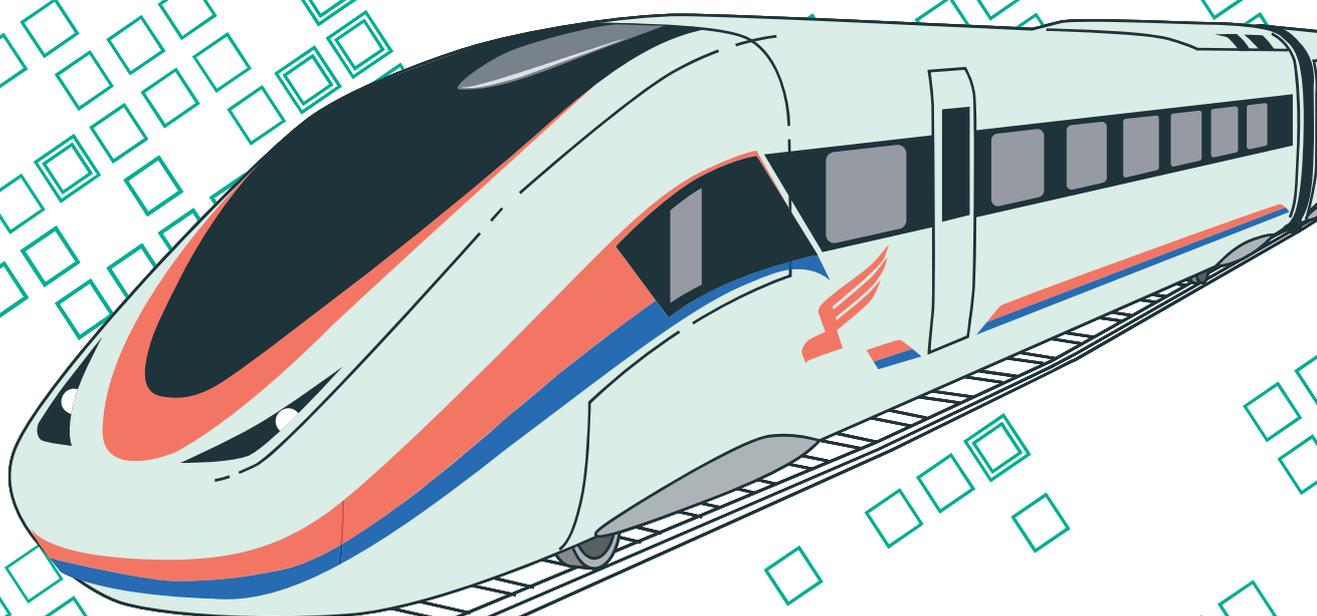


Интеллектуальные технологии на транспорте

Intellectual Technologies
on Transport



Выпуск 3
2024

ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ НА ТРАНСПОРТЕ

(сетевой электронный научный журнал)

Выпуск 3 (39), 2024

ISSN 2413-2527

Сетевой электронный научный журнал, свободно распространяемый через интернет. Публикуются статьи на русском и английском языках с результатами исследований и практических достижений в области интеллектуальных технологий и сопутствующих им научных исследований. Журнал основан в 2015 году.

Учредитель

Федеральное государственное бюджетное образовательное учреждение высшего образования «Петербургский государственный университет путей сообщения Императора Александра I» (ФГБОУ ВО ПГУПС)

Издатель

ООО «Медиа-Сервис» по договору № ЭА00271 от 19.12.2023.

Периодичность выхода — 4 номера в год.

Адрес редакции:

190031, Санкт-Петербург, Московский пр., 9
e-mail: itt-pgups@yandex.ru
Телефон: +7 (812) 457-86-06

Сетевое издание «Интеллектуальные технологии на транспорте (сетевой электронный научный журнал), Intellectual Technologies on Transport» зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций. Свидетельство Эл № ФС77-61707 от 7.05.2015.

Журнал зарегистрирован в Российском индексе научного цитирования (РИНЦ).

Информация предназначена для читателей старше 12 лет.

Выпуски журнала доступны на сайте <http://itt-pgups.ru>

Хомоненко А. Д., д. т. н., проф., С.-Петербург, РФ — главный редактор

Божко Л. М., д. э. н., доц., ПГУПС, С.-Петербург, РФ — заместитель главного редактора

Баталов Д. И., к. т. н., ПГУПС, С.-Петербург, РФ — научный редактор

РЕДАКЦИОННЫЙ СОВЕТ

Сопредседатели редакционного совета:

Валинский О. С., к. т. н., ректор ПГУПС, С.-Петербург, РФ
Чаркин Е. И., зам. гендиректора по ИТ ОАО «РЖД», Москва, РФ

Ададуров С. Е., д. т. н., проф., ВНИИЖТ, Москва, РФ

Дудин А. Н., д. ф.-м. н., проф., БГУ, Минск, Беларусь

Корниенко А. А., д. т. н., проф., ПГУПС, С.-Петербург, РФ

Макаренко С. И., д. т. н., доц., ПАО «Интелтех», С.-Петербург, РФ

Меркурьев Ю. А., Dr. Habil., проф., член Латвийской АН, РТУ, Рига, Латвия

Титова Т. С., д. т. н., проф., первый проректор ПГУПС, С.-Петербург, РФ

Юсупов Р. М., д. т. н., проф., чл.-корр. РАН, СПб ФИЦ РАН, С.-Петербург, РФ

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Александрова Е. Б., д. т. н., проф., ПГУПС, С.-Петербург, РФ

Басыров А. Г., д. т. н., проф., ВКА, С.-Петербург, РФ

Безродный Б. Ф., д. т. н., проф., НИИАС, Москва, РФ

Благовещенская Е. А., д. ф.-м. н., проф., ПГУПС, С.-Петербург, РФ

Бубнов В. П., д. т. н., проф., С.-Петербург, РФ

Булавский П. Е., д. т. н., доц., ПГУПС, С.-Петербург, РФ

Василенко М. Н., д. т. н., проф., ПГУПС, С.-Петербург, РФ

Глухов А. П., д. т. н., ПГУПС, С.-Петербург, РФ

Гуда А. Н., д. т. н., проф., РГУПС, Ростов-на-Дону, РФ

Ермаков С. Г., д. т. н., проф., ПГУПС, С.-Петербург, РФ

Заборовский В. С., д. т. н., проф., СПбПУ, С.-Петербург, РФ

Канаев А. К., д. т. н., проф., ПГУПС, С.-Петербург, РФ

Котенко А. Г., д. т. н., проф., ВНИИЖТ, Москва, РФ

Куренков П. В., д. э. н., к. т. н., проф., РУТ (МИИТ), Москва, РФ

Лецкий Э. К., д. т. н., проф., РУТ (МИИТ), Москва, РФ

Наседкин О. А., к. т. н., доц., ПГУПС, С.-Петербург, РФ

Никитин А. Б., д. т. н., проф., ПГУПС, С.-Петербург, РФ

Новиков Е. А., д. т. н., доц., ВКА, С.-Петербург, РФ

Охтилев М. Ю., д. т. н., проф., НИО ЦИТ «Петрокомета», С.-Петербург, РФ

Привалов А. А., д. воен. н., проф., С.-Петербург, РФ

Соколов Б. В., д. т. н., проф., СПб ФИЦ РАН, С.-Петербург, РФ

Таранцев А. А., д. т. н., проф., ИПТ РАН, С.-Петербург, РФ

Утепбергенов И. Т., д. т. н., проф., АУЭС, Алма-Ата, Казахстан

Фазылов Ш. Х., д. т. н., проф., НИИ развития цифровых технологий и ИИ, Ташкент, Узбекистан

Хабаров В. И., д. т. н., проф., СГУПС, Новосибирск, РФ

Ходаковский В. А., д. т. н., проф., ПГУПС, С.-Петербург, РФ

Чехонин К. А., д. ф.-м. н., доц., ХВИЦ ДВО РАН, Хабаровск, РФ

INTELLECTUAL TECHNOLOGIES ON TRANSPORT
(Network electronic scientific journal)
Issue 3 (39), 2024
ISSN 2413-2527

Network electronic scientific journal, open access.
It publishes articles in Russian and English with the
results of research and practical achievements in the field
of intelligent technologies and associated research.
Founded in 2015.

Founder

Federal State Budgetary Educational Institution
of Higher Education “Emperor Alexander I
St. Petersburg State Transport University”

Publisher

Media Service LLC No. EA00271,
19.12.2023

Frequency of release — 4 issues per year.

Editorial address:

190031, St. Petersburg, Moskovsky ave., 9
e-mail: itt-pgups@yandex.ru
Phone: +7 812 457 86 06

The online journal “Intellectual Technologies
on Transport” is registered by the Federal Service
for Supervision of Communications, Information
Technologies, and Mass Media.
El No. FS77-61707 Testimony from 7.05.2015.

The journal is registered in the Russian Science Citation
Index (RSCI).

The content is for readers over the age of 12.
Issues of the magazine are available at <http://itt-pgups.ru>

Khomonenko A. D., Dr. Sci. in Engineering, Prof., St.
Petersburg, Russia — Editor-in-Chief

Bozhko L. M., As. Prof., PSTU, St. Petersburg, Russia —
Deputy Editor-in-Chief

Batalov D. I., PhD in Engineering, PSTU, St. Petersburg,
Russia — Science Editor

EDITORIAL COUNCIL MEMBERS

Co-chairs of the Editorial Council:

Valinsky O. S., *Co-chairs of the Editorial Council, rector
of PSTU, St. Petersburg, Russia*
Charkin E. I., *CIO of JSC “Russian Railways”, Moscow,
Russia*

Adadurov S. E., Prof., VNIIZHT, Moscow, Russia

Dudin A. N., Prof., BSU, Minsk, Belarus

Kornienko A. A., Prof., PSTU, St. Petersburg, Russia

Makarenko S. I., As. Prof., Inteltech, St. Petersburg, Russia

Merkuryev Yu. A., Prof., Academician of the Latvian
Academy of Sciences, RTU, Riga, Latvia

Titova T. S., Prof., First Vice-Rector PSTU, St. Petersburg,
Russia

Yusupov R. M., Prof., Corr. Member of RAS, SPC RAS,
St. Petersburg, Russia

EDITORIAL BOARD MEMBERS

Aleksandrova E. B., Prof., SPbPU, St. Petersburg, Russia

Basyrov A. G., Prof., MSA, St. Petersburg, Russia

Bezrodny B. F., Prof., NIIAS, Moscow, Russia

Blagoveshchenskaya E. A., Prof., PSTU, St. Petersburg,
Russia

Bubnov V. P., Prof., St. Petersburg, Russia

Bulavsky P. E., As. Prof., PSTU, St. Petersburg, Russia

Vasilenko M. N., Prof., PSTU, St. Petersburg, Russia

Glukhov A. P., PSTU, St. Petersburg, Russia

Guda A. N., Prof., RSTU, Rostov-on-Don, Russia

Ermakov S. G., Prof., PSTU, St. Petersburg, Russia

Zaborovsky V. S., Prof., SPbPU, St. Petersburg, Russia

Kanaev A. K., Prof., PSTU, St. Petersburg, Russia

Kotenko A. G., Prof., VNIIZHT, Moscow, Russia

Kurenkov P. V., Prof., RUT (MIIT), Moscow, Russia

Letsky E. K., Prof., RUT (MIIT), Moscow, Russia

Nasedkin O. A., As. Prof., PSTU, St. Petersburg, Russia

Nikitin A. B., Prof., PSTU, St. Petersburg, Russia

Novikov E. A., As. Prof., MSA, St. Petersburg, Russia

Okhtilev M. Yu., Prof., JSC “Petrokometa”, St. Petersburg,
Russia

Privalov A. A., Prof., PSTU, St. Petersburg, Russia

Sokolov B. V., Prof., SPC RAS, St. Petersburg, Russia

Tarantsev A. A., Prof., IPT RAS, St. Petersburg, Russia

Utebergenov I. T., Prof., AUPET, Almaty, Kazakhstan

Fazilov Sh. X., Prof., AIRI, Tashkent, Uzbekistan

Khabarov V. I., Prof., STU, Novosibirsk, Russia

Khodakovsky V. A., Prof., PSTU, St. Petersburg, Russia

Chekhonin K. A., Prof., Khabarovsk FRC RAS,
Khabarovsk, Russia

Содержание

Искусственный интеллект и машинное обучение

Самонов А. В.

Методы оптимизации процесса обучения и тонкой настройки больших языковых моделей (in English) 5

Математическое моделирование, численные методы и комплексы программ

Алимов Б. М., Султонов Ш. Х., Кодирова Е. В.

Математическое моделирование бипланетарного механизма для привода смесительных машин. 13

Черняховская Е. С.

Математическая модель выкладки товаров на полочном пространстве складского стеллажа с одновременной горизонтальной и вертикальной категоризацией товаров 21

Липанов И. Д., Хомоненко А. Д.

Технологии и методы планирования перемещения БПЛА по маршрутным точкам 30

Системный анализ, управление и обработка информации, статистика

Соколов С. С., Антонова А. Е.

Применение имитационного моделирования в процессах постройки судна 44

Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей

Забродин А. В., Добровольская М. А., Петрова Е. Ю., Федоров Д. Ю., Хонина О. И.

Разработка и внедрение системы управления информационными киосками в университете: подходы, технологии. 52

Методы и системы защиты информации, информационная безопасность

Ковтуненко А. А.

Разработка DLP-модуля для защиты персональных данных в распределенной автоматизированной системе. 65

Интеллектуальные транспортные системы

Зеленцов В. А., Мочалов В. Ф.

Показатели и алгоритмы оценивания качества результатов идентификации состояния прилегающих территорий 73

Contents

Artificial intelligence and machine learning

Samonov A. V.

Methods for Optimizing the Training and Fine-Tuning Large Language Models (in English) 5

Mathematical modeling, numerical methods and software complexes

Alimov B. M., Sultanov Sh. Kh., Kodirova E. V.

Mathematical Modeling of the Planetary Mechanism for the Drive of Mixing Machines. 13

Chernyakhovskaya E. S.

Mathematical Model of Product Allocation on Warehouse Rack with Simultaneous Horizontal and Vertical Product Categorization 21

Lipanov I. D., Khomonenko A. D.

Technologies and Methods for Planning the Movement of UAVs Along Waypoints. 30

System analysis, management and information processing, statistics

Sokolov S. S., Antonova A. E.

Application of Simulation Modeling in Ship Construction Processes 44

Mathematical and software of computer systems, complexes and computer networks

Zabrodin A. V., Dobrovolskaya M. A., Petrova E. Yu., Fedorov D. Yu., Honina O. I.

Development and Implementation of a Management System for Information Kiosks at the University: Approaches, Technologies 52

Methods and systems of information protection, information security

Kovtunenکو A. A.

Development of a DLP Module to Protect Personal Data in a Distributed Automated System 65

Intelligent transport systems

Zelentsov V. A., Mochalov V. F.

Indicators and Algorithms for Assessing the Identification Results Quality of the Adjacent Territories State 73

УДК 004.85

Methods for Optimizing the Training and Fine-Tuning Large Language Models

A. V. Samonov — PhD in Engineering. Research interests: system analysis, computer science, system and software engineering, methods and means of information security. E-mail: a.samonov@mail.ru

Mozhaisky Military Aerospace Academy, 13 Zhdanovskaya str., St. Petersburg, 197198, Russia

For citation: Samonov A. V. Methods for Optimizing the Training and Fine-Tuning Large Language Models // Intellectual Technologies on Transport. 2024. No. 3 (39). Pp. 5–12. DOI: 10.20295/2413-2527-2024-339-5-12

Abstract. *The main problematic issues in the development and specialization of LLM are: catastrophic forgetting, the risk of overfitting, hallucinations, incorrect interpretations, incorrect processing of exceptional situations as well as exceptionally high performance requirements for the computing tools used in this case. The purpose of the study is to select and develop methods for optimizing the training and fine-tuning process LLM, providing a significant reduction in the computing resources required for this. To achieve this goal, it is proposed to use the following methods of optimizing LLMs and their learning algorithms: LoRA and QLoRA, Batch size choice, Gradient Accumulation, Gradient Checkpoint, Mixed precision training, FlashAttention-2. To obtain a cumulative positive effect when using these methods together, it is necessary to perform a number of practical experiments. When setting up LLM learning hyperparameters, you should first determine which package size gives the best results, and then choose adequate methods to optimize the computing resources used. The application of the presented methods will increase the efficiency of using computing resources when training and fine-tuning large language models and will reduce the time and financial costs necessary for this.*

Keywords: *fine-tuning, gradient accumulation, graphics processing unit, Large Language Model, Low-Rank Adaptation, mixed precision*

Introduction

The current stage of global development is characterized by the active introduction of artificial intelligence technologies into industry, science, education and other spheres of economic and social life, vivid examples of which are deep learning methods and generative artificial intelligence. Large language models (LLM) created with their help are capable of processing and creating texts, understanding and synthesizing speech, images, generating program code, solving analytical, mathematical and other non-trivial tasks. The main problematic issues in the development and specialization of LLM are: catastrophic forgetting, the risk of overfitting, hallucinations, incorrect interpretations, correct

handling of exceptional situations, ensuring the integrity of models, exceptionally high performance requirements for computing tools used in LLM training.

This article discusses methods for optimizing the training and fine-tuning process LLM, which ensure a significant reduction in the computing resources required for this with minimal loss of quality of the created models.

Stages of development of specialized systems based on large language models

The technological chain of the process of creating specialized software systems based on large language models includes three main stages: the development

of a basic model, preliminary specialization and fine-tuning to solve problems in a specific subject area.

At the initial stage, the model is trained on unstructured and unlabeled data. The main data sources that were used in the development of most modern LLMs are: Wikipedia, Common Craw, BooksCorpus (a collection of book texts), OpenWebText (a set of articles from the Internet). The result is a basic general purpose LLM. Examples of such models are GPT 4, GPT 3.5, Claude2, Gemini, Falcon, Llama, T5 [1].

At the second stage, the LLM is finalized through self-study on specially prepared data, which adjusts the model to solve problems of a certain class. At the third stage, such models undergo additional training with reinforcement based on expert feedback (RLHF, Reinforcement Learning from Human Feedback). Such open-source LLMs as Llama, Vicuna 13B, Mixtral 7B, T5 can be used to create specialized LLMs [1, 2]. As a result, domain-specific AI systems are being created, designed to solve certain tasks in specific areas of application.

LLM training and fine-tuning require high-performance computers, large amounts of RAM and disk memory, and graphics processing unit — GPU. During LLM training, GPU memory is used to store model weights, optimizer states, gradients (parameter derivatives), and direct activations stored to calculate gradients. When calculating weight coefficients in fp32 format, 4 bytes of GPU memory are required per LLM parameter. Thus, to load a model with three billion parameters, 12 GB of memory ($4 \text{ bytes} \times 3000000$) will be required. In addition, an additional 6 GB ($2 \text{ bytes} \times 3000000$) will be required to store the states of the AdamW 8-bit optimizer, and 12 GB of memory ($4 \text{ bytes} \times 3000000$) will be required to accommodate gradients. Thus, it turns out that to train a model with three billion parameters, at least 30 GB of RAM will be required.

