using ATLAS or not. Thus, PAF management needs to think about motivating its employees so that they are more motivated to use the ATLAS system properly. When auditors are motivated to use the ATLAS system, they will tend to continue using the ATLAS system.

On the other hand, system quality and service quality do not affect satisfaction when using ATLAS. This finding is inconsistent with previous research [25–28]. This empirical evidence is supported by the study of [37]. This empirical evidence is suspect because the satisfaction felt by the auditor is subjective so that the perception of each auditor can be different. This study proves that the relationship between information quality and satisfaction in the DeLone and McLean model is not sufficient to explain the phenomenon of someone accepting or rejecting use of the ATLAS system.

With the development of technology, ATLAS was introduced as an information system that combines the entire audit process from the initial engagement to the independent auditor's report. The significance of this study is to identify the factors that influence the intention to con-

tinue using ATLAS. From the results, this study suggests a strategy to increase interest in continuing the use of ATLAS by verifying the causal relationship between related factors, and the use of ATLAS for auditors is voluntary.

The limitation of this study is that it only uses one construct related to the organizational context, namely top management commitment. Based on the limitations of this study, further studies can be carried out, namely further studies can add other constructs in the organizational context that were not examined in this study, for example organizational culture. ■

Acknowledgements

The work described in this paper was supported by Educational Fund Management Institution – Ministry of Finance.

References

1. Krismonanda C., Widyastuti S., Nugraheni R. (2021) Analysis of the application of the audit tools and linked archives system (ATLAS) to the financial statement audit process. *Jurnal Penelitian Ekonomi dan Akuntansi (JPENSI)*, vol. 6, no. 3 (in Indonesian). https://doi.org/10.30736/jpensi.v6i3.794
2. Ismail A.M. (2019) *The effect of using information technology on auditor performance (empirical study at public accounting firms in Makassar city)*. Thesis, Economics Faculty in State University of Makassar. Available at: http://eprints.unm.ac.id/14065/ (accessed 8 Feb. 2024).
3. Setiawan I., Alfie A.A., Astuti W.B. (2022) Application of the ATLAS application, competence and independence of auditors and audit quality of public accounting firms in Semarang. *The Journal of Accounting and Finance (JAFIN)*, vol. 1, no. 1, pp. 67–77 (in Indonesian).
4. Bhattacherjee A. (2001) Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly*, vol. 25, no. 3, pp. 351–370.
5. Aldholay A.H., Isaac O., Abdullah Z., Ramayah T. (2018) The role of transformational leadership as a mediating variable in DeLone and McLean information system success model: The context of online learning usage in Yemen. *Telematics and Informatics*, vol. 35, no. 5, pp. 1421–1437. https://doi.org/10.1016/j.tele.2018.03.012
6. Jeyaraj A., Zadeh A.H. (2020) Evolution of information systems research: Insights from topic modelling. *Information & Management*, vol. 57, no. 4, 103207. https://doi.org/10.1016/j.im.2019.103207
7. Joo Y.J., Park S., Shin E.K. (2017) Students' expectation, satisfaction, and continuance intention to use digital textbooks. *Computers in Human Behavior*, vol. 69, pp. 83–90. https://doi.org/10.1016/j.chb.2016.12.025
8. Yousaf A., Mishra A., Taheri B., Kesgin M. (2021) A cross-country analysis of the determinants of customer recommendation intentions for over-the-top (OTT) platforms. *Information and Management*, vol. 58, no. 8, 103543. https://doi.org/10.1016/j.im.2021.103543
9. Hsu C.L., Chen M.C. (2021) Advocating recycling and encouraging environmentally friendly habits through gamification: An empirical investigation. *Technology in Society*, vol. 66, 101621. https://doi.org/10.1016/j.techsoc.2021.101621
10. Si H., Duan X., Cheng L., Zhang Z. (2022) Determinants of consumers' continuance intention to use dynamic ride-sharing services. *Transportation Research Part D: Transport and Environment*, vol. 104, 103201. https://doi.org/10.1016/j.trd.2022.103201
11. Maduku D.K., Thusi P. (2023) Understanding consumers' mobile shopping continuance intention: New perspectives from South Africa. *Journal of Retailing and Consumer Services*, vol. 70, 103185. https://doi.org/10.1016/j.jretconser.2022.103185
12. DeLone W.H., McLean E.R. (2003) The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, vol. 19, no. 4, pp. 9–30. https://doi.org/10.1080/07421222.2003.11045748
13. Oliver R.L. (1980) A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, vol. 17, no. 4, pp. 460–469. https://doi.org/10.2307/3150499

14. Limayem M., Hirt S.G., Cheung C.M.K. (2007) How habit limits the predictive power of intention: The case of information systems continuance. *MIS Quarterly*, vol. 31, no. 4, pp. 705–737. https://doi.org/10.2307/25148817

15. Chen Y.-C., Hu L.-T., Tseng K.-C., Juang W.-J., Chang C.-K. (2019). Cross-boundary e-government systems: Determinants of performance. *Government Information Quarterly*, vol. 36, no. 3, pp. 449–459. https://doi.org/10.1016/j.giq.2019.02.001

16. Mohamadali N.A.K.S., Garibaldi J.M. (2010) A novel evaluation model of user acceptance of software technology in healthcare sector. Proceedings of the *Third International Conference on Health Informatics BIOSTEC, Valencia, Spain*, pp. 392–397. https://doi.org/10.5220/0002695703920397

17. Alyoussef I.Y. (2023) Acceptance of e-learning in higher education : The role of task-technology fit with the information systems success model. *Heliyon*, vol. 9, no. 3, e13751. https://doi.org/10.1016/j.heliyon.2023.e13751

18. Roky H., Meriouh A.Y. (2015) Evaluation by users of an industrial information system (XPPS) based on the DeLone and McLean model for IS success. *Procedia Economics and Finance*, vol. 26, pp. 903–913. https://doi.org/10.1016/s2212-5671(15)00903-x

19. Alnemer H.A. (2022) Determinants of digital banking adoption in the Kingdom of Saudi Arabia: A technology acceptance model approach. *Digital Business*, vol. 2, no. 2, 100037. https://doi.org/10.1016/j.digbus.2022.100037

20. Addis S., Dvivedi A., Beshah B. (2019) Quality management as a tool for job satisfaction improvement in low-level technology organizations: The case of Ethiopia. *Production Planning & Control*, vol. 30, no. 8, pp. 665–681. https://doi.org/10.1080/09537287.2019.1574510

21. Davis F.D., Bagozzi R.P., Warshaw P.R. (1989) User acceptance of computer-technology: A comparison of two theoretical models. *Management Science*, vol. 35, no. 8, pp. 982–1003. Available at: http://www.jstor.org/stable/2632151 (accessed 8 Feb. 2024).

22. Hadji B., Degoulet P. (2016) Information system end-user satisfaction and continuance intention: A unified modelling approach. *Journal of Biomedical Informatics*, vol. 61, pp. 185–193. https://doi.org/10.1016/j.jbi.2016.03.021

23. Sasongko D.T., Handayani P.W., Satria R. (2022) Analysis of factors affecting continuance use intention of the electronic money application in Indonesia. *Procedia Computer Science*, vol. 197, pp. 42–50. https://doi.org/10.1016/j.procs.2021.12.116

24. Oghuma A.P., Chang Y., Libaque-Saenz C.F., Park M.C., Rho J.J. (2015) Benefit-confirmation model for post-adoption behavior of mobile instant messaging applications: A comparative analysis of KakaoTalk and Joyn in Korea. *Telecommunications Policy*, vol. 39, no. 8, pp. 658–677. https://doi.org/10.1016/j.telpol.2015.07.009

25. Lee K.C., Chung N. (2009) Understanding factors affecting trust in and satisfaction with mobile banking in Korea: A modified DeLone and McLean's model perspective. *Interacting with Computers*, vol. 21, nos. 5–6, pp. 385–392. https://doi.org/10.1016/j.intcom.2009.06.004

26. Pamugar H., Winarno W.W., Najib W. (2014) Evaluation model of success and acceptance of e-learning information systems in government training institutions. *Scientific Journal of Informatics*, vol. 1, no. 1 (in Indonesian). https://doi.org/10.15294/sji.v1i1.3638

27. Tam C., Oliveira T. (2016) Understanding the impact of m-banking on individual performance: DeLone & McLean and TTF perspective. *Computers in Human Behavior*, vol. 61, pp. 233–244. https://doi.org/10.1016/j.chb.2016.03.016

28. Wang Y.S., Liao Y.W. (2008) Assessing eGovernment systems success: A validation of the DeLone and McLean model of information systems success. *Government Information Quarterly*, vol. 25, no. 4, pp. 717–733. https://doi.org/10.1016/j.giq.2007.06.002

29. Akrong G.B., Yunfei S., Owusu E. (2022) Development and validation of an improved DeLone-McLean IS success model – application to the evaluation of a tax administration ERP. *International Journal of Accounting Information Systems*, vol. 47, 100579. https://doi.org/10.1016/j.accinf.2022.100579

30. Sinha N., Garg A.K., Dhingra S., Dhall N. (2016) Mapping the linkage between organizational culture and TQM: The case of Indian auto component industry. *Benchmarking: An International Journal*, vol. 23, no. 1, pp. 208–235. https://doi.org/10.1108/BIJ-12-2014-0112

31. Mardiana S., Tjakraatmadja J.H., Aprianingsih A. (2015) Validating the conceptual model for predicting intention to use as part of Information System success model: The case of an Indonesian government agency. *Procedia Computer Science*, vol. 72, pp. 353–360. https://doi.org/10.1016/j.procs.2015.12.150

32. Prakash A.V., Das S., Pillai K.R. (2021) Understanding digital contact tracing app continuance: Insights from India. *Health Policy and Technology*, vol. 10, no. 4, 100573. https://doi.org/10.1016/j.hlpt.2021.100573

33. Huang Y.M. (2019) Examining students' continued use of desktop services: Perspectives from expectation-confirmation and social influence. *Computers in Human Behavior*, vol. 96, pp. 23–31. https://doi.org/10.1016/j.chb.2019.02.010

34. Alraimi K.M., Zo H., Ciganek A.P. (2015) Understanding the MOOCs continuance: The role of openness and reputation. *Computers & Education*, vol. 80, pp. 28–38. https://doi.org/10.1016/j.compedu.2014.08.006

35. Cheng Y.-M. (2014) Why do users intend to continue using the digital library? An integrated perspective. *Aslib Journal of Information Management*, vol. 66, no. 6, pp. 640–662. https://doi.org/10.1108/AJIM-05-2013-0042