The main characteristics of the GPU are: memory capacity, performance on special floating-point tasks, scalability, virtualization support, power consumption and price. Depending on the characteristics, the cost of the GPU varies from hundreds of thousands (4090 ADA 24 GB 3.5 Slot FP16 = 83 TFLOPS) to several million rubles (H100 Hopper 80 GB, FP16 = 204.9 TFLOPS) [3, 4]. At the same time, according to

estimates [2], to train LLM with 3 billion parameters, it is necessary to use 128 GPU A100 40 GB for 7 days. Given that modern LLMs have tens and hundreds of billions of parameters, it becomes obvious that there is an urgent need to develop methods to optimize the processes of their training, configuration and use.

The research task

In connection with the above, it becomes obvious that there is an urgent need to develop and apply less expensive methods and tools for training and configuring LLM that can implement these processes without significantly reducing the quality of the intelligent systems created at the same time. The following subsections of the article provide a brief description of such methods and tools, as well as suggestions for their optimal use.

Method for fine-tuning large language models

In order to reduce the performance requirements of computing tools used in LLM training, methods Parameter-Efficient Fine-Tuning (PEFT) have been developed and are being used. A description of modern methods and tools for fine-tuning LLM is presented in [5–9]. The software implementation of these methods is presented in the *peft* library on site *huggingface.co*.

Methods for fine-tuning effective LLM parameters, unlike full model tuning, provide training for only a small set of parameters (PEFT), which can be a subset of existing model parameters or a set of added parameters. Usage PEFT methods allows you to reduce training time, reduce the cost of computing and storing models, reduce the risks of overtraining, overcome catastrophic forgetting, correctly handle exceptional situations, and ensure ease of deployment and transfer to other devices.

Currently, the most promising methods of fine tuning are: LoRA (Low-Rank Adaptation) and QLoRA (Quantization-Aware LoRA). The LoRA method focuses on changing the weights of only certain layers and parameters of the basic LLM, focusing on those that are most useful for solving problems of this class. This is achieved by applying matrices with significantly lower rank than the matrices of the basic model to adjust the weights.

LoRa applies only to transformer query and value matrices, which means that the multilayer perceptron is frozen and only attention weights are adapted. The loss function is optimized by passing the gradient through the frozen model to the adapters. The formula describing the LoRA method in tensor notation has the following form:

$$W0 + \Delta W = W0 + BA,$$

where $W0$ is the weight matrix of the pre-trained model;

ΔW is the updated and added weight coefficients during the adaptation of the original model;

$A \in Rr \times k$ is a matrix of size $r \times k$, the elements of which are random variables corresponding to the normal distribution law $N(\mu, \sigma^2)$, where $\mu = 0$ (the average value of the value), σ is the standard deviation; $B \in Rd \times r$ is a matrix of size $d \times r$, the elements of which are assigned zeros at the initial stage of training.

An important advantage of LoRa is the ability to use the same model for different tasks by replacing the weights in matrices A and B, reducing the amount of memory needed to store different models.

The QLoRA method (LoRA with quantization) is designed to deploy models in environments with limited resources. It allows you to significantly reduce the requirements for the necessary amounts and performance of GPU and CPU, as well as computing power, for deploying and configuring models. QLoRA is a modification of the LoRA method by quantifying the model parameters, i.e. reducing the accuracy of the weighting coefficients,

while maintaining the necessary correctness and performance.

The number of parameters is determined by the rank and shape of the original weights. In practice, trainable parameters vary as low as 0.1 % to 1 % of all the parameters. As the number of parameters needing fine-tuning decreases, the size of gradients and optimizer states attached to them decrease accordingly. Thus, the overall size of the loaded model reduces. For example, the Llama 2 7B model parameters could be loaded in int8 (1 byte), with 1 GB trainable parameters loaded in fp16 (2 bytes). Hence, the size of the gradient (fp16), optimizer states (fp32), and activations (fp32) aggregates to approximately 7–9 GB. This brings the total size of the loaded model to be fine-tuned to 15–17 GB, as illustrated in Fig. 1.

Thus, thanks to the use of LoRA, it was possible to reduce the amount of memory required to configure Llama 2 7B by 4 times.

Methods for optimizing the learning process and configuring large language models

In order to increase the efficiency of using computing resources when training large language models to solve problems in a certain subject area, the following approaches and methods are currently used: Gradient Accumulation, Gradient Checkpoint, Mixed precision training, FlashAttention-2 [10–12].

The Gradient Accumulation method provides the calculation of gradients in smaller increments instead of calculating them for the entire batch at once. Iterative calculation of gradients is performed in small batches by performing forward and reverse

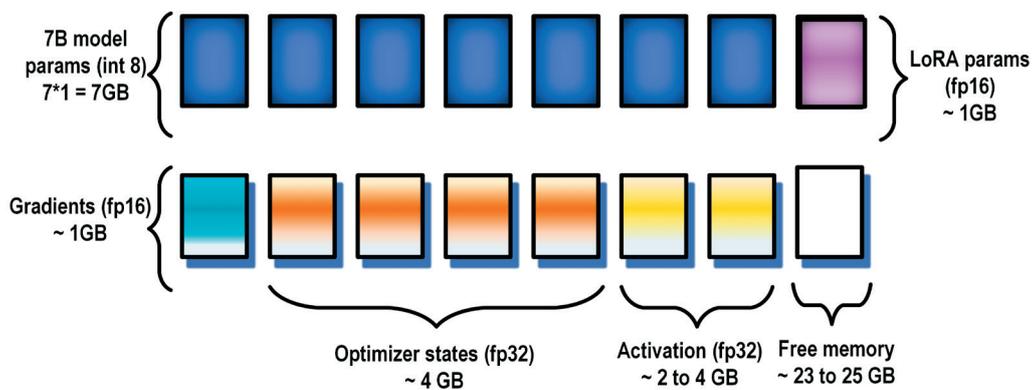


Fig. 1. Schematic showing an example of memory footprint of LoRA fine tuning with Llama 2 7B model

passes through the model and accumulating gradients in the process. Once a sufficient number of gradients have been accumulated, the model is optimized. Using this method, the effective packet size can be increased beyond the limits imposed by the amount of memory of the GPU. At the same time, it should be borne in mind that additional passes forward and backward, implemented in the process of gradient accumulation, can slow down the learning process. To use this method when training a model, the *gradient_accumulation_steps* argument must be included in the *TrainingArguments* configuration file:

```
training_args = TrainingArguments(per_device_train_batch_size=1, gradient_accumulation_steps=4, **default_args)
```

Using the “Gradient Accumulation” method allows you to maximize the use of GPU resources. Examples and results of the application of the “*gradient accumulation*” method are presented in [12].

To save all forward pass activations, significant amounts of memory must be allocated to calculate gradients during the reverse pass. If you do not save the activations, then their re-calculation during the reverse passage through the graph of the model will lead to significant computational costs and slow down the learning process. The “*Gradient Checkpointing*” method offers a compromise between these two approaches and preserves strategically important activations for the entire computational graph at certain control points. Due to this, only a part of the activations needs to be calculated again. To use this method, the “*gradient_checkpointing=True*” argument must be included in the *TrainingArguments* configuration file. The use of this method increases the efficiency of memory usage, but slows down learning by about 20 % [12].

Using the “*Mixed precision training*” method, the efficiency of the model training process is increased by using lower precision numerical formats for certain variables. Most models use 32-bit floating-point precision (fp32 or float32) to represent and process variables.

However, not all variables require such a high level of accuracy to achieve good results. By reducing the precision of some variables, for example, to 16-bit floating point values (fp16 or float16), calculations can be accelerated. The main advantage of mixed-precision

learning is the storage of half-precision activations (fp16). Although gradients are also calculated with half accuracy, they are converted back to full accuracy during the optimization stage. Therefore, there is no memory saving in this case.

Thus, learning with mixed accuracy, on the one hand, leads to faster calculations, and on the other hand, it can lead to an increase in the amount of GPU memory used, especially with small packet sizes. This is due to the fact that the model is now present on the GPU with both 16-bit and 32-bit precision, i.e. 1.5 times more than the original model. The scheme of the algorithm mixed precision method, used for calculating LLM parameters is shown in Fig. 2.

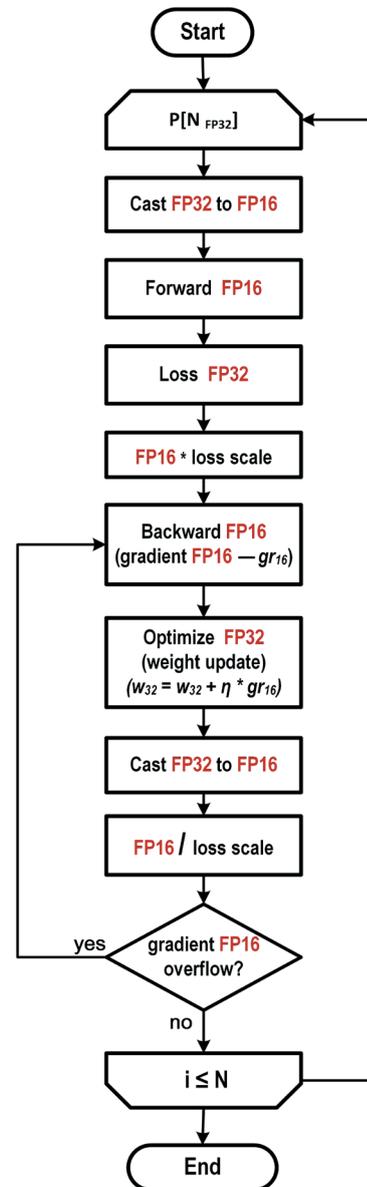


Fig. 2. Scheme of the algorithm mixed precision method

The parameters $P[N_{FP32}]$ received at the input of the algorithm in FP32 format are converted to FP16 format. The loss level is set to FP32. During backward computation, the value is multiplied by the loss scale to avoid overflow due to a small gradient value. A parameter in FP16 format is used to calculate the gradient, and the result is converted to FP32. Then the value is divided by the loss scale to restore the multiplied gradient. The optimizer checks if the gradient is overflowing. If yes, the optimizer skips the update. If not, the optimizer uses FP32 to update the initial parameters. To use the mixed precision method, you must set the “*fp16=True*” parameter in the *Training Arguments* configuration file.

FlashAttention-2 is a faster and more efficient implementation of the standard attention mechanism, which can significantly speed up logical inference due to:

- additional parallelization of attention calculations along the length of the sequence;
- separation of work between GPU threads to reduce data exchange and read/write operations in shared memory between them.

FlashAttention-2 can only be used if the model format is fp16 or bf16. To use the FlashAttention-2 method, the “*attn_implementation=“flash_attention_2”*” parameter must be included in the model description. FlashAttention-2 can be combined with other optimization methods such as quantization. Below is an example of using this method in combination with 8-bit quantization:

```
# load in 8bit
model = AutoModelForCausalLM.from_pretrained
(model_id,
load_in_8bit=True,
attn_implementation="flash_attention_2")
```

Reducing the requirements for the amount of GPU memory required for LLM training can be achieved by choosing the optimal optimizer. Below are the GPU memory requirements required by three different optimizers when learning LLM with 3 billion parameters [10]:

- the standard AdamW optimizer will require 24 GB of GPU memory, since 8 bytes are used for each parameter ($8 \times 3 \Rightarrow 24$ GB);
- adafactor optimizer will require more than 12 GB, because a little more than 4 bytes are used for each parameter;

- an 8-bit quantized BNB optimizer will use only $(2 \times 3) \times 6$ GB if all the states of the optimizer are quantized.

Adafactor does not store moving averages for each element in weighting matrices. Instead, it stores aggregated information (sums of moving averages by rows and columns), which significantly reduces the memory used.

However, compared to Adam, Adafactor may have slower convergence in some cases. Which optimizer will be used is determined in the *Training Arguments* configuration file using the “*optim=“adafactor”*” parameter.

In combination with other approaches (gradient accumulation, gradient checkpoint detection, and training with mixed accuracy), you can get a three-fold reduction in the size of the required memory while maintaining bandwidth. However, as mentioned earlier, the convergence coefficient of Adafactor may be worse than Adam.

An example of program implementing a learning cycle of a large language model

This section provides an example of a program that uses the methods described above to training basic LLM using the functions of the Accelerate library of the Pytorch framework. The configuration file includes a description of the following parameters of the learning process:

```
training_args = TrainingArguments
(per_device_train_batch_size=1,
gradient_accumulation_steps=4,
gradient_checkpointing=True
fp16=True,
**default_args)
```

A fragment of a program that implements a learning cycle using the Accelerator module:

```
from accelerate import Accelerator
from torch.utils.data.dataloader import DataLoader
dataloader = DataLoader(ds, batch_size=training_args.per_device_train_batch_size)
if training_args.gradient_checkpointing:
model.gradient_checkpointing_enable()
accelerator = Accelerator(fp16=training_args.fp16)
model, optimizer, dataloader = accelerator.prepare(model, adam_bnb_optim, dataloader)
```

First, we load the training dataset using the DataLoader data loader. To optimize the learning process of the model, we use the `gradient_checkpointing_enable()` method and the mixed precision learning mode — `fp16`. In the call to the `prepare` method, it is determined that the data loader will be distributed between processes if we use several GPUs, and the 8-bit `adam_bnb_optim` optimizer will be used for training. Below is a fragment of the program that implements the main learning cycle:

```
model.train ()
for step, batch in enumerate (dataloader, start=1):
    loss = model(**batch).loss
    loss = loss / training_args.gradient_accumulation_
steps
accelerator.backward(loss)
if step % training_args.gradient_accumulation_
steps == 0:
    optimizer.step()
    optimizer.zero_grad()
```

Step is the number of accumulation steps, *batch* is the batch size assigned in the dataloader. The error back propagation method is performed using the `accelerator.backward(loss)`. Gradient accumulation is performed as follows: we normalize the losses, get the average value at the end of accumulation, and as soon as we have enough steps, we start the model learning process – optimizing the model weights using the `optimizer.step()`. `optimizer.zero_grad()` – resets the gradients of all optimized tensors (weight coefficients).

The methods discussed in this article allow you to reduce the learning time and increase efficiency the use of GPU and CPU. To obtain a cumulative positive effect when using these methods together, it is necessary

to plan and perform a number of practical experiments. When setting up LLM learning hyperparameters, you should first determine which package size gives the best results, and then choose adequate methods to optimize the computing resources used. Examples and results of such studies and experiments are presented in [12–15].

Conclusion

This article presents methods to reduce the requirements for the number and performance of computing tools necessary for teaching large language models by optimizing models and algorithms for their training, as well as methods to increase the efficiency of using available computing resources when they are fine-tuned to solve problems in a specific subject area. The most significant results presented in the article, from the point of view of scientific novelty and practical significance, are:

1. Methodological recommendations for the use of LoRA (Low-Rank Adaptation) and QLoRA (Quantization-Aware LoRA) methods for fine-tuning large language models, which provide a significant (by an order of magnitude or more) reduction in computing performance requirements.

2. Algorithmic and software support of LLM learning processes using the methods Gradient Accumulation, Gradient Checkpoint, Mixed precision training and FlashAttention-2, which provides an increase in the efficiency of using available computing resources when teaching and using large language models to solve problems in a specific subject area. With the complex and correct application of these methods, it is possible to obtain a threefold reduction in the size of the required memory while maintaining bandwidth and the quality of the results obtained.

REFERENCES

1. A Survey of Large Language Models / W. Zhao [et al.] // ArXiv. 2023. Vol. 2303.18223. 124 p. DOI: 10.48550/arXiv.2303.18223
2. Down to Scale Up: A Guide to Parameter-Efficient Fine-Tuning / V. Lialin [et al.] // ArXiv. 2023. Vol. 2303.15647. 21 p. DOI: 10.48550/arXiv.2303.15647
3. Matrix Multiplication Background User's Guide // NVIDIA Documentation Hub. URL: <http://docs.nvidia.com/deeplearning/performance/dl-performance-matrix-multiplication> (accessed 26 Mar 2024).
4. Bekman S. Benchmarking Transformers with HF Trainer on a Single A100 40GB // Github. URL: <http://github.com/huggingface/transformers/issues/15026> (accessed 26 Mar 2024).

5. LORA: Low-Rank Adaptation of Large Language Models / E. Hu [et al.] // ArXiv. 2021. Vol. 2106.09685. 26 p. DOI: 10.48550/arXiv.2106.09685
6. LLaMA-Adapter: Efficient Fine-Tuning of Language Models with Zero-Init Attention / R. Zhang [et al.] // ArXiv. 2023. Vol. 2303.16199. 22 p. DOI: 10.48550/arXiv.2303.16199
7. Delta Tuning: A Comprehensive Study of Parameter Efficient Methods for Pre-Trained Language Models / N. Ding [et al.] // ArXiv. 2022. Vol. 2203.06904. 49 p. DOI: 10.48550/arXiv.2203.06904
8. QA-LoRA: Quantization-Aware Low-Rank Adaptation of Large Language Models / Y. Xu [et al.] // ArXiv. 2023. Vol. 2309.14717. 16 p. DOI: 10.48550/arXiv.2309.14717
9. QDyLoRA: Quantized Dynamic Low-Rank Adaptation for Efficient Large Language Model Tuning / H. Rajabzadeh [et al.] // ArXiv. 2024. Vol. 2402.10462. 6 p. DOI: 10.48550/arXiv.2402.10462
10. Methods and Tools for Efficient Training on a Single GPU // Hugging Face Community. URL: http://huggingface.co/docs/transformers/perf_train_gpu_one (accessed 26 Mar 2024).
11. Goodfellow I., Bengio Y., Courville A. Optimization for Training Deep Model // Deep Learning. Cambridge (MA): MIT Press, 2016. Pp. 267–320.
12. Bekman S. Benchmarking Transformers with HF Trainer on RTX-309 // Github. URL: <http://github.com/huggingface/transformers/issues/14608> (accessed 26 Mar 2024).
13. Linear/Fully Connected Layers User’s Guide // NVIDIA Documentation Hub. URL: <http://docs.nvidia.com/deeplearning/performance/dl-performance-fully-connected> (accessed 26 Mar 2024).
14. Exploring Parameter-Efficient Fine-Tuning Techniques for Code Generation with Large Language Models / M. Weyssow [et al.] // ArXiv. 2023. Vol. 2308.10462. 23 p. DOI: 10.48550/arXiv.2308.10462
15. PTraining FP8 Large Language Models / H. Peng [et al.] // ArXiv. 2023. Vol. 2310.18313. 23 p. DOI: 10.48550/arXiv.2310.18313

Received: 18.07.2024

Accepted: 13.08.2024

Методы оптимизации процесса обучения и тонкой настройки больших языковых моделей

А. В. Самонов — канд. техн. наук. Область научных интересов: системный анализ, информатика, математическое и имитационное моделирование и разработка сложных программно-технических систем, методы и средства обеспечения информационной безопасности. E-mail: a.samonov@mail.ru

Военно-космическая академия имени А. Ф. Можайского, Россия, 197198, Санкт-Петербург, ул. Ждановская, 13

Для цитирования: Самонов А. В. Методы оптимизации процесса обучения и тонкой настройки больших языковых моделей // Интеллектуальные технологии на транспорте. 2024. № 3 (39). С. 5–12. DOI: 10.20295/2413-2527-2024-339-5-12

Аннотация. Основными проблемными вопросами при разработке и специализации больших языковых моделей (Large Language Model — LLM,) являются катастрофическое забывание, риск переобучения, галлюцинации, некорректная обработка исключительных ситуаций, а также исключительно высокие требования к производительности используемых при этом вычислительных средств. Целями исследования являются выбор и разработка методов оптимизации процесса обучения и настройки LLM, обеспечивающих существенное снижение необходимых для этого вычислительных ресурсов. Для достижения

данной цели предложено использовать следующие методы оптимизации LLM и алгоритмов их обучения: *LoRA* и *QLoRA*, *Batch size choice* (выбор оптимального размера пакета), *Gradient Accumulation* (накопление градиента), *Gradient Checkpointing* (контрольные точки градиента), *Mixed precision training* (смешанная точность), *FlashAttention 2*. Для получения кумулятивного положительного эффекта при совместном использовании этих методов необходимо выполнить ряд практических экспериментов. При настройке гиперпараметров обучения LLM сначала следует определить, какой размер пакета дает наилучшие результаты, а затем выбрать адекватные методы оптимизации используемых вычислительных ресурсов. Применение представленных методов позволит повысить эффективность использования вычислительных ресурсов при настройке больших языковых моделей и обеспечит сокращение необходимых для этого временных и финансовых затрат.