36. Chen S.C., Yen D.C., Peng S.-C. (2018) Assessing the impact of determinants in e-magazines acceptance: An empirical study. *Computer Standards & Interfaces*, vol. 57, pp. 49–58. https://doi.org/10.1016/j.csi.2017.11.004

37. Wara L.S., Kalangi L., Gamaliel H. (2021) Testing the success model of the Delone and Mclean information systems on the examination application system (SIAP) at the supreme audit agency of the Republic of Indonesia representative for North Sulawesi province. *Jurnal Riset Akuntansi dan Auditing "GOODWILL"*, vol. 12, no. 1, pp. 1–15 (in Indonesian). https://doi.org/10.35800/jjs.v12i1.31885

About the authors

Kurniasari Novi Hardanti

Doctoral Student, Department of Accounting, Brawijaya University, Veteran Street, Malang 65145, Indonesia;

E-mail: kurniasarinovih@gmail.com

ORCID: 0009-0004-9505-3224

Sutrisno T.

Doctor of Accounting (PhD), Professor;

Professor, Department of Accounting, Brawijaya University, Veteran Street, Malang 65145, Indonesia;

E-mail: sutrisno@ub.ac.id

ORCID: 0000-0002-0147-5070

Erwin Saraswati

Doctor of Accounting (PhD), Professor;

Professor, Department of Accounting, Brawijaya University, Veteran Street, Malang 65145, Indonesia;

E-mail: erwin@ub.ac.id

ORCID: 0000-0002-8799-8605

Arum Prastiwi

Doctor of Accounting (PhD);

Senior Lecturer, Department of Accounting, Brawijaya University, Veteran Street, Malang 65145, Indonesia;

E-mail: arum@ub.ac.id

ORCID: 0000-0002-0993-8850

DOI: 10.17323/2587-814X.2024.1.79.88

Product information recognition in the retail domain as an MRC problem

Tho Chi Luong [a]
E-mail: tholc@vnu.edu.vn

Oanh Thi Tran [b] *
E-mail: oanhtt@gmail.com

[a] Institut Francophone International, Vietnam National University, Hanoi
Address: E5, 144 Xuan Thuy St., Cau Giay Dist., Hanoi, Vietnam

[b] International School, Vietnam National University, Hanoi
Address: G7, 144 Xuan Thuy St., Cau Giay Dist., Hanoi, Vietnam

Abstract

This paper presents the task of recognizing product information (PI) (i.e., product names, prices, materials, etc.) mentioned in customer statements. This is one of the key components in developing artificial intelligence products to enable businesses to listen to their customers, adapt to market dynamics, continuously improve their products and services, and improve customer engagement by enhancing effectiveness of a chatbot. To this end, natural language processing (NLP) tools are commonly used to formulate the task as a traditional sequence labeling problem. However, in this paper, we bring the power of machine reading comprehension (MRC) tasks to propose another, alternative approach. In this setting, determining product information types is the same as asking "Which PI types are referenced in the statement?" For example, extracting product names (which corresponds to the label PRO_NAME) is cast as retrieving answer spans to the question "Which instances of product names are mentioned here?" We perform extensive experiments on a Vietnamese public dataset. The experimental results show the robustness of the proposed alternative method. It boosts the performance of the recognition model over the two robust baselines, giving a significant improvement. We achieved 92.87% in the F1 score on recognizing product descriptions at Level 1. At Level 2, the model yielded 93.34% in the F1 score on recognizing each product information type.

* Corresponding Author

Keywords: product information recognition, MRC framework, retail domain, large-language models, viBERT, vELECTRA

Citation: Luong T.C., Tran O.T. (2024) Product information recognition in the retail domain as an MRC problem. *Business Informatics*, vol. 18, no. 1, pp. 79–88. DOI: 10.17323/2587-814X.2024.1.79.88

Introduction

Product Information (PI) is all the data about the products that a company sells. It includes a product's technical specifications, size, materials, prices, photos, schematics, etc. E-commerce requires companies to collect clear basic PI that consumers can actually understand and place orders. Without PI[1], the product could not be found and sold online at all.

Recognizing PI is crucial for widespread applications. For example, in the e-commerce field, it is vital to integrate this component to develop AI products like chatbots [1] to enhance customers' experiences. Chatbots significantly help reduce customer support costs, while increasing customer satisfaction with an AI chatbot that can recognize customer intents, instantly provide information, on any channel, and never take a day off. Identifying PI also helps to better analyze the sentiments [2] in comments/reviews of their customers. With PI, we can associate specific sentiments with different aspects of a product to analyze customer sentiments and opinions from reviews. This would improve the product and help make better marketing campaigns.

Conventionally, the task of PI recognition is formulated as a sequence labeling problem. It is a supervised learning problem that involves predicting an output sequence for a given input sequence. Most research in this field has proposed different machine learning approaches using handcrafted features or neural network approaches [3, 4] without using handcrafted features.

In this paper, we bring the power of machine reading comprehension (MRC) to this task. This idea is significantly inspired by a recent trend of transforming natural language processing (NLP) tasks to answering MRC questions. Specifically, Levy et al. [5] formulated the relation extraction task as a QA task. McCann et al. [6] transformed the tasks of summarization or sentiment analysis into question answering. For example, the task of summarization can be formalized as answering the question *"What is the summary?"* Li et al. [7] formalized the task of entity-relation extraction as a multi-turn question-answering problem.

So far, most current work has focused on high-resource languages. Therefore, to narrow the gap between low and high-resource languages, this paper also targets the Vietnamese language. This paper proposed an alternative way to extract PI by modeling it as a MRC problem. We conduct many extensive experiments on a public dataset by Tran et al. [1] and the results demonstrate that this approach introduces a significant performance boost over robust existing systems. The main contribution of this paper can be highlighted as follows:

- We proposed an alternative method to recognize PI by tailoring the MRC framework to suit the specific requirement of the task.
- We have conducted extensive experiments to verify the effectiveness of the proposed approach on a public Vietnamese benchmark dataset[2].

The remainder of this paper is organized as follows. Related work is presented in Section 1. Section 2 shows how to formulate the task as an MRC problem and then describes the method for generating questions, as

[1] In this paper, we consider seven types of PI types including *categories, attributes, extra-attributes, brands, packsizes, numbers,* and *unit-of-measurements (uoms)*.

[2] https://github.com/oanhtt84/PI_dataset/tree/main

well as the model architecture. Section 3 describes the experimental setups, experimental results, and some discussions. Finally, we conclude the paper and figure out some future lines of work.

1. Related work

This section first presents the work on PI identification, and then describes related work about the machine reading comprehension (MRC) tasks.

1.1. Work on PI recognition

Information retrieval chatbots are widely applied as assistants, to support customers formulate their requirements about the products they want when placing an order online. In order to develop such chatbots, most current systems use information retrieval techniques [8, 9] or a concept-based knowledge model [10] to identify product information details mentioned by their customers. Towards building task-oriented chatbots, Yan et al. [11] presented a general solution for online shopping. To extract PI asked by customers, the system matched the question to basic PI using the DSSM model. Unfortunately, these studies do not support customers who are performing orders online, and some external data resources exploited in their research are intractable in many actual applications.

Most work has been done for rich-resource languages such as English and Chinese; work for poor-resource languages is much rarer. In Vietnam, there is only one work focusing on recognizing PI types in the retail domain. Specifically, Tran et al. [1] introduced a study on understanding what the users say in chatbot systems. They concentrated on recognizing PI types implied in users' statements. In that work, they modeled the task as a sequence labelling problem and then explored different deep neural networks such as CNNs and LSTMs to solve the task.

1.2. Work on MRC

MRC refers to the ability of a machine learning model to understand and extract relevant information from written texts. It is similar to how a human reader would do this and accurately answer questions related to the content of the texts. The power of the MRC model is evaluated by the ability to extract the correct answer to the user question.

Many published novel datasets inspired a large number of new neural MRC models. In the past several years, we have witnessed many neural network models created such as BERT [12, 16] RoBERTa [13] and XLNet [10]. Many large language models utilize transformers [14] to pre-train representations by considering both the left and right context across all layers. Due to their remarkable success, this approach has progressively evolved into a mainstream method, involving pre-training large language models on extensive corpora and subsequently fine-tuning them on datasets specific to the target domain. Deep learning neural networks, particularly those based on transfer learning, are widely employed to address diverse challenges in natural language processing (NLP). Transfer learning methods emphasize the retention of data and knowledge acquired during the exploration of one problem, then applying this acquired knowledge to address different yet related questions. The effectiveness of these cutting-edge neural network models is noteworthy. For example, Lithe SOTA neural network models by Therasa et al. [12] has already exceeded human performance over many related MRC benchmark datasets.

In this paper, we borrow the idea of MRC to propose another alternative approach to this task. To prove the effectiveness of the approach, we conduct extensive experiments on a public Vietnamese dataset released by Tran et al. [1]. The results showed a new SOTA result over the traditional existing techniques.

2. Recognizing PI as an MRC problem

In this section, we first formulate the task of recognizing PI as an MRC problem. Then, we show the method to generate questions/queries for finding the answers (which could be the product information instances) appearing in the users' input utterances. Finally, the model architecture is presented and explained in more details.

2.1. Problem formulation

Given a users' statement x including n syllables $\{x_1, x_2, \ldots, x_n\}$, we need to build a model to identify every product information mentioned in x. For each instance of a product information type found in x we assign a label y to it. Here, y belongs to one of the pre-defined PI list including *product names*, *product size*, *product unit-of-measurements*, *product attribute*, *product brand*, *product number*, and *product extra attribute*.

To exploit the MRC approach, it is necessary to recast the task as an MRC problem. To this end, we construct triples of questions, answers, and contexts $\{q_{pi}, x_{start:end}, \ldots, x\}$ for each label pi mentioned in x as follows:

- x: the user's statement
- $x_{start:end}$: the product information mentioned in x. It is a sequence of syllables within x identified by the specified start and end indexes $\{x_{start}, \ldots x_{end}\}$, where the condition start <= end holds true. Expert knowledge is required to annotate this data.
- q_{pi}: the question to ask the model to find $x_{start:end}$ corresponding to the label pi. This is a natural question consisting of m syllables $\{q_1, q_2, \ldots, q_m\}$. Various approaches will be investigated in order to generate such questions.

This exactly establishes the triple (*Question, Answer, Context*) to be exploited in the proposed framework. And now, the task can be recast as an MRC problem as follows: Given a collection of k training examples $\{q^i, x^i_{start:end}, x^i\}$ (where $i = 1..k$). The purpose is to train a predictor which receives the statement x and the corresponding question q, and outputs the answer $x_{start:end}$. It is formulated as the following formula:

$$x^i_{start:end} = f(q^i, x^i).$$

2.2. Question generation

Each PI is associated with a specific question generated by combining the predefined templates and its training example values. It is a natural language question. In order to provide more prior knowledge about the label, we add some examples to the questions so that the model can recognize answers easier. These examples are randomly withdrawn from the training data set. Some typical generated questions for product information types are shown in *Table 1.*

Here we just provide some examples of each product information type to help the model find all of its instances appearing in the input statement.