Ключевые слова: большая языковая модель, графический процессор, накопление градиента, смешанная точность, точная настройка LLM, *Large Language Model*, *Low-Rank Adaptation*

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. A Survey of Large Language Models / W. Zhao [et al.] // ArXiv. 2023. Vol. 2303.18223. 124 p. DOI: 10.48550/arXiv.2303.18223
2. Down to Scale Up: A Guide to Parameter-Efficient Fine-Tuning / V. Lialin [et al.] // ArXiv. 2023. Vol. 2303.15647. 21 p. DOI: 10.48550/arXiv.2303.15647
3. Matrix Multiplication Background User's Guide // NVIDIA Documentation Hub. URL: <http://docs.nvidia.com/deeplearning/performance/dl-performance-matrix-multiplication> (accessed 26 Mar 2024).
4. Bekman S. Benchmarking Transformers with HF Trainer on a Single A100 40GB // Github. URL: <http://github.com/huggingface/transformers/issues/15026> (accessed 26 Mar 2024).
5. LORA: Low-Rank Adaptation of Large Language Models / E. Hu [et al.] // ArXiv. 2021. Vol. 2106.09685. 26 p. DOI: 10.48550/arXiv.2106.09685
6. LLaMA-Adapter: Efficient Fine-Tuning of Language Models with Zero-Init Attention / R. Zhang [et al.] // ArXiv. 2023. Vol. 2303.16199. 22 p. DOI: 10.48550/arXiv.2303.16199
7. Delta Tuning: A Comprehensive Study of Parameter Efficient Methods for Pre-Trained Language Models / N. Ding [et al.] // ArXiv. 2022. Vol. 2203.06904. 49 p. DOI: 10.48550/arXiv.2203.06904
8. QA-LoRA: Quantization-Aware Low-Rank Adaptation of Large Language Models / Y. Xu [et al.] // ArXiv. 2023. Vol. 2309.14717. 16 p. DOI: 10.48550/arXiv.2309.14717
9. QDyLoRA: Quantized Dynamic Low-Rank Adaptation for Efficient Large Language Model Tuning / H. Rajabzadeh [et al.] // ArXiv. 2024. Vol. 2402.10462. 6 p. DOI: 10.48550/arXiv.2402.10462
10. Methods and Tools for Efficient Training on a Single GPU // Hugging Face Community. URL: http://huggingface.co/docs/transformers/perf_train_gpu_one (accessed 26 Mar 2024).
11. Goodfellow I., Bengio Y., Courville A. Optimization for Training Deep Model // Deep Learning. Cambridge (MA): MIT Press, 2016. Pp. 267–320.
12. Bekman S. Benchmarking Transformers with HF Trainer on RTX-309 // Github. URL: <http://github.com/huggingface/transformers/issues/14608> (accessed 26 Mar 2024).
13. Linear/Fully Connected Layers User's Guide // NVIDIA Documentation Hub. URL: <http://docs.nvidia.com/deeplearning/performance/dl-performance-fully-connected> (accessed 26 Mar 2024).
14. Exploring Parameter-Efficient Fine-Tuning Techniques for Code Generation with Large Language Models / M. Weysow [et al.] // ArXiv. 2023. Vol. 2308.10462. 23 p. DOI: 10.48550/arXiv.2308.10462
15. PTraining FP8 Large Language Models / H. Peng [et al.] // ArXiv. 2023. Vol. 2310.18313. 23 p. DOI: 10.48550/arXiv.2310.18313.10.48550/arXiv.2310.18313

Дата поступления: 18.07.2024

Решение о публикации: 13.08.2024

УДК 621.01

Математическое моделирование бипланетарного механизма для привода смесительных машин

**Алимов
Бахтияр
Милябаевич**

— канд. техн. наук, доцент кафедры «Информатика и компьютерная графика». Область научных интересов: теория механизмов и деталей машин. E-mail: alimov46@internet.ru

**Султонов
Шохрух
Холмурзаевич**

— докт. философии, канд. наук, заведующий кафедрой «Информатика и компьютерная графика». Область научных интересов: информационные системы, обработка больших данных и моделирование. E-mail: sultonovsh@yandex.ru

**Кодирова
Елена
Владимировна**

— ассистент кафедры «Информатика и компьютерная графика». Область научных интересов: информационные системы, обработка больших данных и моделирование. E-mail: lena.kadirova@mail.ru

Ташкентский государственный транспортный университет, Республика Узбекистан, Ташкент, 1-й пр. Куддус, 1

Для цитирования: Алимов Б. М., Султонов Ш. Х., Кодирова Е. В. Математическое моделирование бипланетарного механизма для привода смесительных машин // Интеллектуальные технологии на транспорте. 2024. № 3 (39). С. 13–20. DOI: 10.20295/2413-2527-2024-339-13-20

Аннотация. Представлено исследование по разработке кинематической схемы планетарного механизма с размыкающей кинематической цепью для привода рабочего органа строительной мешалки миксера. **Основной целью** исследования является создание в планетарном механизме двух кинематических циклов, в которых в течение одного оборота водила сателлиты последовательно выполняют движения эпициклической траектории гипоциклического движения. Для этого в переходных зонах между двумя разноименными сегментами использованы тормозные устройства для гашения кинематической энергии сателлитов и устройства синхронизатора, а также плавного входа шестерни сателлита в зубчатое зацепление данных передач. Рассматривается колебательный процесс синхронизатора при входе зуба шестерни сателлита в зацепление с передними зубьями синхронизатора. **Методы:** в статье предлагаются теоретические разработки планетарного механизма для привода рабочего органа лопастей смесительных машин, которые осуществлялись по методике основных положений теории механизмов и машин, колебаний и устойчивости конструкции, технологических процессов работы строительных машин. Использовано уравнение Лагранжа при колебательном движении синхронизатора. Результаты получены при помощи методов математической статистики и компьютерной графики. **Практическая значимость:** обеспечение возможности сборки планетарного механизма с учетом выбора угла поворота синхронизатора вокруг оси опоры. В обсуждении высказываются рекомендации по дальнейшему совершенствованию разработанного бипланетарного механизма для привода смесительных машин.

Ключевые слова: колебание, частота, сателлит, водило, синхронизатор, момент инерции, планетарный механизм, смесительная машина, миксер, привод

Введение

В планетарных механизмах применяются передачи с зубчатыми, цевочными [1], фрикционно-ременными передачами [2], которые используются в приводе шпинделей аппарата хлопкоуборочной машины. Также планетарные механизмы имеют компактную конструкцию, что нашло широкое применение в народном хозяйстве: в растворосмесительных механизмах [3], пищевой промышленности, особенно при проектировании различных миксеров [4–7], в очистительных сооружениях и водоемах [8, 9] при вентилировании очистки сточных вод [10]. В работе [11] дана кинематическая схема эпициклического планетарного механизма для двигателя внутреннего сгорания с составлением уравнения кинематического движения сателлитов. В работах [12–14] приводится исследование планетарного механизма для привода рабочего органа лопастей строительной мешалки миксера.

Результаты и их обсуждение

В технологических процессах растворомешалки миксера применяют планетарные механизмы для привода рабочего органа лопастей смесительных машин. На рис. 1 приводится кинематическая схема бипланетарного механизма для привода рабочего органа лопастей смесительной машины. В бипланетарных механизмах сателлиты непрерывно реверсируются: за один оборот водила сателлиты рабочего органа лопастей дважды меняют направление вокруг своей оси. Передачи выполнены с зубчатыми сегментами внутреннего (1) и внешнего (2) зацепления, в переходной зоне α планетарного механизма установлен синхронизатор (6, 7) — между двух зубчатых сегментов 1 и 2 для плавного входа зубьев шестерни 3-го сателлита в зубчатое зацепление и тормозное устройство (4, 5) для реверса шестерни 3-го сателлита.

Бипланетарный механизм для привода рабочего органа лопастей смесительных машин работает следующим образом. При вращении водила по часовой стрелке с угловой скоростью ω_b шестерня 3 сателлита, выйдя из зубчатого зацепления внешнего сегмента с угловой скоростью ω_c , по инерции

продолжает вращаться по часовой стрелке и в зоне холостого участка α до тормозного устройства 4, где происходит реверс шестерни 3 сателлита, то есть шестерня сателлита меняет вращение с угловой скоростью ω_c против часовой стрелки. После реверса шестерня 3 соприкасается с первым зубом синхронизатора 6, который имеет измененный профиль зуба, способствующий точному входу зубьев шестерни 3 в зубчатое зацепление. Далее шестерня 3 сателлита, после прохождения по длине дуги синхронизатора 6, соединяется с зубчатым зацеплением сегмента 1 внутреннего зацепления. Аналогичный процесс происходит при подходе шестерни 3 сателлита к зоне зубчатого сегмента 2 внешнего зацепления: происходит торможение шестерни 3 сателлита тормозными устройствами 5, затем после реверса шестерни происходит соприкосновение зубчатых передач шестерни с синхронизатором, и в дальнейшем шестерня сателлита соприкасается с зацеплением сегмента 2 внешнего зацепления.

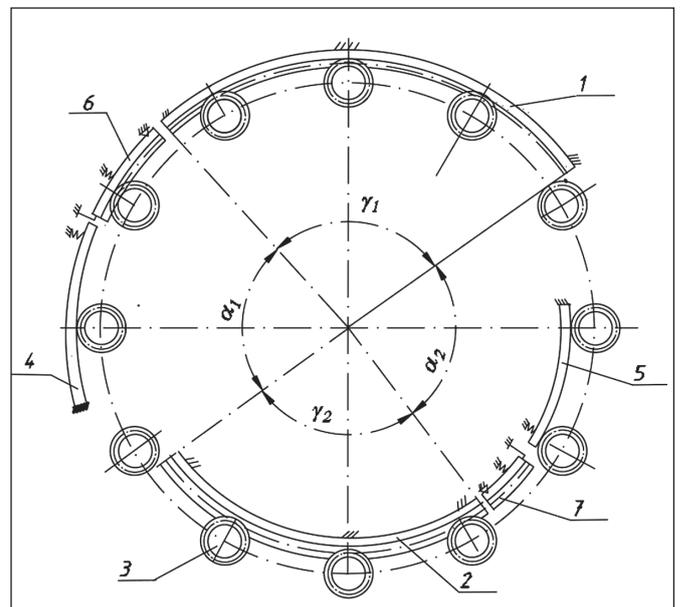


Рис. 1. Схема бипланетарного механизма с размыканием кинематической цепи

Силу F , действующую на зуб синхронизатора, находим по экспериментальным данным, учитывая периодический характер колебаний системы под действием внешних сил шестерни сателлита при частоте $n = 110$ об/мин вращательного движения

водила бипланетарного механизма, воздействующего на синхронизатор, который является импульсным, описывается полусинусоидальной кривой и представляет собой возмущающую силу F . Возмущающая сила F шестерни сателлита, воздействующая на передний зуб синхронизатора, равна:

$$F = F_A \cdot \sin(\pi t_1 / t),$$

где F_A — осевое усилие, действующее в зубчатом зацеплении цилиндрических передач, $F_A = 8\text{Н}$; t_1 — переменная величина времени срабатывания синхронизатора, с ($0 < t_1 < t$);

t — время срабатывания синхронизатора, с.

Подставляя численные значения и решая уравнение в MathCAD, получим график колебаний действительного возбуждения системы в течение времени срабатывания синхронизатора $t = 0,048$ с (рис. 2).

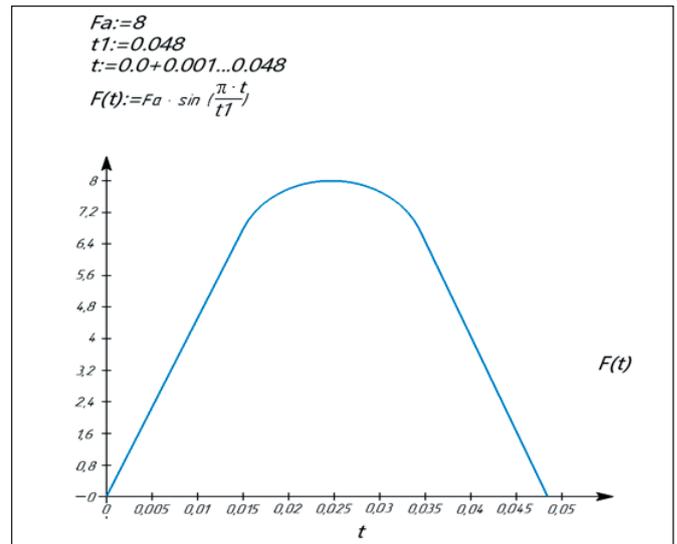


Рис. 2. График кривой попадания зуба сателлита по зубу синхронизатора

Возмущающую силу можно разложить в ряд Фурье, тогда

$$F = F_A \left[\frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \left(\frac{\cos 2\omega t}{3} + \frac{\cos 4\omega t}{15} + \frac{\cos 6\omega t}{35} \right) \right]. \quad (1)$$

Здесь ω — угловая скорость сателлита, $\omega = 130,08$ рад/с.

Используя уравнение Лагранжа, получим уравнение колебаний синхронизатора:

$$\mathcal{I}_D \cdot \ddot{\varphi} + c l^2 \cdot \varphi = M_{sp}, \text{ или } \mathcal{I}_D \cdot \ddot{\varphi} + c l^2 \cdot \varphi = F_t \cdot l. \quad (2)$$

Подставляя значения из уравнения (1) в (2), получим

$$\mathcal{I}_D \cdot \ddot{\varphi} + c l^2 \cdot \varphi = F_A \cdot l \left[\frac{1}{4} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \left(\frac{\cos 2\omega t}{3} + \frac{\cos 4\omega t}{4} + \frac{\cos 6\omega t}{35} \right) \right],$$

или

$$\varphi + \frac{c l^2}{\mathcal{I}_D} \varphi = \frac{F_A \cdot l}{\mathcal{I}_D} \left[\frac{1}{4} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \left(\frac{\cos 2\omega t}{3} + \frac{\cos 4\omega t}{4} + \frac{\cos 6\omega t}{35} \right) \right]. \quad (3)$$

Уравнение без правой части, соответствующее уравнению (3),

$$\ddot{\varphi} + \frac{c l^2}{\mathcal{I}_D} \cdot \varphi = 0, \text{ и характеристическое уравнение } r^2 + \frac{c l^2}{\mathcal{I}_D} = 0,$$

корни которого равны $r = \pm p_i$, где p_i — собственная частота системы, $p_i = l \sqrt{\frac{c}{\mathcal{I}_D}}$.

Отсюда $\varphi_1 = \cos pt$, $\varphi_2 = \sin pt$.

Запишем решение уравнения (3) в следующем виде:

$$\varphi = c_1(t) \cos pt + c_2(t) \sin pt.$$

Составим систему уравнений для отыскания $c_1(t)$ и $c_2(t)$:

$$\dot{c}_1(t) \cos pt + \dot{c}_2(t) \sin pt = 0,$$

тогда

$$-p \dot{c}_1(t) \cdot \sin pt + p \dot{c}_2(t) \cdot \cos pt = \frac{F_A \cdot l}{\mathcal{I}_D} \left[\frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \left(\frac{\cos 2\omega t}{3} + \frac{\cos 4\omega t}{15} + \frac{\cos 6\omega t}{35} \right) \right].$$

Решая эту систему, находим: $\dot{c}_1(t) = -\dot{c}_2(t) \cdot \tan pt$;

$$\begin{aligned} \dot{c}_2(t) &= \dot{c}_2(t) \cdot \sin pt \cdot \tan pt + \dot{c}_2(t) \cos pt = \\ &= \frac{F_A \cdot l}{I_D \cdot p} \left[\frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \left(\frac{\cos 2\omega t}{3} + \frac{\cos 4\omega t}{15} + \frac{\cos 6\omega t}{35} \right) \right]. \end{aligned}$$

Далее находим:

$$\begin{aligned} c_2(t) &= \frac{\cos pt}{\sin^2 pt + \cos^2 pt} \cdot \frac{F_A \cdot l}{I_D \cdot p} \left[\frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \left(\frac{\cos 2\omega t}{3} + \frac{\cos 4\omega t}{15} + \frac{\cos 6\omega t}{35} \right) \right]; \\ \dot{c}_1(t) &= \frac{F_A \cdot l \cdot \sin pt}{I_D \cdot p} \left[\frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \left(\frac{\cos 2\omega t}{3} + \frac{\cos 4\omega t}{15} + \frac{\cos 6\omega t}{35} \right) \right]. \end{aligned}$$

Интегрирование этих значений дает:

$$\begin{aligned} \dot{c}_1(t) &= \frac{F_A \cdot l}{I_D \cdot p} \int \sin pt \left[\frac{1}{4} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \left(\cos 2\omega t + \frac{1}{15} \cos 4\omega t + \frac{1}{35} \cos 6\omega t \right) \right] dt = \\ &= \frac{F_A \cdot l}{I_D \cdot p} \left\{ \frac{1}{\pi \cdot p} \cos pt - \frac{1}{4} \left[\frac{1}{p-\omega} \cdot \sin(p-\omega)t - \frac{1}{p+\omega} \sin(p+\omega)t \right] - \right. \\ &\quad \left. - \frac{1}{3\pi} \left[\frac{1}{p+2\omega} \cos(p+2\omega)t + \frac{1}{p+\cos 2\omega} (p-2\omega)t \right] - \right. \\ &\quad \left. - \frac{1}{15\pi} \left[\frac{1}{p+4\omega} \cos(p-4\omega)t - \frac{1}{35} \left[\frac{1}{p-6\omega} \cos(p-6\omega)t + \frac{1}{p-6\omega} \cos(p-6\omega)t \right] \right] \right\} + k_1. \end{aligned}$$

Также находим:

$$\begin{aligned} \dot{c}_2(t) &= \frac{F_A \cdot l}{I_D \cdot p} \int \cos pt \left[\frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \left(\frac{1}{3} \cos 2\omega t + \frac{1}{15} \cos 4\omega t + \frac{1}{35} \cos 6\omega t \right) \right] dt = \\ &= \frac{F_A \cdot l}{I_D \cdot p} \left\{ \frac{1}{\pi \cdot p} \sin pt + \frac{1}{4} \left[-\frac{1}{p+\omega} \cos(p+\omega)t + \frac{1}{p-\omega} \cos(p+\omega)t \right] - \right. \\ &\quad \left. - \frac{1}{3\pi} \left[\frac{1}{p-2\omega} \cdot \sin(p-2\omega)t + \frac{1}{p+2\omega} \sin(p-2\omega)t \right] + \right. \\ &\quad \left. + \frac{1}{15} \left[\frac{1}{p-4\omega} \sin(p-4\omega)t + \frac{1}{p+4\omega} \sin(p-4\omega)t + \right. \right. \\ &\quad \left. \left. + \frac{1}{35} \left[\frac{1}{p-6\omega} \sin(p-6\omega)t + \frac{1}{p-6\omega} \sin(p-6\omega)t \right] \right] \right\} + k_2, \end{aligned}$$

где k_1 и k_2 — произвольные интегрирования.

После подставления величин $c_1(t)$ и $c_2(t)$ в уравнение (5), получим угол поворота φ синхронизатора вокруг точки опоры D .

$$\begin{aligned} \varphi &= \frac{F_A \cdot l}{I_D \cdot p} \left\{ \frac{1}{\pi \cdot p} \cos pt - \frac{1}{4} \left[\frac{1}{\pi-\omega} \sin(p-\omega)t - \frac{1}{p+\omega} \sin(p-\omega)t \right] - \right. \\ &\quad \left. - \frac{1}{3\pi} \left[\frac{1}{p-2\omega} \cos(p-4\omega)t - \frac{1}{p-2\omega} \cos(p-2\omega)t \right] - \right. \\ &\quad \left. - \frac{1}{15} \left[\frac{1}{p-4\omega} \cos(p-4\omega)t + \frac{1}{p-4\omega} \cos(p-4\omega)t \right] - \right. \\ &\quad \left. - \frac{1}{35\pi} \left[\frac{1}{p-6\omega} \cos(p-6\omega)t + \frac{1}{p-6\omega} \cos(p-6\omega)t \right] \right\} \cos pt + k_1 \cos pt + \\ &\quad + \frac{F_A \cdot l}{I_D \cdot p} \left\{ \frac{1}{\pi \cdot p} \sin pt + \frac{1}{4} \left[\frac{1}{p-\omega} \cos(p-\omega)t - \frac{1}{p-\omega} \cos(p+\omega)t \right] - \right. \end{aligned}$$

$$-\frac{1}{35}\left[\frac{1}{p-2\omega}\sin(p-2\omega)t+\frac{1}{p+2\omega}\sin(p+2\omega)t\right]-$$

$$-\frac{1}{15\pi}\left[\frac{1}{p-4\omega}\sin(p-4\omega)t+\frac{1}{p+4\omega}\sin(p+4\omega)t\right]-$$

$$-\frac{1}{35\pi}\left[\frac{1}{p-6\omega}\sin(p-6\omega)t+\frac{1}{p+6\omega}\sin(p+6\omega)t\right]\}\sin pt+k_2\sin pt.$$

После соответствующих преобразований получим:

$$\varphi = \frac{F_A \cdot l}{I_D \cdot p} \left\{ \frac{1}{\pi \cdot p} + \frac{p \sin \omega t}{2(p^2 - \omega^2)} - \frac{2p \cos 2\omega t}{3\pi(p^2 - 4\omega^2)} - \frac{2p \cos 4\omega t}{15\pi(p^2 - 16\omega^2)} - \frac{2p \cos 6\omega t}{35\pi(p^2 - 36\omega^2)} \right\} + k_1 \cos pt + k_2 \sin pt.$$

Угол перемещения синхронизатора:

$$\varphi = \frac{F_A \cdot l}{I_D \cdot p} \left\{ \frac{p\omega \cos \omega t}{2(p^2 - \omega^2)} - \frac{2p}{\pi} \left[\frac{-2\omega \sin 2\omega t}{3(p^2 - 4\omega^2)} - \frac{4\omega \sin 4\omega t}{15(p^2 - 16\omega^2)} - \frac{6\omega \sin 6\omega t}{35(p^2 - 36\omega^2)} \right] \right\} - k_1 p \sin pt + k_2 p \cos pt.$$

При $t = 0$ и $\varphi = 0$ для определения произвольных постоянных k_1 и k_2 решаем следующую систему уравнений:

$$\begin{cases} \frac{F_A \cdot l}{I_D \cdot p} \left\{ \frac{1}{\pi} - \frac{2p}{\pi} \left[\frac{1}{3(p^2 - 4\omega^2)} + \frac{1}{15(p^2 - 16\omega^2)} + \frac{1}{35(p^2 - 36\omega^2)} + k_1 \right] \right\} = 0 \\ \frac{F_A \cdot l}{I_D} \cdot \frac{\omega}{2(p^2 - \omega^2)} + k_2 p = 0. \end{cases} \tag{4}$$

Имеем:

$$k_1 = \frac{F_A \cdot l}{I_D \cdot p} \left\{ \frac{1}{\pi p} - \frac{2p}{\pi} \left[\frac{1}{3(p^2 - 4\omega^2)} + \frac{1}{15(p^2 - 16\omega^2)} + \frac{1}{35(p^2 - 36\omega^2)} \right] \right\},$$

$$k_2 = \frac{F_A \cdot l \omega}{2I_D p (p^2 - \omega^2)}.$$

После подстановки произвольных постоянных k_1 и k_2 в уравнение (4) получим форму колебаний угла поворота φ синхронизатора вокруг оси опоры при рабочем режиме привода шпинделей уборочного аппарата:

$$\varphi = \frac{F_A \cdot l}{I_D \cdot p} \left\{ \frac{(1 - \cos pt)}{\pi p} + \frac{p \sin \omega t}{2(p^2 - \omega^2)} - \frac{2p}{\pi} \left[\frac{\cos 2\omega t - \cos pt}{3(p^2 - 4\omega^2)} - \frac{\cos 4\omega t - \cos pt}{15(p^2 - 16\omega^2)} + \frac{\cos 6\omega t - \cos pt}{35(p^2 - 36\omega^2)} \right] - \frac{\omega \sin pt}{2(p^2 - \omega^2)} \right\}. \tag{5}$$

Практическим значением полученной формулы (5) является возможность использовать ее при сборке планетарного механизма, в котором синхронизатор в колебательном процессе со- блюдает нестеснительное движение в системе при последовательном расположении зубчатых передач по линии зацепления синхронизатора и сегмента.

Заключение

Приводится кинематическая схема бипланетарного механизма для привода рабочего органа смесительной растворомешалки с установлением в переходных зонах между двумя разноименными зубчатыми сегментами внутреннего и внешнего зацепления синхронизаторов и тормозных устройств.

Теоретические исследования показали, что при выборе подвижной схемы синхронизатора будет обеспечиваться работоспособность бипланетарного механизма с размыканием кинематической цепи. Дальнейшее исследование имеет важное значение для изготовления смесительных растворомешалок с применением планетарного механизма.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Киреев С. О., Ковалев В. Н. Структура, кинематика и геометрия планетарных передач с внецентроидным цевочным зацеплением. Новочеркасск: Новочеркасский гос. техн. ун-т, 1995. 98 с.
2. Матчанов Р. Д. Хлопкоуборочные машины (1929–2010 гг.). Ташкент: Фан ва технология, 2011. 354 с.
3. Coulissee Mechanism with Rotating Link for Operating Part Drive Unit of the Mortar Pump / B. Alimov [et al.] // Construction Mechanics, Hydraulics and Water Resources Engineering (CONMECHYDRO–2021): Proceedings of the International Scientific Conference (Tashkent, Uzbekistan, 1–3 April, 2021). E3S Web of Conferences. 2021. Vol. 264. Art. No. 05050. 7 p. DOI: 10.1051/e3sconf/202126405050.
4. Kinematic Analysis of a Planetary Mixer Mechanism / V. S. Bogdanov [et al.] // High-Tech and Innovations in Research and Manufacturing (HIRM-2019): Proceedings of the International Conference (Krasnoyarsk, Russia, 06 May 2019). Journal of Physics: Conference Series. 2019. Vol. 1353. Art. No. 012030. 6 p. DOI: 10.1088/1742-6596/1353/1/012030.
5. Kammel D. W. Design, Selection and Use of TMR Mixers. URL: <http://www.researchgate.net/publication/268272514> (дата обращения 27.05.2024).
6. Design and Analysis of Planetary Gearbox for Industrial Concrete Mixer / S. Dange [et al.] // International Journal of Latest Research in Engineering and Technology. 2016. Vol. 2, iss. 4–1. Pp. 41–45.
7. Valigi M. C., Gasperini I. A Preliminary Examination of Blades Wear in a Planetary Concrete Mixer. URL: <http://www.researchgate.net/publication/266458874> (дата обращения 27.05.2024).
8. Pawar P. V., Kulkarni P. R. Critical Review of Design of Planetary Gears and Gear Box // International Journal of Mechanical Engineering and Information Technology. 2015. Vol. 3, iss. 3. Pp. 1067–1073.
9. Methodology for Calculating Power Consumption of Planetary Mixers / Antsiferov S. I. [et al.] // Proceedings of the International Conference on Mechanical Engineering, Automation and Control Systems (Tomsk, Russia, 4–6 December 2017). IOP Conference Series: Materials Science and Engineering. 2018. Vol. 327, iss. 4. Art. No. 042007. 9 p. DOI: 10.1088/1757-899X/327/4/042007.
10. Sutar S. S., Sutar A. V., Rawal M. R. Torque Measurement in Epicyclic Gear Train // International Journal of Engineering and Advanced Technology. 2016. Vol. 5, iss. 6. Pp. 185–189.
11. Turanov Kh., Shaumarova M. Incorrect Application of the Epicycloid Equation to the Planetary Mechanism of the Cotton Harvester // Topical Problems of Green Architecture, Civil and Environmental Engineering 2019 (TPACEE 2019): Proceedings of the International Conference (Moscow, Russia, 19–22 November 2019). E3S Web of Conferences. 2020. Vol. 164. Art. No. 06034. 10 p. DOI: 10.1051/e3sconf/202016406034.
12. Alimov B. M., Abdikayimova G. Dynamic Calculation of Gears of the Planetary Mixer Mechanism // Problems in the Textile and Light Industry in the Context of Integration of Science and Industry and Ways to Solve Them (PTLICISIWS-2): Proceedings of the International and Practical Conference (Namangan, Uzbekistan, 04–05 May 2023). AIP Conference Proceedings. 2024. Vol. 3045. Art. No. 060004. DOI: 10.1063/5.0197543.
13. Epi and Hypocyclic Planetary Gear to Drive the Working Organ of the Mixer / B. M. Alimov [et al.] // APAM–2021: Proceedings of the International Conference on Actual Problems of Applied Mechanics (Samarkand, Uzbekistan, 27–29 October 2021). AIP Conference Proceedings. 2022. Vol. 2637. Art. No. 060010. DOI: 10.1063/5.0120095.

14. Rizayev A., Alimova D., Alimov B. Modeling Spindle Drive of the Cotton Harvester // ETESD–2022: Proceedings of the International Conference on Environmental Technologies and Engineering for Sustainable Development (Tashkent, Uzbekistan, 12–15 October 2022). IOP Conference Series: Earth and Environmental Science. 2022. Vol. 1112. Art. No 012042. 8 p. DOI: 10.1088/1755-1315/1112/1/012042.

Дата поступления: 10.06.2024

Решение о публикации: 17.07.2024

Mathematical Modeling of the Planetary Mechanism for the Drive of Mixing Machines

- Bakhtiyar M. Alimov** — PhD in Engineering, Associate Professor of the Department of Computer Science and Computer Graphics. Research interests: theory of mechanisms and machine parts. E-mail: alimov46@internet.ru
- Shokhrukh K. Sultanov** — Doctor of Philosophy, PhD, Head of the Department of Computer Science and Computer Graphics. Research interests: information systems, big data processing and modeling. E-mail: sultonovsh@yandex.ru
- Elena V. Kodirova** — assistant at the Department of Computer Science and Computer Graphics. Research interests: information systems, big data processing and modeling. E-mail: lena.kadirova@mail.ru

Tashkent State Transport University, 1st ave. Kuddus, 1, Tashkent, Republic of Uzbekistan

For citation: Alimov B. M., Sultanov Sh. Kh., Kadyrova E. V. Mathematical Modeling of the Planetary Mechanism for the Drive of Mixing Machines // Intellectual Technologies on Transport. 2024. No. 3 (39). P. 13–20. DOI: 10.20295/2413-2527-2024-339-13-20

Abstract. *A study on the development of a kinematic scheme of a planetary mechanism with an opening kinematic circuit for driving the working body of mixing machines is presented. The main purpose of the study is to create two kinematic cycles in the planetary mechanism, in which, during one revolution, the satellite drivers consistently perform movements of the epicyclic trajectory of hypocyclic motion. To achieve this goal, braking devices for damping the kinematic energy of satellites and synchronizer devices were used in the transition zones between two segments of different names, for smooth entry of the satellite gear into the gear engagement of these gears. The oscillatory process of the synchronizer at the entrance of the tooth of the satellite gear with the front teeth of the synchronizer is considered. Practical significance includes ensuring the possibility of assembling a planetary mechanism, taking into account the choice of the angle of rotation of the synchronizer around the axis of the support. The discussion makes recommendations for further improvement of the developed biplanetary mechanism for driving mixing machines.*

Keywords: *oscillation, frequency, satellite, driver, synchronizer, moment of inertia, planetary mechanism, mixing machine, mixer, drive*

REFERENCES

1. Kireev S. O., Kovalev V. N. Struktura, kinematika i geometriya planetarnyh peredach s vnecentroidnym cevochnym zacepleniem. Novochoerkassk: Novochoerkasskij gos. tekhn. un-t, 1995. 98 s.

2. Matchanov R. D. Hlopkouborochnye mashiny (1929–2010 gg.). Tashkent: Fan va tekhnologiya, 2011. 354 s.
3. Coulisse Mechanism with Rotating Link for Operating Part Drive Unit of the Mortar Pump / B. Alimov [et al.] // Construction Mechanics, Hydraulics and Water Resources Engineering (CONMECHYDRO-2021): Proceedings of the International Scientific Conference (Tashkent, Uzbekistan, 1–3 April, 2021). E3S Web of Conferences. 2021. Vol. 264. Art. No. 05050. 7 p. DOI: 10.1051/e3sconf/202126405050.
4. Kinematic Analysis of a Planetary Mixer Mechanism / V. S. Bogdanov [et al.] // High-Tech and Innovations in Research and Manufacturing (HIRM-2019): Proceedings of the International Conference (Krasnoyarsk, Russia, 06 May 2019). Journal of Physics: Conference Series. 2019. Vol. 1353. Art. No. 012030. 6 p. DOI: 10.1088/1742-6596/1353/1/012030.
5. Kammel D. W. Design, Selection and Use of TMR Mixers. URL: <http://www.researchgate.net/publication/268272514> (data obrashcheniya 27.05.2024).
6. Design and Analysis of Planetary Gearbox for Industrial Concrete Mixer / S. Dange [et al.] // International Journal of Latest Research in Engineering and Technology. 2016. Vol. 2, iss. 4–1. Pp. 41–45.
7. Valigi M. C., Gasperini I. A Preliminary Examination of Blades Wear in a Planetary Concrete Mixer. URL: <http://www.researchgate.net/publication/266458874> (data obrashcheniya 27.05.2024).
8. Pawar P. V., Kulkarni P. R. Critical Review of Design of Planetary Gears and Gear Box // International Journal of Mechanical Engineering and Information Technology. 2015. Vol. 3, iss. 3. Pp. 1067–1073.
9. Methodology for Calculating Power Consumption of Planetary Mixers / Antsiferov S. I. [et al.] // Proceedings of the International Conference on Mechanical Engineering, Automation and Control Systems (Tomsk, Russia, 4–6 December 2017). IOP Conference Series: Materials Science and Engineering. 2018. Vol. 327, iss. 4. Art. № 042007. 9 p. DOI: 10.1088/1757-899X/327/4/042007.
10. Sutar S. S., Sutar A. V., Rawal M. R. Torque Measurement in Epicyclic Gear Train // International Journal of Engineering and Advanced Technology. 2016. Vol. 5, iss. 6. Pp. 185–189.
11. Turanov Kh., Shaumarova M. Incorrect Application of the Epicycloid Equation to the Planetary Mechanism of the Cotton Harvester // Topical Problems of Green Architecture, Civil and Environmental Engineering 2019 (TPACEE 2019): Proceedings of the International Conference (Moscow, Russia, 19–22 November 2019). E3S Web of Conferences. 2020. Vol. 164. Art. № 06034. 10 p. DOI: 10.1051/e3sconf/202016406034.
12. Alimov B. M., Abdikayimova G. Dynamic Calculation of Gears of the Planetary Mixer Mechanism // Problems in the Textile and Light Industry in the Context of Integration of Science and Industry and Ways to Solve Them (PTLI-CISIWS-2): Proceedings of the International and Practical Conference (Namangan, Uzbekistan, 04–05 May 2023). AIP Conference Proceedings. 2024. Vol. 3045. Art. № 060004. DOI: 10.1063/5.0197543.
13. Epi and Hypocyclic Planetary Gear to Drive the Working Organ of the Mixer / B. M. Alimov [et al.] // APAM–2021: Proceedings of the International Conference on Actual Problems of Applied Mechanics (Samarkand, Uzbekistan, 27–29 October 2021). AIP Conference Proceedings. 2022. Vol. 2637. Art. № 060010. DOI: 10.1063/5.0120095.
14. Rizayev A., Alimova D., Alimov B. Modeling Spindle Drive of the Cotton Harvester // ETESD–2022: Proceedings of the International Conference on Environmental Technologies and Engineering for Sustainable Development (Tashkent, Uzbekistan, 12–15 October 2022). IOP Conference Series: Earth and Environmental Science. 2022. Vol. 1112. Art. № 012042. 8 p. DOI: 10.1088/1755-1315/1112/1/012042.

Received: 10.06.2024

Accepted: 17.07.2024

УДК 656.073

Математическая модель выкладки товаров на полочном пространстве складского стеллажа с одновременной горизонтальной и вертикальной категоризацией товаров

Черняховская Екатерина Сергеевна — PhD, канд. наук по менеджменту и качеству, преподаватель Вроцлавского университета экономики и бизнеса. Область научных интересов: оптимизация, программирование, распределение полочного пространства. E-mail: kateryna.czerniachowska@ue.wroc.pl

Вроцлавский университет экономики и бизнеса, Вроцлав, Польша

Для цитирования: Черняховская Е. С. Математическая модель выкладки товаров на полочном пространстве складского стеллажа с одновременной горизонтальной и вертикальной категоризацией товаров // Интеллектуальные технологии на транспорте. 2024. № 3 (39). С. 21–29. DOI: 10.20295/2413-2527-2024-339-21-29

Аннотация. Организация выкладки товаров на полочном пространстве складского стеллажа с выделением отдельных категорий позволяет более рационально использовать доступные ресурсы. Это способствует оптимизации пространства на складе, обеспечивая компактное и логичное размещение товаров, что облегчает их поиск и извлечение. Такой подход также уменьшает количество времени, затрачиваемого на обработку заказов, и снижает вероятность ошибок при комплектации. В результате повышается общая эффективность складских операций и улучшается управление запасами. Целью исследования является разработка математической модели выкладки товара на складском стеллаже с одновременной категоризацией товаров и полок по вертикали и по горизонтали. Исследование имеет важное значение для складской логистики.