Table 1.

Some questions generated for each PI types using templates

No.	Product information types	Generated questions
1	Product names	Which product names are mentioned in the text such as *smoothies* and *cakes*?
2	Product sizes	Which product sizes are mentioned in the text such as *big* and *small*?
3	Product colors	Which product colors are mentioned in the text such as *green* and *blue*?
4	Product uoms	Which product uoms are mentioned in the text such as *cup* and *cm*?
5	Product attributes	Which product attributes are mentioned in the text such as *extra ice* and *little sugar*?
6	Product extra attributes	Which product extra attributes are mentioned in the text such as *strawberry flavor* and *orange flavor*?
7	Product brand	Which brands are mentioned in the text such as *Samsung* and *Toyota*?

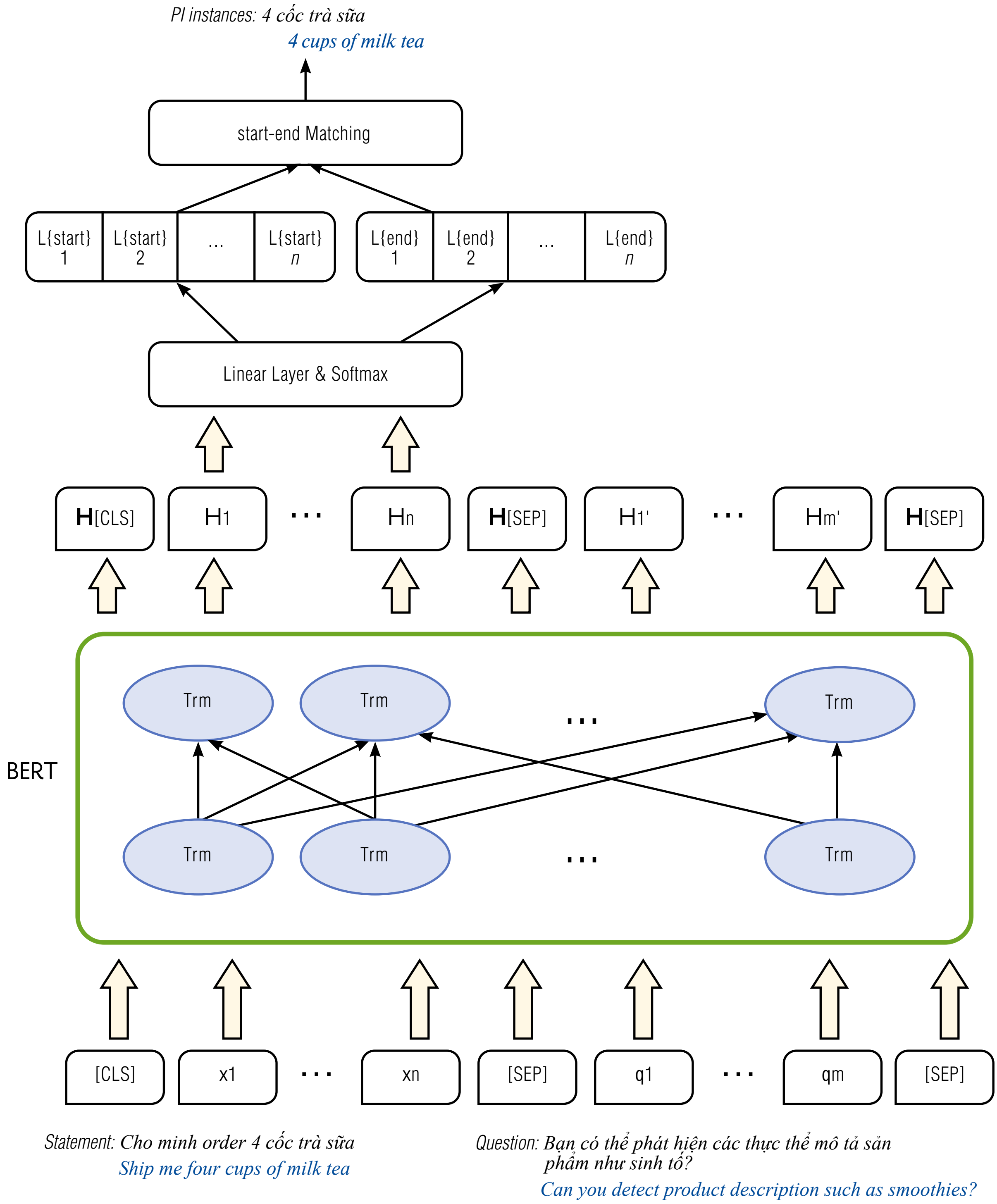

Fig. 1. An architecture using BERT to solve the PI recognition task as an MRC problem.
(English translation is given right below the Vietnamese texts).

2.3. Model architecture

Figure 1 shows the general architecture which includes several main components. The model in this framework is built with a pre-trained large language model (i.e., BERT encoder) and a network designed to produce candidate for *start* and *end* indexes, along with their associated confidence scores indicating the likelihood of being product information.