Ключевые слова: математическое моделирование, оптимизация, распределение полочного пространства, категоризация товаров

Введение

Склады являются важными элементами современных цепочек поставок. На складах хранится продукция и собираются заказы, тем самым обеспечивается бесперебойная работа всей логистической системы. Правильная организация складского пространства и оптимизация процессов комплектации играют ключевую роль в эффективности цепочек поставок.

Оптимизация операций комплектации заказов включает выполнение различных задач, перечисленных ниже [1, 2]:

- Выбор метода комплектации: этот процесс включает в себя комплектацию единичного заказа, комплектацию партиями, зональную комплектацию и волновую комплектацию. Метод выбирается в соответствии с политикой склада при комплектации заказов.
- Пакетирование заказов: процедура, при которой заказы клиентов собираются и одновременно выдаются в четко отведенное время.
- Маршрутизация комплектовщика: определяется порядком, в котором товары включаются

в список для комплектования, чтобы обеспечить наиболее эффективную траекторию перемещения комплектовщика по складу.

- Сортировка/накопление заказов: операция, включающая объединение выбранных продуктов по заказу клиента или по месту доставки [2].

Процесс извлечения продуктов со склада в ответ на конкретный запрос клиента называется комплектованием заказов [3]. Известно, что комплектование заказов, являющееся наиболее трудоемким процессом на складах с ручными системами, может составлять до 55 % от общих операционных расходов склада [4]. В результате планирование процесса комплектования заказов следует считать одной из важнейших задач для повышения производительности склада в целом [5].

Комплектование заказов — это одна из ключевых функций склада. Оно представляет собой подпроцесс более общего складского процесса, который включает извлечение товаров из запасов, организованных в блоки или на стеллажах, для выполнения качественных и количественных запросов клиента [6, 7].

Заказ клиента может быть собран полностью одним человеком или группой сотрудников, работающих поочередно или одновременно. Выбор оптимальной стратегии зависит от множества факторов, среди которых одним из наиболее значимых является скорость обработки заказов [8].

Таким образом, склады играют ключевую роль в современных цепочках поставок, обеспечивая хранение продукции и сборку заказов. Эффективная организация складского пространства с использованием горизонтальной и вертикальной категоризации, а также с присвоением нескольких меток товарам значительно улучшает видимость и доступность продукции, ускоряет процессы комплектации и позволяет гибко реагировать на изменения спроса. В результате это способствует повышению общей эффективности и устойчивости логистической системы.

Постановка задачи

Описываемая задача ранее была опубликована в [9–12]. Параметры и индексы, используемые

в модели: S — количество полок; P — количество товаров; K — количество категорий; T — количество меток; i — индекс полки, $i = 1, \dots, S$; j — индекс товара, $j = 1, \dots, P$; k — индекс категории, $k = 1, \dots, K$; t — индекс метки, $t = 1, \dots, T$; r — индекс положения, $r \in \{0, 1\}$;

$$r = \begin{cases} 0, & \text{прямое положение} \\ 1, & \text{боковое положение} \end{cases}.$$

Параметры полки i : s_i^l — длина; s_i^d — глубина; s_{ii}^g — бинарная метка;

$$s_{ii}^g = \begin{cases} 1, & \text{если полка } i \text{ обозначена меткой } t \\ 0, & \text{otherwise} \end{cases}.$$

Параметры товара j : p_j^w — ширина; p_j^d — глубина; p_j^u — прибыль; p_j^l — кластер; p_{ij}^t — метка t ; p_j^k — категория; p_j^s — лимит поставки; p_{jr}^w — ширина или глубина в зависимости от положения;

$$p_{jr}^w = \begin{cases} p_{j0}^w, & \text{если } r = 0, \\ \text{ширина при прямом положении} \\ p_{j1}^w, & \text{если } r = 1, \\ \text{глубина при боковом положении} \end{cases}.$$

p_j^{o2} — бинарный параметр возможности бокового положения;

$$p_j^{o2} = \begin{cases} 1, & \text{если боковое положение допустимо} \\ 0, & \text{в противном случае} \end{cases}.$$

f_j^{\min} — минимальное количество единиц товара; f_j^{\max} — максимальное количество единиц товара; s_j^{\min} — минимальное количество полок, на которых можно разместить товар; s_j^{\max} — максимальное количество полок, на которых можно разместить товар.

Параметры категории: c_k^m — минимальная ширина категории в процентах от длины полки; c_k^t — допуск ширины категории на разных полках в процентах от длины полки.

Параметры метки: b_t^n — специфика метки t , $b_t^n = \{H, H^+, V^+\}$; b_{ij}^t — метка совместимости продукта с полкой;

$$b_{ij}^t = \begin{cases} 1, & \text{если } s_{ii}^t = p_{ij}^t \wedge b_t^n = \{H\} \\ 0, & \text{в противном случае} \end{cases},$$

$t = 1, \dots, T$ — горизонтальный уровень полки для крупногабаритных товаров;

$$b_{ij}^t = \left\{ \begin{array}{l} \min (p_{ij}^t; 1) \wedge b_i^n = \{V^+\} \\ 1, \text{ если } p_{ij}^t = 1 \wedge s_{ii}^t = p_{ij}^t \wedge b_i^n = \{H^+\} \\ 0, \text{ если } p_{ij}^t = 1 \wedge s_{ii}^t \neq p_{ij}^t \wedge b_i^n = \{H^+\} \\ 1, \text{ если } p_{ij}^t = 0 \wedge b_i^n = \{H^+\} \end{array} \right\},$$

$t = 1, \dots, T$ — горизонтальные и вертикальные уровни полок для мелкогабаритных товаров.

Переменные задачи:

$$x_{ijr} = \left\{ \begin{array}{l} 1, \text{ если товар } j \text{ положен на полку } i \\ \text{ в положении } r \\ 0, \text{ в противном случае} \end{array} \right\} —$$

бинарная переменная расположения товара на полке, $i = 1, \dots, S, j = 1, \dots, P: r \in \{0, 1\}, x_{ijr} \in \{0, 1\}$

f_{ijr} — количество единиц товара j на полке i в положении r ;

$$y_j = \left\{ \begin{array}{l} 0, \text{ если товар } j \text{ в прямом положении} \\ 1, \text{ если товар } j \text{ в боковом положении} \end{array} \right\} —$$

положение товара $j, j = 1, \dots, P: y_i \in \{0, 1\}$.

Логистическая цепь поставок на складе организуется по зонам для оптимизации процесса комплектации товаров. Каждая зона отвечает за определенную группу товаров, что ускоряет сборку заказов и снижает вероятность ошибок. Разделение склада на зоны позволяет более эффективно управлять запасами и оперативно реагировать на изменения спроса. Использование зональной системы на складах также облегчает работу сотрудников, делая процесс комплектации более структурированным и предсказуемым.

В данном исследовании задача определения размеров полочного пространства на стеллаже представлена таким образом, что стеллаж визуально разделен на вертикальные категории, чтобы комплектовщику заказов было проще находить товар. Дистрибьюторы по-разному расставляют свою продукцию на полках. Они могут расположить товар в заметных горизонтальных или вертикальных категориях в зависимости от типа товара, упаковки или массы. Каждая полка исследуемого стеллажа маркируется горизонтально, и каждой полке присваивается соответствующая метка. Полки стеллажа также маркируются вертикально, то есть метка присваивается каждой полке в категории. Когда

контейнер подъезжает к стеллажу, комплектовщик может быстро найти заказанный товар на стеллаже и положить его в контейнер.

Основную проблему можно сформулировать следующим образом. Определенное количество товаров должно быть разложено на полках стеллажа в складской зоне. Вертикальные и горизонтальные категории используются для категоризации товаров на стеллаже. Цель состоит в том, чтобы определить необходимое количество места на полке для каждой категории на стеллаже, а также количество единиц товара, максимизируя общую прибыль.

Обычно на складских стеллажах представлены различные виды продукции. Кластеризация или категоризация используется для их группировки. Каждая товарная категория выложена по вертикали. По горизонтали предметы и полки также маркируются. При этом у каждого товара может быть несколько меток. Каждой полке может быть присвоено несколько меток одновременно.

В качестве иллюстрации полочных меток рассмотрим следующие случаи:

- полка используется для крупногабаритных/мелкогабаритных товаров;
- на полку ставится товар в конкретной упаковке (коробка, полиэтиленовый пакет, банка, бутылка);
- полка находится на уровне руки/глаз/головы.

В данном исследовании предлагается три метки для обозначения товаров и полок $b_i^n = \{H; H^+; V^+\}$:

- H — горизонтальная метка для обозначения весовой категории (крупногабаритные, мелкогабаритные). Не допускается располагать вместе товары, слишком различающиеся по весу (большие коробки и мелкие винтики).
- H^+ — горизонтальный уровень для обозначения высоты (руки/глаз/головы). Определенные товары должны размещаться на своем уровне. При наличии места другие товары без указания уровня также могут быть поставлены на эту полку.
- V^+ — вертикальная метка для обозначения всех полок в данной категории на стеллаже (по типу товаров, цвету, размеру, упаковке).

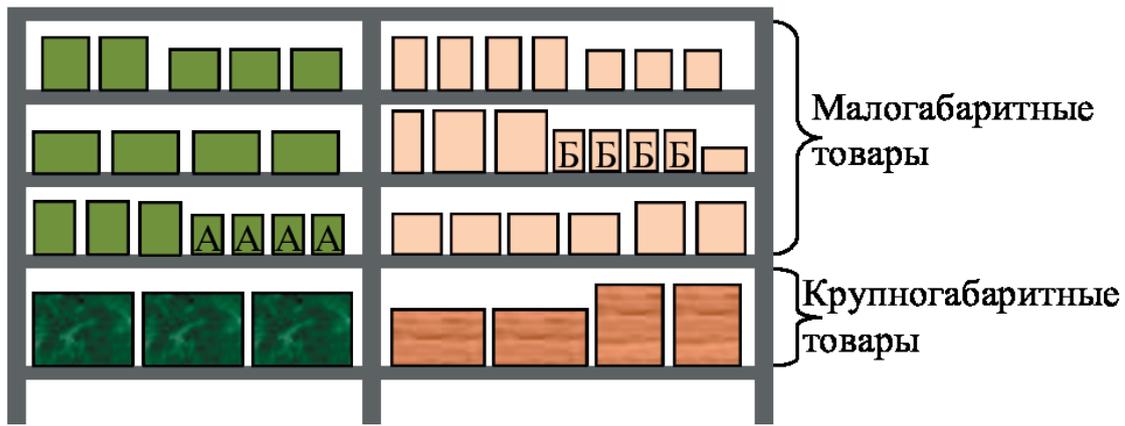


Рис. 1. Стеллаж с вертикальными (цвет) и горизонтальными (темные, светлые тона) метками: две вертикальные категории (зеленый, коричневый), две горизонтальные категории (малогоабаритные, крупногабаритные товары)

Допускается использование одновременно нескольких меток как для продукта, так и для полки.

На рис. 1 показан пример вертикальных и горизонтальных меток на складском стеллаже. В качестве примера можно рассмотреть вертикальную категоризацию товаров в две категории. Самая нижняя полка отведена под крупногабаритные и тяжелые товары. На других полках можно разместить малогабаритные товары.

На данном стеллаже указаны следующие уровни: нижний уровень для крупногабаритных товаров, уровень руки для малогабаритных товаров, уровень глаз для малогабаритных товаров и самая высокая полка без метки (там можно разместить разные малогабаритные товары). Некоторым товарам на стеллаже присвоены следующие метки:

- сформированы две вертикальные категории (V^+);
- нижний уровень — для крупногабаритных товаров обеих категорий (H);
- все полки, кроме нижней, предназначены для малогабаритных товаров обеих категорий (H);
- товар Б во второй категории (V^+) малогабаритный (H) на уровне глаз (H^+);
- товар А в первой категории (V^+) малогабаритный (H) на уровне руки (H^+).

Товары допускается маркировать одновременно несколькими метками. Метки $b_i^n = \{H; H^+; V^+\}$ характеризуют только тип группировки. Может

быть несколько значений (названий) определенного типа метки. Некоторые товары могут быть сгруппированы в кластер, поэтому их следует размещать на одной и той же полке (например, зарядное устройство и кабель). Таким образом, дистрибьюторы предотвращают ситуации отсутствия товара на складе, заполняя полки дополнительными, обычно покупаемыми вместе товарами. Некоторым товарам может быть не присвоена ни одна метка, должна быть лишь определена вертикальная категория.

Если продукт упакован в коробку, его можно расположить на полке двумя способами: лицевой стороной и боковой. Очевидно, что параметр положения товара не используется для бутылок, потому что ширина и глубина идентичны. По умолчанию все товары возможно поставить на полку лицевой стороной. В результате бинарный параметр положения определяет, можно ли разместить товар боком, то есть достаточно ли видны этикетка и название товара с его боковой стороны.

Чтобы решить задачу, дистрибьютор должен определить, на какой полке, на скольких полках следует разместить товар, затем определить количество единиц каждого товара, которое должно быть помещено на каждую полку, затем определить, какой стороной поставить товар на полку (лицевой или боковой), а также учесть набор ограничений, разделенных на четыре категории: ограничения на полку, ограничения на товар, ограничения

на положение и ограничения на метки. Целью размещения товаров на стеллаже является максимизация общей прибыли со стеллажа.

В данной задаче используются следующие переменные:

x_{ijr} — товар j размещен на полке i в положении r ;

f_{ijr} — количество единиц товара j на полке i в положении r ;

y_i — положение товара j .

Проблему можно сформулировать следующим образом:

$$\max \sum_{j=1}^P \sum_{i=1}^S \sum_{r=0}^1 p_j^u f_{ijr}.$$

1. Ограничения по полке:

- длина полки:

$$\forall(i) \left[\sum_{j=1}^P \sum_{r=0}^1 p_{jr}^w f_{ijr} \leq s_i^w \right];$$

- глубина полки, если товар поставлен в прямом положении:

$$\forall(i, j, p_{j1}^w > s_i^d) [f_{ij0} = 0];$$

- глубина полки, если товар поставлен в боковом положении:

$$\forall(i, j, p_{j0}^w > s_i^d) [f_{ij1} = 0].$$

2. Ограничения на товар:

- минимальное и максимальное количество единиц товара:

$$\forall(i, j) \left[f_j^{\min} x_{ijr} \leq \sum_{r=0}^1 f_{ijr} \leq f_j^{\max} x_{ijr} \right];$$

- лимит поставки, если один и тот же товар размещен на нескольких полках:

$$\forall(j) \left[\sum_{i=1}^S \sum_{r=0}^1 f_{ijr} \leq p_j^s \right];$$

- товар стоит на полке:

$$\forall(i, j, r) [f_{ijr} \leq x_{ijr} f_j^{\max}];$$

- товары, сгруппированные в кластеры, размещаются на одной полке:

$$\forall(i) \forall(a, b : p_a^l = p_b^l, a, b = 1, \dots, P) \left[\sum_{r=0}^1 x_{iar} = \sum_{r=0}^1 x_{ibr} \right].$$

3. Ограничения на расположение товара на нескольких полках:

- минимальное и максимальное количество полок:

$$\forall(j) \left[s_j^{\min} \leq \sum_{i=1}^S \sum_{r=0}^1 x_{ijr} \leq s_j^{\max} \right];$$

- при расположении товара на нескольких полках его можно поставить только на подряд идущей полке:

$$\forall(j) \forall(a, b : |a - b| \neq 1 \wedge a < b, a, b = 1, \dots, S) \left[\sum_{r=0}^1 x_{ajr} + \sum_{r=0}^1 x_{bjr} \leq 1 \right].$$

4. Ограничения на положение товара:

- боковое положение товара допустимо:

$$\forall(i, j) [y_j \leq p_j^{o2}];$$

- доступно только одно положение для одного товара (прямо или боком):

$$\forall(i, j) \left[\sum_{r=0}^1 x_{ijr} \leq 1 \right].$$

5. Ограничения на метки:

- метки совместимости полок и товаров:

$$\forall(i, j) \left[\prod_{t=1}^T b_{ij}^t \geq \sum_{r=0}^1 x_{ijr} \right];$$

- минимальная ширина категории, если категория имеется на полке:

$$\forall(i, k) \left[\left(\sum_{j=1, p_j^k=k}^P \sum_{r=0}^1 p_{jr}^w f_{ijr} \geq \left[s_i^l \cdot c_k^m \right] \right) \vee \left(\sum_{j=1, p_j^k=k}^P \sum_{r=0}^1 f_{ijr} = 0 \right) \right];$$

- допуск по ширине категории на разных полках:

$$\forall(k) \left[\max_{i=1, \dots, S} \left(\sum_{j=1, \dots, S}^P \sum_{r=0}^1 p_{jr}^w f_{ijr} \right) - \min_{i=1, \dots, S} \left(\sum_{j=1, \dots, S}^P \sum_{r=0}^1 p_{jr}^w f_{ijr} \right) \leq \left[\max_{i=1, \dots, S} (s_i^l) \cdot c_k^t \right] \right].$$

6. Ограничения отношений:

- соотношение единиц товара и нахождения товара на полке:

$$\forall(i, j, r) [f_{ijr} \geq x_{ijr}];$$

- соотношение единиц товара и прямого положения:

$$\forall(i, j) [f_{ij0} \leq (1 - y_j) f_j^{\max}];$$

- соотношение единиц товара и бокового положения:

$$\forall(i, j) [f_{ij1} \leq y_j f_j^{\max}].$$

7. Переменные решения:

- товар размещен на полке:

$$\forall(i, j, r) [x_{ijr} \in \{0, 1\}];$$

- количество единиц товара на полке:

$$\forall(i, j, r) [f_{ijr} = \{f_j^{\min} \dots f_j^{\max}\}];$$

- прямое или боковое положение единиц товара на полке:

$$\forall(j) [y_j \in \{0, 1\}].$$

Обсуждение

В данном исследовании приводится математическая модель распределения полочного пространства складского стеллажа. Организация выкладки товаров на полочном пространстве с выделением отдельных категорий позволяет более рационально использовать доступные ресурсы. Это достигается путем логического распределения товаров по категориям, что упрощает их нахождение и извлечение. Категоризация может основываться на различных факторах, таких как тип продукции, частота покупки или сезонность спроса.

Организация выкладки товаров на полочном пространстве складов с выделением отдельных категорий и определением конкретных по-

лок для каждого типа товаров обеспечивает ряд ключевых преимуществ для логистической цепи поставок. Во-первых, такая организация способствует лучшей видимости товаров для комплектовщика. Размещение товаров определенных категорий вместе с другими сопутствующими продуктами облегчает их поиск, что сокращает время, затрачиваемое на сборку заказа. Это особенно важно в условиях высокого объема заказов и ограниченного времени на их выполнение, что позволяет увеличить производительность и снизить затраты. Во-вторых, вертикальная и горизонтальная категоризация товаров дает возможность гибко адаптировать выкладку в зависимости от сезонных изменений спроса и маркетинговых акций. Например, в период праздничных распродаж или сезонных пиков спроса можно легко переместить наиболее востребованные товары на более доступные и видимые полки. Это позволяет быстро реагировать на изменения рынка и потребительских предпочтений, обеспечивая своевременную доступность нужных товаров.

Кроме того, такая система категоризации и размещения товаров улучшает управление запасами. Четкое разграничение полочного пространства по категориям и меткам упрощает мониторинг остатков и процессы пополнения полок. Это позволяет более точно планировать закупки и минимизировать риск дефицита или излишков товаров.

Однако необходимо учитывать и некоторые сложности. Внедрение и поддержание системы с одновременной горизонтальной и вертикальной категоризацией требует значительных организационных усилий. Для успешной реализации такой системы необходимо тщательно продумать планировку склада, обучить сотрудников и постоянно обновлять информацию о местоположении товаров. Это может потребовать дополнительных финансовых и временных затрат, особенно на этапе внедрения.

Таким образом, организация выкладки товаров с использованием горизонтальной и вертикальной категоризации, а также присвоением нескольких меток товарам и полкам предоставляет значительные

преимущества в управлении складом. Это улучшает видимость и доступность товаров, ускоряет сборку заказов и позволяет гибко реагировать на изменения спроса, однако требует тщательного планирования и постоянного контроля для эффективного функционирования.

Заключение

В данной работе дается математическая модель выкладки товаров на складском стеллаже с применением вертикальной и горизонтальной категоризации товаров и полок. Такой подход представляет существенный вклад в управление товарами на складах и организацией сборки заказов комплектовщиками.

Одновременная горизонтальная и вертикальная категоризация товаров улучшает видимость и доступность продукции для комплектовщиков. Размещение товаров на полках с учетом их размеров, веса и частоты заказов позволяет сократить время, необходимое для их извлечения. Например, более часто востребованные товары можно разместить на уровне глаз, что облегчает их нахождение и уменьшает физическую нагрузку на работников.

Кроме того, присвоение товарам нескольких меток при категоризации дает возможность гибко адаптировать выкладку в зависимости от текущих потребностей. Сезонные изменения спроса или маркетинговые акции могут требовать быстрой перестановки товаров, и наличие системы меток позволяет сделать это без значительных усилий. Например, в период новогодних праздников популярные подарочные наборы могут быть размещены на наиболее доступных полках, чтобы ускорить процесс их комплектации.

Дальнейшие исследования будут направлены на разработку способов многоуровневой категоризации различных товаров, что необходимо для обеспечения работоспособности приведенной модели. Важность правильной категоризации заключается в ее способности существенно повысить эффективность и точность операций на складе.

Исследования также затронут аспекты эргономики и удобства для работников склада, чтобы новая система не только повышала производительность, но и улучшала условия труда. Это включает в себя оптимизацию маршрутов комплектовщиков и обеспечение легкого доступа к наиболее востребованным товарам.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Davarzani H., Norrman A. Toward a Relevant Agenda for Warehousing Research: Literature Review and Practitioners' Input // *Logistics Research*. 2015. Vol. 8, iss. 1. Art. 1. 18 p. DOI: 10.1007/s12159-014-0120-1
2. Bahrami B., Aghezzaf E.-H., Limère V. Enhancing the Order Picking Process Through a New Storage Assignment Strategy in Forward-Reserve Area // *International Journal of Production Research*. 2019. Vol. 57, iss. 21. Pp. 6593–6614. DOI: 10.1080/00207543.2019.1567953
3. De Koster R., Le-Duc T., Roodbergen K. J. Design and Control of Warehouse Order Picking: A Literature Review // *European Journal of Operational Research*. 2007. Vol. 182, iss. 2. Pp. 481–501. DOI: 10.1016/j.ejor.2006.07.009
4. Warehousing in the Global Supply Chain: Advanced Models, Tools and Applications for Storage Systems / R. Manzini (ed.) // London: Springer-Verlag, 2012. 508 p. DOI: 10.1007/978-1-4471-2274-6
5. Chen W., Zhang Y., Zhou Y. Integrated Scheduling of Zone Picking and Vehicle Routing Problem with Time Windows in the Front Warehouse Mode // *Computers and Industrial Engineering*. 2022. Vol. 163. Art. 107823. 14 p. DOI: 10.1016/j.cie.2021.107823
6. Emmett S. Excellence in warehouse management: How to minimise costs and maximise value. Chichester: John Wiley & Sons, 2005. 320 p.
7. Parikh P. J., Meller R. D. Selecting Between Batch and Zone Order Picking Strategies in a Distribution Center // *Transportation Research Part E: Logistics and Transportation Review*. 2008. Vol. 44, iss. 5. Pp. 696–719. DOI: 10.1016/j.tre.2007.03.002

8. Bartholdi III J. J., Hackman S. T. Warehouse & Distribution Science: Release 0.94. Atlanta (GA): Supply Chain and Logistics Institute, Georgia Institute of Technology, 2011. 321 p.
9. Czerniachowska K., Lutosławski K., Fojcik M. Heuristics for Shelf Space Allocation Problem with Vertical and Horizontal Product Categorization // Procedia Computer Science. 2022. Vol. 207. Pp. 195–204. DOI: 10.1016/j.procs.2022.09.052
10. Czerniachowska K., Wichniarek R., Żywicki K. Heuristics for Dimensioning the Shelf Space on the Rack with Vertical and Horizontal Product Categorisation in the Distribution Centre with Zone Picking // Management and Production Engineering Review. 2022. Vol. 13. No. 2. Pp. 31–47. DOI: 10.24425/mper.2022.142053
11. Czerniachowska K., Lutosławski K., Hernes M. Linear and Non-Linear Shelf Space Allocation Problems with Vertical and Horizontal Bands // Journal of Economics and Management. 2022. Vol. 44. Pp. 119–141. DOI: 10.22367/jem.2022.44.06
12. Czerniachowska K., Wichniarek R., Żywicki K. Industry Expertise Heuristics for Dimensioning Shelf Space of Rack Storage Location in a Distribution Centre with Zone Picking // Management and Production Engineering Review. 2023. Vol. 14. No. 1. Pp. 43–60. DOI: 10.24425/mper.2023.145365

Дата поступления: 06.06.2024

Решение о публикации: 17.07.2024

Mathematical Model of Product Allocation on Warehouse Rack with Simultaneous Horizontal and Vertical Product Categorization

Ekaterina S. Cherniakhovskaia

— PhD in Management and Quality, Lecturer at the Wrocław University of Economics and Business. Research interests: optimization, programming, shelf space allocation. E-mail: kateryna.czerniachowska@ue.wroc.pl

Wrocław University of Economics and Business, Wrocław, Poland

For citation: Cherniakhovskaia E. S. Mathematical Model of Product Display on the Shelf Space of a Retail Chain with the Presence of Expandable and Contractible Segments for Definite Types of Products // Intellectual technologies on transport. 2024. No. 3 (39). Pp. 21–29. DOI: 10.20295/2413-2527-2024-339-21-29 (In Russian)

Abstract. *The organization of the product allocation on the shelf space of the warehouse rack with the creation of separate categories allows for a more rational use of available resources. This allows for optimization of warehouse space by ensuring a compact and logical product arrangement, making them easy to find and retrieve. This approach also reduces the amount of time spent on processing orders and reduces the likelihood of picking errors. As a result, the overall efficiency of warehouse operations and inventory management is improved. The purpose of the study is to develop a mathematical model of product allocation on a warehouse rack with simultaneous categorization of goods and shelves vertically and horizontally. The study is important for warehouse logistics.*

Keywords: *mathematical modelling, optimization, shelf space allocation, product categorization*

REFERENCES

1. Davarzani H., Norrman A. Toward a Relevant Agenda for Warehousing Research: Literature Review and Practitioners' Input // *Logistics Research*. 2015. Vol. 8, iss. 1. Art. 1. 18 p. DOI: 10.1007/s12159-014-0120-1
2. Bahrami B., Aghezzaf E.-H., Limère V. Enhancing the Order Picking Process Through a New Storage Assignment Strategy in Forward-Reserve Area // *International Journal of Production Research*. 2019. Vol. 57, iss. 21. Pp. 6593–6614. DOI: 10.1080/00207543.2019.1567953
3. De Koster R., Le-Duc T., Roodbergen K. J. Design and Control of Warehouse Order Picking: A Literature Review // *European Journal of Operational Research*. 2007. Vol. 182, iss. 2. Pp. 481–501. DOI: 10.1016/j.ejor.2006.07.009
4. Warehousing in the Global Supply Chain: Advanced Models, Tools and Applications for Storage Systems / R. Manzini (ed.) // London: Spinger-Verlag, 2012. 508 p. DOI: 10.1007/978-1-4471-2274-6
5. Chen W., Zhang Y., Zhou Y. Integrated Scheduling of Zone Picking and Vehicle Routing Problem with Time Windows in the Front Warehouse Mode // *Computers and Industrial Engineering*. 2022. Vol. 163. Art. 107823. 14 p. DOI: 10.1016/j.cie.2021.107823
6. Emmett S. Excellence in warehouse management: How to minimise costs and maximise value. Chichester: John Wiley & Sons, 2005. 320 p.
7. Parikh P. J., Meller R. D. Selecting Between Batch and Zone Order Picking Strategies in a Distribution Center // *Transportation Research Part E: Logistics and Transportation Review*. 2008. Vol. 44, iss. 5. Pp. 696–719. DOI: 10.1016/j.tre.2007.03.002
8. Bartholdi III J. J., Hackman S. T. Warehouse & Distribution Science: Release 0.94. Atlanta (GA): Supply Chain and Logistics Institute, Georgia Institute of Technology, 2011. 321 p.
9. Czerniachowska K., Lutosławski K., Fojcik M. Heuristics for Shelf Space Allocation Problem with Vertical and Horizontal Product Categorization // *Procedia Computer Science*. 2022. Vol. 207. Pp. 195–204. DOI: 10.1016/j.procs.2022.09.052
10. Czerniachowska K., Wichniarek R., Żywicki K. Heuristics for Dimensioning the Shelf Space on the Rack with Vertical and Horizontal Product Categorisation in the Distribution Centre with Zone Picking // *Management and Production Engineering Review*. 2022. Vol. 13. No. 2. Pp. 31–47. DOI: 10.24425/mper.2022.142053
11. Czerniachowska K., Lutosławski K., Hernes M. Linear and Non-Linear Shelf Space Allocation Problems with Vertical and Horizontal Bands // *Journal of Economics and Management*. 2022. Vol. 44. Pp. 119–141. DOI: 10.22367/jem.2022.44.06
12. Czerniachowska K., Wichniarek R., Żywicki K. Industry Expertise Heuristics for Dimensioning Shelf Space of Rack Storage Location in a Distribution Centre with Zone Picking // *Management and Production Engineering Review*. 2023. Vol. 14. No. 1. Pp. 43–60. DOI: 10.24425/mper.2023.145365

Received: 06.06.2024

Accepted: 17.07.2024

УДК 629

Технологии и методы планирования перемещения БПЛА по маршрутным точкам

Липанов Илья Дмитриевич — аспирант кафедры «Информационные и вычислительные системы». E-mail: illipanov@mail.ru

Хомоненко Анатолий Дмитриевич^{1,2} — докт. техн. наук, профессор, профессор кафедры «Информационные и вычислительные системы», профессор кафедры «Математическое и программное обеспечение». E-mail: khomonenko@pgups.ru

¹ Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

² Военно-космическая академия имени А. Ф. Можайского, Россия, 197198, Санкт-Петербург, ул. Ждановская, 13

Для цитирования: Липанов И. Д., Хомоненко А. Д. Технологии и методы планирования перемещения БПЛА по маршрутным точкам // Интеллектуальные технологии на транспорте. 2024. № 3 (39). С. 30–43. DOI: 10.20295/2413-2527-2024-339-30-43

Аннотация. В нашем мире особое место занимают беспилотные летательные аппараты (БПЛА). Их способность перемещаться по заданным маршрутам открывает перспективы в разных сферах. **Цель исследования:** обзор и анализ навигационных систем и алгоритмов маршрутизации БПЛА, методов, позволяющих БПЛА с высокой точностью следовать по маршруту. Исследуются системы GPS и инерциальной навигации (INS), обеспечивающие точное определение местоположения. Анализируются возможности сенсорных систем — камер, лидаров и ультразвуковых датчиков — для обнаружения препятствий и корректировки траектории; воксельные карты для трехмерного моделирования окружающей среды и методы одновременной локализации и картографирования (SLAM); алгоритм A^* (A -star); генетический алгоритм маршрутизации, алгоритмы избегания препятствий на основе потенциалов и RRT. **Практическая значимость:** применение указанных алгоритмов и технологий может существенно повысить безопасность и точность маршрутизации БПЛА, возможность автономно перемещаться в сложных и динамически изменяющихся ландшафтах. В заключение **обсуждаются** преимущества и ограничения навигационных подходов и технологий, значимость интеграции сенсорных систем и методов SLAM для повышения автономности и эффективности БПЛА, направления дальнейших исследований.

Ключевые слова: маршрутизация БПЛА, GPS, INS, SLAM, ADS-B, воксельные карты, сенсорные системы, генетический алгоритм, алгоритм RRT

Введение

Современные беспилотные летательные аппараты (БПЛА, или дроны) становятся все более популярными в самых разных сферах, включая логистику, экологический мониторинг, сельское

хозяйство и спасательные операции. Одной из основных задач, обеспечивающих успешное выполнение миссий беспилотных летательных аппаратов, является их способность точно

и надежно перемещаться по заранее заданным точкам маршрута. Для этого требуется применять современные методы и технологии, которые позволяют планировать и контролировать движение аппарата в реальном времени, принимая во внимание различные факторы, такие как погодные условия, наличие препятствий и ограниченные ресурсы на борту.

Для достижения высокой точности навигации беспилотных летательных аппаратов активно используются глобальные системы позиционирования (Global Positioning System, GPS), инерциальные навигационные системы (Inertial Navigation System, INS) и разнообразные сенсорные технологии, которые помогают определить местоположение устройства в пространстве и корректировать его траекторию. Кроме того, для успешного передвижения по сложным и динамичным маршрутам необходимо применять интеллектуальные алгоритмы, способные адаптироваться к меняющимся условиям и обеспечивать автономное принятие решений на борту БПЛА.

Кроме того, остаются актуальными задачи по интеграции различных навигационных систем и разработке методов для обхода препятствий так, чтобы БПЛА могли безопасно и эффективно выполнять свои функции даже в условиях ограниченной видимости и плотной городской застройки.

Таким образом, исследование алгоритмов и технологий планирования перемещения БПЛА по заданным маршрутным точкам является важным направлением, имеющим большое значение для развития беспилотных технологий и расширения их применения в различных областях. В настоящей статье проведен обзор и рассмотрены ключевые подходы к планированию маршрутов, интеграции навигационных систем, а также перспективы внедрения новых технологий для повышения эффективности и безопасности полетов БПЛА.

Системы позиционирования БПЛА

Принципы навигации и маршрутизации БПЛА имеют важное значение для их эффективного и безопасного перемещения по установленным маршрутам.

Навигация БПЛА основывается на применении различных технологий, которые позволяют точно определять местоположение и путь движения устройства. Одной из основных технологий является система глобального позиционирования GPS. Принцип работы GPS заключается в измерении времени, необходимого для того, чтобы сигнал прошел от спутника до приемника. Зная это время и скорость света, можно рассчитать расстояние до спутника. При получении сигналов от нескольких спутников (обычно не менее четырех) приемник способен определить свои координаты с помощью метода триангуляции.

Формула вычисления расстояния до спутника [1]:

$$d = c \cdot t,$$

где d — расстояние до спутника, c — скорость света, t — время прохождения сигнала.

GPS обеспечивает точное определение местоположения БПЛА в режиме реального времени, что дает возможность отслеживать его перемещение и при необходимости вносить изменения в маршрут.

Обозначим вектор координат в GPS:

$$P = (x, y, z),$$

где x, y, z — координаты БПЛА в глобальной системе координат.

Для повышения точности позиционирования применяется дифференциальный GPS (Differential GPS, DGPS), который учитывает погрешности, возникающие из-за атмосферных явлений и других факторов. В системе DGPS используются наземные станции, обладающие точными координатами, которые способны вычислять ошибки в спутниковых сигналах. Эти станции передают корректирующую информацию, что позволяет GPS-приемникам снижать погрешности в определении местоположения.

Инерциальные навигационные системы INS также имеют важное значение в навигации беспилотных летательных аппаратов. INS фиксируют ускорения и угловые скорости устройства, что дает возможность определить его текущее местоположение и ориентацию. Схема интеграции GPS и INS приведена на рис. 1. Сочетание данных GPS

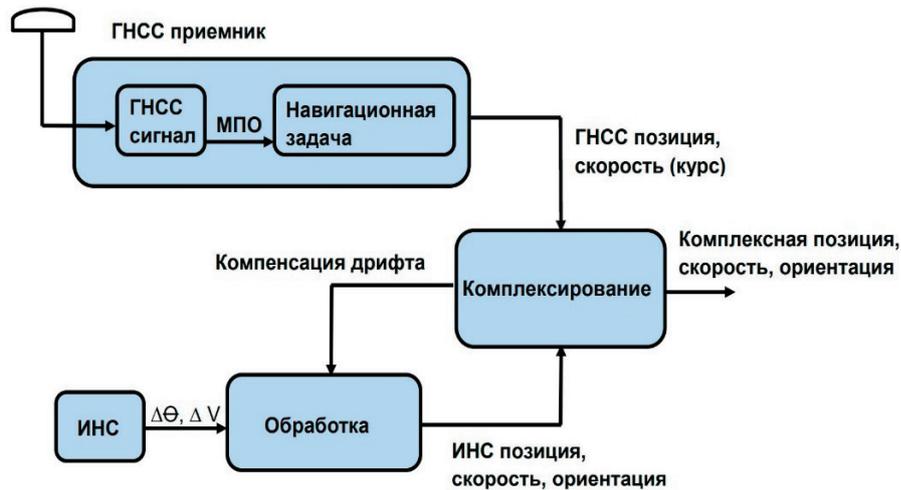


Рис. 1. Схема интеграции GPS и INS

и INS обеспечивает высокую точность и надежность навигации, что особенно критично в ситуациях, когда сигнал GPS заметно ослаблен [2].

Основные шаги включают интеграцию ускорений для получения скорости, а затем интеграцию скорости для обновления координат. Простейшая форма расчета координат может быть представлена следующими формулами:

1. Обновление скорости:

$$v(t) = v(t-1) + a(t) \cdot \Delta t,$$

где $v(t)$ — вектор скорости на текущий момент времени t , $v(t-1)$ — вектор скорости на предыдущий момент времени $t-1$, $a(t)$ — вектор ускорения, измеренный акселерометром в момент времени t , а Δt — временной интервал между измерениями.

2. Обновление координат:

$$r(t) = r(t-1) + v(t) \cdot \Delta t,$$

где $r(t)$ — вектор положения (координаты) на текущий момент времени t , $r(t-1)$ — вектор положения на предыдущий момент времени $t-1$, $v(t)$ — вектор скорости на текущий момент времени t .

Эти уравнения описывают процесс двойной интеграции ускорений, измеряемых INS, для определения текущего положения БПЛА.