Given the question q_{pi}, the purpose is to find the text span $x_{start:end}$$ categorized as the product information type *pi*. In the first step, q_{pi} and x are concatenated to establish the string {[*CLS*], q_1, q_2, ..., q_m; [*SEP*], [x_1, x_2, ..., x_n},where [*CLS*] and [*SEP*] are special tokens employed in the conventional pre-trained LLMs. Then, the string is inputted into BERT to generate a contextual representation matrix $E \in R^{n \times d}$, (here d indicates the vector dimension of the final layer). Here, we do not make any prediction for the question, so its final vector representation is ignored.

2.4. Producing the indexes of start:end

To this end, we follow the method proposed by Li et al. [7] to build two corresponding binary classifiers. These two classifiers estimate the probability of each token to be a start or an end index using a softmax function. Specifically, p^{start}, $p^{end} \in R^n$ indicate the vectors that show the likelihood in probability of each token being the start index and end index, respectively:

$$[p^{start}, p^{end}] = softmax\ (EW + B),$$

where both W and $B \in R^{n^2}$ are trainable parameters.

Then, a ranked list of potential Product Information (PI) along with corresponding confidence scores is produce by the model. These scores are computed as the sum of the probabilities associated with their *start* and *end* tokens.

In training, the overall objective is to minimize the global loss of three types which are losses for start index, end index and start-end index matching. These losses are simultaneously trained in an end-to-end framework. We use [15] to optimize the loss.

3. Experiments

This section first shows the general information about the public benchmark dataset used for experiments. Then, it tells us about the setups of experiments. Finally, the experimental results and discussion are shown.

3.1. Dataset

In this paper, we used the dataset released by Tran and Luong [1] to perform comparative experiments. This data was collected from a history log of a retail restaurant, some forums and social websites. It was annotated with seven main types of PI which are *product category, product attribute, product extra-attribute, product brand, product packsizes, product number,* and *product uoms*. Two levels of annotation were provided. At the first level, descriptions of products (Level 1) are extracted. Then, these product descriptions are further decomposed into some detailed PI types (Level 2). An example is given in *Table 2.*

3.2. Experimental setups

The models are evaluated using popular metrics such as precision, recall, and F1 scores [17]. The best parameters were fine-tuned on development sets. The best values for parameters and hyper-parameters are listed as follows:

- Train sequence length: 768
- Number of epochs: 300
- Batch size: 8
- Learning rate: 3e-5
- Adam epsilon: 1e-8
- Max gradient norm: 1.0
- Bert embedding: 768 dimensions.

Table 2.

One example of a user's statement annotated at two levels.

(English translation is provided right after the Vietnamese statement at the first row)

Utterances	*Cho em đặt*	1	*hộp*	*bánh kem*	*vị xoài*	*Cỡ lớn*
	Let me order	1	*pack*	*cream cake*	*mango flavor*	*big size*
Level 1	other	Product description				
Level 2	other	number	uom	category	attribute	size

We adapted the MRC framework[3] to this task and exploited the viBERT[4] and vELECTRA[5], a pre-trained large language model optimized for Vietnamese, to build the PI recognition model. In case the pre-trained models are not available to optimize for a specific language, it is also feasible to use a multilingual pre-trained model, such as mBERT (a.k.a multi-lingual BERT) in order to get the vector representations for its sentences. We trained the model on the GPU Tesla V100 SXM2 32GB.

3.3. Experimental results

Tables 3 and *4* show the experimental results of the proposed model in comparison to the two baselines which are BiLSTM-CRF and CNN-CRF [1].

At Level 1, we can see that the MRC approach boosted the performance by a large margin on all evaluation metrics. In comparison to the best baseline CNN-CRF, it enhanced F1 score by nearly 2% in the case of using viBERT and 2.4% in the case of using vELECTRA. This suggested that the MRC approach is very promising and yields a better performance than other traditional approaches.

Table 3.

Experimental results of the models at Level 1 — Product descriptions

	Precision	**Recall**	**F1-scores**
biLSTM-CRF	89.71	91.35	90.52
CNN-CRF	90.6	91.24	90.91
MRC-viBERT	94.1	91.68	92.87
MRC-vELECTRA	**94.5**	**92.18**	**93.33**

At Level 2, in comparison to biLSTM-CRF, it significantly outperformed this baseline in all product information types. The MRC-viBERT approach also slightly increased the F1 score by about 0.3% in comparison the best baseline of CNN-CRF method. Among seven PI types, it achieved a significant improvement over the two baselines by a large margin on three PI types (i.e. *product branch, product category*, and *product extra_attribute*). For the type of attribute, the proposed approach got the competitive results.

[3] https://github.com/CongSun-dlut/BioBERT-MRC

[4] https://github.com/fpt-corp/viBERT

[5] https://github.com/fpt-corp/vELECTRA

Table 4.