Радионавигационные системы дополняют функционал GPS и инерциальных навигационных систем INS, предоставляя дополнительные сведения о местоположении и корректировке маршрутов. Для определения координат беспилотных летательных аппаратов применяются радиомаяки

и наземные станции, а системы автоматического зависимого наблюдения (Automatic Dependent Surveillance-Broadcast, ADS-B) позволяют получать информацию о местоположении и движении других воздушных судов, что помогает избежать столкновений.

Визуальная навигация и лидарные технологии становятся все более востребованными благодаря прогрессу в области обработки изображений и лазерного сканирования. Камеры и лидары позволяют собирать данные об окружающей среде и создавать карты местности. Обработка изображений способствует распознаванию маркеров и объектов на пути БПЛА, что повышает точность и безопасность навигационных процессов.

Взаимосвязь алгоритмов и методов навигации БПЛА

Алгоритмы и методы навигации БПЛА тесно связаны и служат одной общей цели — обеспечению точного и безопасного перемещения беспилотных летательных аппаратов в различных средах.

Основные аспекты взаимосвязи алгоритмов и методов навигации:

Основа для принятия решений. Алгоритмы являются основой для всех методов навигации БПЛА, определяя логику и шаги для выполнения задач навигации, таких как взлет, полет по заданному маршруту, обход препятствий, посадка и т.д. Они обрабатывают данные с датчиков и внешних

источников для принятия решений в реальном времени.

Встроенные системы. Методы навигации БПЛА часто используют встроенные системы, такие как GPS, инерциальные навигационные системы (ИНС), системы видения и лидары для определения положения и ориентации аппарата. Эти системы нуждаются в алгоритмах для интерпретации получаемых данных и принятия соответствующих навигационных решений.

Адаптация к среде. Алгоритмы позволяют БПЛА адаптироваться к изменениям в среде и условиям полета, используя методы, такие как SLAM для картографирования и локализации в неизвестной среде, или алгоритмы потенциалов и RRT для обхода препятствий.

Эффективность и безопасность. Современные алгоритмы стремятся повысить не только эффективность полетов БПЛА, минимизируя время и расход энергии, но и их безопасность, обеспечивая избежание столкновений и остальных опасностей.

Разработка и оптимизация. Алгоритмы и методы навигации разрабатываются и постоянно оптимизируются с учетом новых технологических достижений, таких как искусственный интеллект и машинное обучение, что позволяет создавать более интеллектуальные и автономные системы навигации БПЛА.

В итоге алгоритмы обеспечивают «мозг» навигационным системам БПЛА, позволяя эффективно использовать сенсорные данные и методы навигации для выполнения задачи.

Маршрутизация и алгоритмы планирования маршрута

Маршрутизация беспилотных летательных аппаратов включает несколько этапов, начиная с разработки маршрута и заканчивая его оптимизацией и корректировкой в реальном времени. На этапе планирования маршрута определяются точки пути (waypoints), через которые должен пройти БПЛА, при этом учитываются ограничения по высоте, зоны полетов и наличие опасных объектов.

Алгоритм A* (A-star) реализует один из самых распространенных одноименный метод плани-

рования маршрутов. Он применяется для поиска кратчайшего пути в графах и представляет собой усовершенствованную версию алгоритма Дейкстры, так как использует эвристическую функцию для оценки расстояния до конечной цели.

Формула для вычисления маршрута с применением модифицированного алгоритма Дейкстры (A*) [3]:

$$f(n) = g(n) + h(n),$$

где $f(n)$ — полная оценка стоимости пути через узел n , $g(n)$ — стоимость пути от начальной точки до узла n , $h(n)$ — эвристическая оценка стоимости пути от узла n до конечной точки. Пример работы алгоритма приведен на рис. 2.

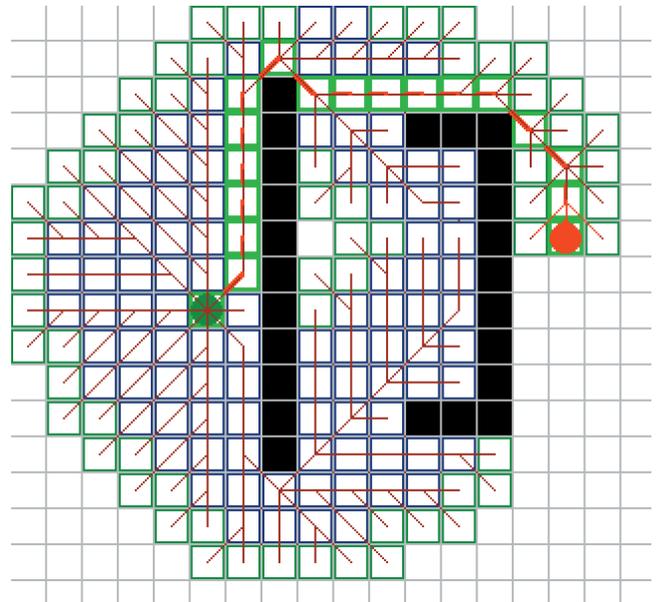


Рис. 2. Пример работы алгоритма A*

Алгоритм Дейкстры предназначен для определения кратчайшего пути от одной вершины графа ко всем остальным вершинам. Он функционирует за счет итеративного пересмотра оценок расстояний и служит основой для множества других алгоритмов [3]:

$$d(v) = \min(d(v), d(u) + w(u, v)),$$

где $d(v)$ — стоимость пути до вершины v , $d(u)$ — стоимость пути до вершины u , $w(u, v)$ — вес ребра между вершинами u и v . Блок-схема и пример работы алгоритма Дейкстры приведены на рис. 3.

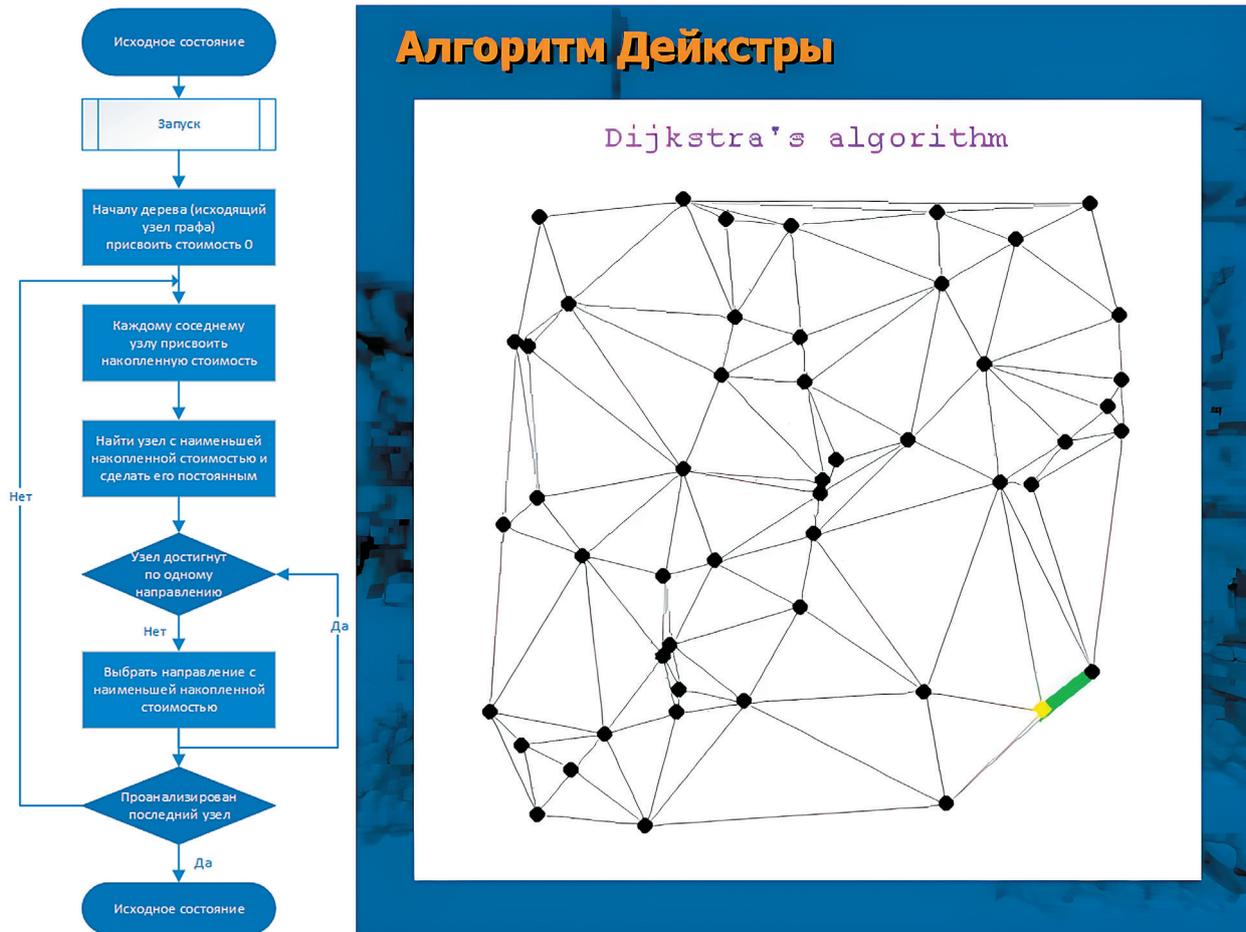


Рис. 3. Блок-схема и пример работы алгоритма Дейкстры

Разработка новых модификаций существующих алгоритмов маршрутизации предлагает оригинальные улучшения алгоритмов, таких как RRT, A*, или алгоритмов на основе потенциалов, адаптировав их под условия динамической среды с реальными временными ограничениями. Например, связь алгоритмов A* и RRT лучше оптимизирует путь с учетом изменения окружающей среды, а также делает это с меньшими вычислительными затратами в сравнении с алгоритмом Дейкстры.

Генетический алгоритм маршрутизации

Генетические алгоритмы (ГА) представляют собой эффективные инструменты для решения сложных задач оптимизации в таких областях, как наука о данных, инженерия, экономика и логистика, включая задачи маршрутизации.

Эти алгоритмы имитируют процесс естественного отбора, применяя принципы наследования,

мутации, селекции и кроссинговера для поиска наилучших решений [4].

Генетический алгоритм включает в себя несколько ключевых этапов:

1. *Создание начальной популяции маршрутов:* формирование исходного набора возможных маршрутов полета БПЛА (вариантов перемещения).
2. *Оценка эффективности маршрута:* анализ каждого маршрута полета БПЛА с точки зрения определенной целевой функции для планирования маршрута.
3. *Отбор маршрутов:* выбор наиболее оптимальных маршрутов для последующего скрещивания, основываясь на их эффективности.
4. *Скрещивание маршрутов:* объединение элементов отобранных маршрутов для создания новых вариантов маршрутов перемещения.
5. *Мутация маршрутов:* внесение случайных изменений в некоторые маршруты для обеспечения

разнообразия и избежания застревания в локальных оптимумах.

6. *Обновление популяции маршрутов*: замена предыдущего поколения маршрутов БПЛА заново сгенерированными вариантами.

7. *Итерация*: повторение процесса оценки эффективности, отбора, скрещивания, мутации и обновления до достижения заданного критерия завершения или нахождения оптимального маршрута перемещения.

В задачах маршрутизации решения обычно представляют в виде хромосом, каждая из которых соответствует определенному маршруту. Например, в задаче коммивояжера хромосома может состоять из последовательности точек маршрута (waypoints), которые необходимо посетить коммивояжеру. Задача коммивояжера и задача планирования перемещения БПЛА тесно связаны друг с другом, поскольку обе являются вариацией классической проблемы маршрутизации. В обоих случаях целью является оптимизация маршрута таким образом, чтобы достигнуть максимальной эффективности посещения ряда точек на карте или узлов в графе.

Функция приспособленности оценивает качество каждого решения по установленным критериям, таким как длина маршрута, время в пути или затраты энергии. В задачах маршрутизации часто в качестве показателя приспособленности используется общее расстояние маршрута.

Формула для оценки приспособленности [4]:

$$f(X) = \sum_{i=1}^{n-1} d(x_i, x_{i+1}),$$

где $X = [x_1, x_2, \dots, x_n]$ — маршрут, $d(x_i, x_{i+1})$ — расстояние между точками x_i и x_{i+1}

Селекция отбирает особей для размножения, основываясь на их адаптивности. Один из популярных подходов — это турнирная селекция, при которой случайным образом выбирается несколько особей и наиболее подходящая из них включается в группу родителей.

Пример турнирной селекции:

1. Выбираются случайные индивиды: ind_1, ind_2, ind_3 .

2. Оценивается их приспособленность: $f(ind_1), f(ind_2), f(ind_3)$.

3. Лучший индивид добавляется в набор родителей.

Кроссинговер объединяет фрагменты двух родительских хромосом для формирования новых организмов. Один из способов этого процесса — одноточечный кроссинговер, при котором случайным образом выбирается одна точка разрыва и происходит обмен частями хромосом.

Пример одноточечного кроссинговера:

Родители: $[A, B, C|D, E]$ и $[V, W, X|Y, Z]$.

Дети: $[A, B, C|Y, Z]$ и $[V, W, X|D, E]$.

Мутация случайным образом изменяет некоторые части хромосом для поддержания генетического разнообразия. Один из методов мутации — инверсия, где случайно выбранный сегмент хромосомы переворачивается.

Пример мутации инверсии:

Исходная хромосома: $[A, B, C, D, E]$

Мутированная хромосома: $[A, D, C, B, E]$

Рассмотрим пример задачи коммивояжера, где необходимо найти кратчайший маршрут, проходящий через заданное количество маршрутных точек. Для этого применим генетический алгоритм, который состоит из следующих шагов:

1. *Инициализация*: создание начальной популяции из случайных маршрутов.

2. *Оценка приспособленности*: вычисление суммарного расстояния для каждого маршрута.

3. *Селекция*: выбор лучших маршрутов для воспроизводства.

4. *Кроссинговер*: комбинирование части выбранных маршрутов для создания новых.

5. *Мутация*: случайное изменение некоторых маршрутов.

6. *Замена*: обновление популяции новыми маршрутами.

7. *Итерация*: повторение процесса до достижения оптимального маршрута.

Генетические алгоритмы являются эффективным и адаптивным средством для решения задач маршрутизации. Они основываются на принципах естественного отбора с применением таких механизмов, как наследование, мутация, селекция

и кроссинговер для поиска наилучших решений. Хотя у них есть некоторые недостатки, такие как высокая вычислительная сложность и необходимость в настройке параметров, генетические алгоритмы находят успешное применение в различных сферах, включая логистику, транспорт и автономные системы.

Пример генетического алгоритма для маршрутизации БПЛА.

Дано:

1. Заранее определенные ключевые точки, которые БПЛА должен пройти, или точки, которые нужно облететь.

2. Окружающая среда, включающая препятствия, которые необходимо избегать (здания, деревья и т.д.).

3. Набор возможных маршрутов, каждый из которых представлен как хромосома (список точек).

4. Фитнес-функция, определяющая качество каждого маршрута, основываясь на его длине и штрафах за приближение к препятствиям.

5. Операторы генетического алгоритма (селекция, кроссовер, мутация).

Найти:

1. Оптимальный маршрут для БПЛА.

2. Лучший маршрут среди популяции, который будет отвечать условиям фитнес-функции.

```
def genetic_algorithm ():
```

```
# 1. Инициализация
```

```
population = initialize_population()
```

```
for generation in range (MAX_GENERATIONS):
```

```
# 2. Оценка фитнеса
```

```
fitness_scores = evaluate_fitness (population)
```

```
# 3. Селекция
```

```
selected_routes = selection (population, fitness_scores)
```

```
# 4. Кроссовер
```

```
new_population = crossover (selected_routes)
```

```
# 5. Мутация
```

```
new_population = mutate(new_population)
```

```
# 6. Оценка новой популяции
```

```
population = new_population
```

```
# 7. Проверка условий завершения
```

```
if stop_condition_met(fitness_scores):
```

```
break
```

```
# Возвращаем лучший маршрут
```

```
return get_best_route(population, fitness_scores)
```

```
# Пример фитнес-функции
```

```
def evaluate_fitness(route):
```

```
return 1 / (route_length(route) + penalty_for_obstacles(route))
```

Для каждого маршрута в популяции вычисляется его фитнес-функция, которая определяет, насколько эффективен маршрут. Обычно в качестве фитнес-функции используется минимизация длины маршрута с учетом избегания препятствий и энергетических затрат на выполнение полета.

Фитнес-функция:

$$f(route) = \frac{1}{L(route) + P_{avoid}(route)},$$

где $L(route)$ — общая длина маршрута, $P_{avoid}(route)$ — штраф за приближение к препятствиям (например, если маршрут близок к препятствиям, фитнес уменьшается).

Системы и алгоритмы избегания препятствий

Реактивное управление и предотвращение столкновений с препятствиями играют ключевую роль в маршрутизации беспилотных летательных аппаратов. Алгоритмы локального планирования дают возможность адаптировать маршрут в зависимости от актуальных условий и избегать столкновений с различными преградами.

Сила избегания препятствия при планировании маршрута БПЛА обычно рассчитывается с использованием алгоритмов искусственных потенциальных полей. В этом подходе пространство вокруг БПЛА рассматривается как поле, в котором препятствия создают отталкивающие потенциальные поля, а цель — притягивающее потенциальное поле. Сила, действующая на БПЛА, является градиентом создаваемого потенциального поля.

Базовая формула для расчета силы избегания препятствий F_{avoid} , возникающей из потенциала U_{avoid} , создаваемого препятствием, выглядит следующим образом [5]:

$$F_{avoid} = -\nabla U_{avoid},$$

где ∇U_{avoid} — градиент потенциальной энергии, связанной с препятствием.

Градиент указывает направление наибольшего возрастания потенциала и определяется как:

$$U_{avoid} = \frac{1}{2} k_{avoid} \left(\frac{1}{d} - \frac{1}{D_0} \right)^2,$$

где k_{avoid} — коэффициент, определяющий величину отталкивания, d — текущее расстояние до препятствия, D_0 — пороговое расстояние, на котором начинает действовать сила.

Таким образом, сила избегания препятствия направлена в сторону, противоположную градиенту потенциала, и возрастает по мере приближения к препятствию.

Системы предотвращения столкновений для беспилотных летательных аппаратов должны гарантировать выявление препятствий и изменение маршрута полета в режиме реального времени. Для достижения этой цели применяются разнообразные датчики и сенсоры, включая камеры, лидары, ультразвуковые и инфракрасные устройства, а также системы GPS и инерциальной навигации (INS) [5].

Датчики и сенсоры, применяемые для оценки маршрутов:

1. *Камеры*: обеспечивают визуальную информацию об окружающей среде и используются для обнаружения и классификации препятствий с помощью алгоритмов компьютерного зрения.

2. *Лидары*: лазерные дальнометры, создающие трехмерную карту окружающей среды, что позволяет точно определять расстояние до объектов.

3. *Ультразвуковые датчики*: измеряют расстояние до ближайших объектов, используя звуковые волны. Они особенно эффективны на коротких дистанциях.

4. *Инфракрасные датчики*: используют инфракрасное излучение для обнаружения объектов, что особенно полезно в условиях низкой освещенности.