Experimental results of the models at Level 2 — Product Information Types

	biLSTM-CRF			CNN-CRF			MRC-viBERT			MRC-vELECTRA		
PI types	**Pre**	**Rec**	**F1**	**Pre**	**Rec**	**F1**	**Pre**	**Rec**	**F1**	**Pre**	**Rec**	**F1**
attribute	93.69	95.63	94.63	95.9	97.24	95.8	95.82	95.25	95.53	96.02	95.71	**95.86**
brand	82.44	83.24	82.77	89.38	88.64	88.98	92.04	89.90	90.90	92.79	90.65	**91.71**
category	86.24	88.45	87.32	91.44	91.90	91.67	93.57	93.88	93.72	94.17	93.98	**94.07**
extra attribute	87.89	86.76	87.26	88.83	86.24	87.39	94.03	88.76	91.31	95.01	89.04	**91.93**
packsize	85.03	86.82	85.84	91.62	93.14	**92.36**	92.23	88.77	90.41	93.04	89.21	91.08
sys number	95.24	95.35	95.28	95.88	95.92	**95.89**	95.29	92.04	93.62	96.12	92.57	94.31
uom	88.80	91.73	90.16	92.12	92.33	**92.19**	89.16	93.07	91.05	90.01	93.11	91.53
Total	89.39	90.86	90.11	92.95	93.21	93.08	93.69	93.01	93.34	94.11	93.76	**93.93**

It surpassed biLSTM-CRF, but could not overcome CNN-CRF on the remaining three PI types (i.e., *product packsize, product sys_number,* and *product uom*).

Among two types of MRC basing on viBERT and vELECTRA as backbone, we witnessed that MRC-vELECTRA performed slightly better than MRC-viBERT. It increased the performance on four PI types (i.e. *product attribute, product branch, product category*, and *product extra_attribute*). However, similar to MRC-viBERT, the MRC-vELECTRA also could not surpass CNN-CRF on the remaining three PI types. Overall, in comparison to the best baseline — CNN-CRF, the MRC-vELECTRA increased the F1 score by 0.85%. This result is quite promising.

3.4. Discussion

Looking at the results shown in *Table 3* and *Table 4*, we acknowledge that using the MRC approach yielded higher F1 scores at both levels. This is because the queries/questions generated provide more prior knowledge to guide the identification process of product information.

It can be also seen that the proposed method yielded better performance on recognizing long PI types (such as *product attributes, product description, product extra attribute*) in comparison to the best baseline — CNN-CRF. This can be explained as follows: the MRC approach captures the sequence information better than CNN. CNN only leverages the local contexts based on n-gram characters and word embeddings. So, it does not have the power of capturing long PI types as compared to the MRC approach. Among two types of word embeddings, the MRC-vELECTRA was slightly better than MRC-viBERT on both two PI levels.

This proposed approach can be generalized to any language. In case BERT is not available to a specific language, we can instead use the mBERT (multi-linguage BERT) as the backbone.

Conclusion

This paper described the task of identifying product information mentioned by customers' statements in a retail domain. This is a vital step in developing many artificial intelligence commercial products. In contrast to many previous studies, we did not formulate the task as a conventional sequence labeling problem. Instead, we make use of the robustness of MRC tasks to propose an alternative approach. The proposed MRC architecture also leverages the knowledge gained during pre-training a large language model and then applies it to a new, related task – MRC. We performed experiments on a Vietnamese public benchmark dataset to verify the effectiveness of the proposed method. We achieved a new SOTA result by boosting the recognition performance over the two strong baselines. Specifically, we achieved 93.33% in the F1 score on recognizing product descriptions at Level 1 (upgraded by 2.4%). At Level 2, the model slightly improved the performance and yielded 93.93% in the F1 score on recognizing each product information type by using MRC-vELECTRA. The results also suggested that this approach is more effective in predicting long PI types with high precision.

In the future, we will continue exploring different kinds of generating questions by providing more clues to help find the product information. Furthermore, we will explore alternative robust pre-trained language models to improve the predictive model. ■

Acknowledgements

This paper was funded by the International School, Vietnam National University Hanoi under the project CS.NNC/2021-07.

References

1. Tran O.T., Luong T.C. (2020) Understanding what the users say in chatbots: A case study for the Vietnamese language. *Engineering Applications of Artificial Intelligence*, vol. 87, 103322. https://doi.org/10.1016/j.engappai.2019.103322

2. Tran O.T., Bui V.T. (2020) A BERT-based hierarchical model for Vietnamese aspect based sentiment analysis. Proceedings of the *12th International Conference on Knowledge and System Engineering (KSE), Can Tho, Vietnam, 2–14 November 2020*, pp. 269–274. https://doi.org/10.1109/KSE50997.2020.9287650

3. Bui V.T., Tran O.T., Le H.P. (2020) Improving sequence tagging for Vietnamese text using transformer-based neural models. *arXiv:2006.15994.* https://doi.org/10.48550/arXiv.2006.15994

4. Lample G., Ballesteros M., Subramanian S., Kawakami K., Dyer C. (2016) Neural architectures for named entity recognition. Proceedings of the *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, California, June 2016*, pp. 260–270. https://doi.org/10.18653/v1/N16-1030

5. Levy O., Seo M., Choi E., Zettlemoyer L. (2017) Zero-shot relation extraction via reading comprehension. *arXiv:1706.04115.* https://doi.org/10.48550/arXiv.1706.04115

6. McCann B., Keskar N.S., Xiong C., Socher R. (2018) The natural language decathlon: Multitask learning as question answering. *arXiv:1806.08730.* https://doi.org/10.48550/arXiv.1806.08730

7. Li X., Yin F., Sun Z., Li X., Yuan A., Chai D., Zhou M., Li J. (2019) Entity-relation extraction as multi-turn question answering. Proceedings of the *57th Conference of the Association for Computational Linguistics, Florence, Italy, July 2019*, pp. 1340–1350. https://doi.org/10.18653/v1/P19-1129

8. Ji Z., Lu Z., Li H. (2014) An information retrieval approach to short text conversation. *arXiv:1408.6988.* https://doi.org/10.48550/arXiv.1408.6988

9. Qiu M., Li F., Wang S., Gao X., Chen Y., Zhao W., Chen H., Huang J., Chu W. (2017) AliMe chat: A sequence to sequence and rerank based chatbot engine. Proceedings of the *55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Vancouver, Canada, July 2017*, pp. 498–503. https://doi.org/10.18653/v1/P17-2079

10. Goncharova E., Ilvovsky D.I., Galitsky B. (2021) Concept-based chatbot for interactive query refinement in product search. Proceedings of the *9th International Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI 2021)*, vol. 2972, CEUR-WS, pp. 51–58. Available at: http://ceur-ws.org/Vol-2972/paper5.pdf (accessed 15 February 2024).

11. Yan Z., Duan N., Chen P., Zhou M., Zhou J., Li Z. (2017) Building task-oriented dialogue systems for online shopping. Proceedings of the *AAAI Conference on Artificial Intelligence*, vol. 31, no. 1. https://doi.org/10.1609/aaai.v31i1.11182

12. Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L., Stoyanov V. (2019) RoBERTa: A robustly optimized BERT pretraining approach. *arXiv:1907.11692.* https://doi.org/10.48550/arXiv.1907.11692

13. Yang Z., Dai Z., Yang Y., Carbonell J., Salakhutdinov R.R., Le Q.V. (2019) XLNet: Generalized autoregressive pretraining for language understanding. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, arXiv:1906.08237.* https://doi.org/10.48550/arXiv.1906.08237

14. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. (2017) Attention is all you need. *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, arXiv:1706.03762.* https://doi.org/10.48550/arXiv.1706.03762

15. Kingma J., Ba J. (2015) Adam: A method for stochastic optimization. *arXiv:1412.6980.* https://doi.org/10.48550/arXiv.1412.6980

16. Devlin J., Chang M.W., Lee K., Toutanova K. (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805.* https://doi.org/10.48550/arXiv.1810.04805

17. Therasa M., Mathivanan G. (2022) Survey of machine reading comprehension models and its evaluation metrics. Proceedings of the *6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India*, pp. 1006–1013. https://doi.org/10.1109/ICCMC53470.2022.9754070

About the authors

Tho Chi Luong

Researcher, Institut Francophone International, Vietnam National University, Hanoi, E5, 144 Xuan Thuy St., Cau Giay Dist., Hanoi, Vietnam;

E-mail: tholc@vnu.edu.vn

ORCID: 0000-0002-7664-705X

Oanh Thi Tran

Associate Professor, PhD;

Lecturer, International School, Vietnam National University, Hanoi, G7, 144 Xuan Thuy St., Cau Giay Dist., Hanoi, Vietnam;

E-mail: oanhtt@gmail.com

ORCID: 0000-0002-3286-3623

25th YASIN (APRIL) INTERNATIONAL ACADEMIC CONFERENCE ON ECONOMIC AND SOCIAL DEVELOPMENT

Dear colleagues,

HSE University is pleased to announce a call for proposals to take part in the **25th Yasin (April) International Academic Conference on Economic and Social Development** (25th Yasin Conference).

The key events of the 25th Yasin Conference will be taking place in Moscow **from April 23 to 26, 2024.**

Reports on new research results will be presented and discussed as part of the Conference's sections.
These reports will be selected through reviews of proposals.
Furthermore, the Conference's programme traditionally features expert discussions of the most pressing economic, social, internal and external issues in the format of roundtables and associated events.

Applications for attending the Conference as a participant can be submitted until **Friday**, **April 19, 2024.**

Further details are available on the Conference website in the section for 'Participants'.

The Conference's events will be held in Russian or English. Certain discussions will be bilingual with simultaneous translation services provided.

With a view to involving participants from Russia's various regions and all over the world, as well as bearing in mind that certain epidemiological restrictions still may be in effect, the 25th Yasin Conference will be held in a hybrid format.

A competitive selection of proposals submitted by young researchers from Russian regions for taking part in the Conference will be traditionally held as part of the 25th Yasin Conference.

25th YASIN (APRIL) INTERNATIONAL ACADEMIC CONFERENCE ON ECONOMIC AND SOCIAL DEVELOPMENT

Proposals for taking part in the Conference with a report will be focused on the following areas:

1. Asian Studies;
2. Geography and Geo-information Technologies;
3. Demography and Labour Markets;
4. Instrumental Methods in Economic and Social Studies;
5. Corporate Finance in the Context of Global Challenges;
6. Macroeconomics and Macroeconomic Policy;
7. International Relations;
8. Economic Methodology;
9. World Economy;
10. Science and Innovation;
11. Education;
12. Political Processes;
13. Law in the Digital Age;
14. Psychology;
15. Healthcare Studies;
16. Regional and Urban Development;
17. Management;
18. Social Policy;
19. Sociocultural Processes;
20. Sociology;
21. Theoretical Economics;
22. Smart City;
23. Financial Institutes, Markets and Payment Systems;
24. Firms and Markets;
25. Data-Driven Economy.

The final deadline for filing proposals online via the HSE University's Conference system is **Friday, April 19, 2024.**

Registration fees

Participation fee for listeners (participants without a report): **RUB 3,000**.
See detailed information in the "Conference fees".

Participation in the Conference will be free of charge for all other participants, including:

- Students and PhD students from any University (upon presentation of their Student ID);
- HSE University's staff (upon presentation of their staff ID badge);
- Honourary guests invited by the Conference Programme and Organizing Committees.