Для безопасного и эффективного полета БПЛА требуется наличие систем, способных избегать препятствия. Эти системы должны в реальном времени выявлять преграды и корректировать маршру-

т движения. Синергия этих систем обеспечивает надежное управление и предотвращение столкновений в различных условиях [6].

Алгоритм избегания препятствий на основе потенциалов

Алгоритм на основе потенциалов воспринимает пространство как поле потенциальной энергии, в котором цель притягивает беспилотный летательный аппарат, а преграды отталкивают его. Главная концепция заключается в формировании искусственного силового поля, которое направляет БПЛА к цели, одновременно предотвращая столкновения с препятствиями. Формула потенциального поля [7]:

$$F = F_{аттракция} + F_{репульсия},$$

где $F_{аттракция}$ — сила притяжения к цели, а $F_{репульсия}$ — сила отталкивания от препятствий.

Алгоритм на основе потенциалов представляет собой эффективное средство для организации маршрутов и навигации беспилотных летательных аппаратов, особенно когда имеются ограничения по вычислительным ресурсам и необходима высокая плавность перемещения [7].

Алгоритм обхода препятствий в реальном времени RRT

Алгоритм RRT (Rapidly-exploring Random Tree) создает дерево потенциальных маршрутов от стартовой позиции к цели, случайным образом создавая новые точки и проверяя их на наличие преград. Это дает возможность эффективно определять пути в сложных и изменяющихся условиях.

Основные шаги алгоритма RRT:

1. Генерация случайной точки в пространстве.
2. Нахождение ближайшей вершины дерева и создание нового ребра к случайной точке.
3. Проверка на столкновение с препятствиями и добавление новой точки в дерево, если путь безопасен.
4. Повторение процесса до достижения цели.

Алгоритм RRT является мощным инструментом для разработки маршрутов беспилотных летательных аппаратов в сложных и изменяющихся условиях. Его способность быстро находить пути делает его особенно ценным для задач, где необходима

быстрая реакция на изменения в окружающей среде. Тем не менее для эффективного применения RRT в реальных условиях часто требуются его адаптация и сочетание с другими методами планирования, чтобы преодолеть такие проблемы, как неоптимальные маршруты и значительная нагрузка на вычислительные мощности [8].

Воксельные карты

Воксельная карта — это трехмерная сетка, в которой каждая ячейка (воксель) может быть либо занята, либо свободна. Такой подход позволяет эффективно обрабатывать данные о препятствиях и разрабатывать безопасные маршруты [9].

Процесс построения воксельной карты включает несколько этапов:

1. *Сбор данных:* с помощью датчиков, таких как лидары, камеры, ультразвуковые и инфракрасные сенсоры, собираются данные об окружающей среде.

2. *Обработка данных:* полученные данные обрабатываются для выделения объектов и определения их положения в пространстве.

3. *Заполнение воксельной карты:* каждая ячейка воксельной карты заполняется информацией о состоянии (занята или свободна) на основе обработанных данных.

Использование трехмерного представления окружающей среды с помощью вокселей способствует эффективному моделированию сложных пространств, позволяя учитывать препятствия и де-

тализовать маршруты. Воксельные карты значительно повышают точность и надежность полетов беспилотников, особенно в условиях ограниченной видимости и сложного рельефа. В сочетании с современными алгоритмами планирования и датчиками воксельные карты обеспечивают высокий уровень автономности и безопасности для БПЛА.

Алгоритм локализации и картографирования SLAM

Одновременная локализация и картографирование (Simultaneous Localization and Mapping, SLAM) — это метод, который позволяет одновременно создавать карту окружающей среды и определять местоположение беспилотного летательного аппарата на этой карте. Воксельные карты являются важным элементом SLAM, так как они обеспечивают точные данные о пространстве и препятствиях [10].

Главная проблема SLAM заключается в том, что для создания карты необходимо точно знать местоположение БПЛА, а для определения этого местоположения требуется наличие точной карты. Этот процесс требует использования методов, которые могут исправлять ошибки, возникающие во время локализации и картографирования. Основные компоненты SLAM:

1. Для реализации SLAM на БПЛА используются различные типы сенсоров, такие как камеры,



Рис. 4. Пример использования воксельной карты

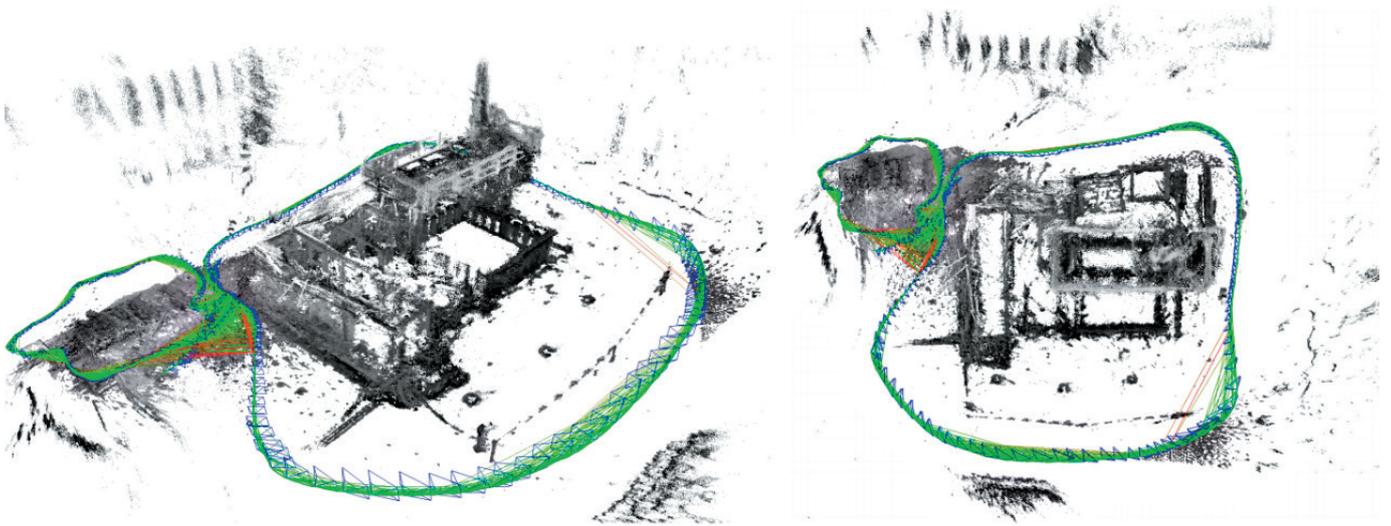


Рис. 5. Пример использования воксельной карты и метода SLAM

лидары, ультразвуковые и инфракрасные датчики. Эти сенсоры собирают данные об окружающей среде, которые затем обрабатываются для создания карты и определения местоположения [11].

2. Для работы с шумными и неполными данными, полученными от сенсоров, используются алгоритмы фильтрации, такие как Калмановский фильтр, расширенный Калмановский фильтр (Extended Kalman Filter — EKF) и фильтр частиц. Эти методы позволяют предсказывать движение БПЛА и корректировать его положение на основе новых данных.

3. Система SLAM должна уметь распознавать и сопоставлять особенности окружающей среды, такие как углы зданий, деревья или другие уникальные объекты, для корректной привязки карты к реальному миру.

4. Построение карты заключается в создании модели окружающей среды на основе данных, собранных с сенсоров. Это может быть двухмерная (2D) или трехмерная (3D) карта, в зависимости от сложности задачи и типа среды.

5. После получения данных и построения начальной карты проводится оптимизация, которая минимизирует ошибки и корректирует как карту, так и местоположение БПЛА.

SLAM представляет собой основополагающую технологию для автономной навигации беспилотных летательных аппаратов, позволяя одновре-

менно создавать карту окружающей среды и определять свое местоположение в режиме реального времени. Прогресс в разработке алгоритмов SLAM открывает новые горизонты для использования БПЛА в сложных и изменяющихся условиях, где классические методы навигации оказываются неэффективными [12].

Заключение

В статье рассмотрены ключевые алгоритмы и подходы, применяемые для перемещения БПЛА по заранее определенным маршрутным точкам. С учетом современных тенденций в развитии технологий БПЛА становятся все более важными в различных сферах, таких как логистика, экологический мониторинг и поисково-спасательные операции. Для эффективного выполнения этих задач требуется обеспечить надежное и точное перемещение аппаратов по заданному маршруту, используя современные алгоритмы планирования и навигации.

Одной из основных задач организации перемещения БПЛА является обеспечение точности и безопасности их полетов. Это достигается благодаря применению различных систем позиционирования, таких как GPS и INS, а также сенсорных технологий для обнаружения и обхода препятствий. Важно подчеркнуть необходимость разработки и внедрения интеллектуальных

алгоритмов маршрутизации, которые способны адаптироваться к изменяющимся условиям окружающей среды, учитывать как динамические, так и статические препятствия, а также эффективно использовать ограниченные вычислительные мощности БПЛА.

В частности, в статье кратко рассмотрены генетические алгоритмы, алгоритм RRT, алгоритм на основе потенциалов, алгоритм A*, воксельные карты и алгоритмы SLAM для моделирования окружающей среды и планирования маршрутов в трехмерном пространстве.

Таким образом, успешное перемещение БПЛА по заданным точкам требует комплексного подхода, который включает в себя точное позиционирование, надежные алгоритмы маршрутизации

и использование сенсорных данных для предотвращения столкновений.

Одним из направлений дальнейших исследований является разработка алгоритмов автоматической навигации и управления, которые позволят БПЛА безопасно и эффективно выполнять заданные маршруты [13, 14]. Еще одним важным аспектом исследований является разработка системы обнаружения и обхода препятствий на пути движения БПЛА, что потенциально позволит уменьшить риск столкновения и повысить безопасность полетов [15, 16]. Кроме того, исследования в области оптимизации маршрутов [17, 18] и управления энергопотреблением БПЛА [19–21] могут улучшить эффективность и экономичность использования этих технологий.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Elmeseiry N., Alshaer N., Ismail T. A. Detailed Survey and Future Directions of Unmanned Aerial Vehicles (UAVs) with Potential Applications // *Aerospace Engineering*. 2021. Vol. 8, iss. 12. No. 363. 29 p. DOI: 10.3390/aerospace8120363
2. Groves P. D. Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems. Second Edition. Norwood (MA): Artech House, 2013. 776 p.
3. Roberge V., Tarbouchi M., Labonte G. Comparison of Parallel Genetic Algorithm and Particle Swarm Optimization for Real-Time UAV Path Planning // *IEEE Transactions on Industrial Informatics*. 2013. Vol. 9, iss. 1. Pp. 132–141. DOI: 10.1109/TII.2012.2198665
4. Лего Т., Фомичев А. В., Лю Я. Решение задачи планирования полета малогабаритного беспилотного летательного аппарата в условиях городской среды // *Автоматизация. Современные технологии*. 2015. № 7. С. 19–24.
5. Kuwata Y. Real-Time Trajectory Design for Unmanned Aerial Vehicles Using Receding Horizon Control: A Thesis for the Degree of Master of Science in Aeronautics and Astronautics. Massachusetts Institute of Technology, 2003. 151 p. URL: <http://www.researchgate.net/publication/242403825> (assessed: 19.08.2024).
6. Ristić-Durrant D., Franke M., Michels K. A Review of Vision-Based On-Board Obstacle Detection and Distance Estimation in Railways // *Sensors*. 2021. Vol. 21, iss. 10. No. 3452. 30 p. DOI: 10.3390/s21103452
7. Recent Advances in Unmanned Aerial Vehicles: A Review / F. Ahmed [et al.] // *Arabian Journal for Science and Engineering*. 2022. Vol. 47, iss. 7. Pp. 7963–7984. DOI: 10.1007/s13369-022-06738-0
8. UAV Path Planning Based on Improved A* and DWA Algorithms / X. Bai [et al.] // *International Journal of Aerospace Engineering*. 2021. Vol. 2021. No. 4511252. 12 p. DOI: 10.1155/2021/4511252
9. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age / C. Cadena [et al.] // *IEEE Transactions on Robotics*. 2016. Vol. 32, iss. 6. Pp. 1309–1332. DOI: 10.1109/TRO.2016.2624754
10. Ubina N. A., Cheng S.-C. A Review of Unmanned System Technologies with Its Application to Aquaculture Farm Monitoring and Management // *Drones*. 2022, Vol. 6, iss. 1. Art. 12. 41 p. DOI: 10.3390/drones6010012
11. Visual SLAM for Unmanned Aerial Vehicles: Localization and Perception / L. Zhuang [et al.] // *Sensors*. 2024. Vol. 24, iss. 10. Art. 2980. 24 p. DOI: 10.3390/s24102980
12. Алагорцев Д. В., Хамухин А. В. Анализ эффективных методов оценки дальности и алгоритмов обработки видеoinформации на БПЛА // *Известия Тульского государственного университета. Технические науки*. 2020. Вып. 12. С. 255–261.

13. Али Б., Садеков Р. Н., Цодокова В. В. Алгоритмы навигации беспилотных летательных аппаратов с использованием систем технического зрения // *Гироскопия и навигация*. 2022. Т. 30, № 4 (119). С. 87–105. DOI: 10.17285/0869-7035.00105
14. Хомоненко А. Д., Яковлев Е. Л. Обоснование архитектуры сверточной нейронной сети для автономного распознавания объектов на изображениях бортовой вычислительной системой // *Наукоемкие технологии в космических исследованиях Земли*. 2018. Т. 10, № 6. С. 86–93. DOI: 10.24411/2409-5419-2018-10190
15. A Review on IoT Deep Learning UAV Systems for Autonomous Obstacle Detection and Collision Avoidance / P. Fraga-Lamas [et al.] // *Remote Sensing*. 2019. Vol. 11, iss. 18. Art. 2144. 29 p. DOI: 10.3390/rs11182144
16. Новиков П. А., Хомоненко А. Д., Яковлев Е. Л. Комплекс программ для навигации мобильных устройств внутри помещений с помощью нейронных сетей // *Информационно-управляющие системы*. 2016. № 1 (80). С. 32–39. DOI: 10.15217/issn1684-8853.2016.1.32
17. Способ построения «субоптимальных» маршрутов мониторинга разнотипных источников беспилотным летательным аппаратом / А. В. Тимошенко [и др.] // *Труды МАИ*. 2020. № 111. 18 с. DOI: 10.34759/trd-2020-111-10
18. Improved Artificial Bee Colony Algorithm-Based Path Planning of Unmanned Autonomous Helicopter Using Multi-Strategy Evolutionary Learning / Z. Han [et al.] // *Aerospace Science and Technology*. 2022. Vol. 122. Art. 7374. 17 p. DOI: 10.1016/j.ast.2022.107374
19. Energy-Efficient UAV-Assisted Mobile Edge Computing: Resource Allocation and Trajectory Optimization / M. Li [et al.] // *IEEE Transactions on Vehicular Technology*. 2020. Vol. 69, iss. 3. Pp. 3424–3438. DOI: 10.1109/TVT.2020.2968343
20. Energy-Efficient Trajectory Optimization for UAV-Assisted IoT Networks / L. Zhang [et al.] // *IEEE Transactions on Mobile Computing*. 2021. Vol. 21, iss. 12. Pp. 4323–4337. DOI: 10.1109/TMC.2021.3075083
21. Resource Allocation and Trajectory Optimization for QoE Provisioning in Energy-Efficient UAV-Enabled Wireless Networks / F. Zeng [et al.] // *IEEE Transactions on Vehicular Technology*. 2020. Vol. 69, iss. 7. Pp. 7634–7647. DOI: 10.1109/TVT.2020.2986776

Дата поступления: 25.08.2024

Решение о публикации: 16.09.2024

Technologies and Methods for Planning the Movement of UAVs Along Waypoints

Ilya D. Lipanov¹ — Graduate student at the Department of Information and Computing systems.
E-mail: illipanov@mail.ru

Anatoly D. Khomonenko^{1,2} — Doctor of Technical Sciences, Full Professor, Professor of the Department of Information and Computing systems; Professor of the Department of Mathematics and Software.
E-mail: khomonenko@pgups.ru

¹ Emperor Alexander I St. Petersburg State Transport University, 9, Moskovsky pr., Saint Petersburg, 190031, Russia

² Mozhaisky Military Aerospace Academy, 13, Zhdanovskaya str., St. Petersburg, 197198, Russia

For citation: Lipanov I. D., Khomonenko A. D. Technologies and Methods for Planning the Movement of UAVs Along Waypoints // *Intellectual Technologies on Transport*. 2024. № 3 (39). Pp. 30–43. DOI: 10.20295/2413-2527-2024-339-30-43. (In Russian)

Abstract. Unmanned aerial vehicles (UAVs) occupy a special place in our world. Their ability to navigate along set routes opens up prospects in various fields. **The purpose of the study** is to review and analyze navigation systems and UAV routing algorithms, methods that allow UAVs to follow the route with high accuracy. GPS and inertial navigation (INS) systems that provide accurate location determination are being investigated. The capabilities of sensor systems — cameras, lidars and ultrasonic sensors — are analyzed to detect obstacles and adjust the trajectory; voxel maps for three-dimensional modeling of the environment and methods of simultaneous localization and mapping (SLAM); algorithm A^* (A-star); genetic routing algorithm, potential-based obstacle avoidance algorithms and RRT. **Practical significance:** the use of these methods and technologies can significantly improve the safety and accuracy of UAV routing, the ability to navigate autonomously in complex and dynamically changing landscapes. In conclusion, the advantages and limitations of navigation approaches and technologies **are discussed**; the importance of integrating sensor systems and SLAM methods to increase the autonomy and efficiency of UAVs; directions for further research.

Keywords: UAV routing, GPS, INS, SLAM, ADS-B, voxel maps, sensor systems, genetic algorithm, RRT algorithm

REFERENCES

1. Elmeseiry N., Alshaer N., Ismail T. A. Detailed Survey and Future Directions of Unmanned Aerial Vehicles (UAVs) with Potential Applications // *Aerospace Engineering*. 2021. Vol. 8, iss. 12. No. 363. 29 p. DOI: 10.3390/aerospace8120363
2. Groves P. D. Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems. Second Edition. Norwood (MA): Artech House, 2013. 776 p.
3. Roberge V., Tarbouchi M., Labonte G. Comparison of Parallel Genetic Algorithm and Particle Swarm Optimization for Real-Time UAV Path Planning // *IEEE Transactions on Industrial Informatics*. 2013. Vol. 9, iss. 1. Pp. 132–141. DOI: 10.1109/TII.2012.2198665
4. Lego T., Fomichev A. V., Lyu YA. Reshenie zadachi planirovaniya poleta malogabaritnogo bespilotnogo letatel'nogo apparata v usloviyah gorodskoj srede // *Avtomatizaciya. Sovremennye tekhnologii*. 2015. № 7. S. 19–24.5.
5. Kuwata Y. Real-Time Trajectory Design for Unmanned Aerial Vehicles Using Receding Horizon Control: A Thesis for the Degree of Master of Science in Aeronautics and Astronautics. Massachusetts Institute of Technology, 2003. 151 p. URL: <http://www.researchgate.net/publication/242403825> (assessed: 19.08.2024).
6. Ristić-Durrant D., Franke M., Michels K. A Review of Vision-Based On-Board Obstacle Detection and Distance Estimation in Railways // *Sensors*. 2021. Vol. 21, iss. 10. No. 3452. 30 p. DOI: 10.3390/s21103452
7. Recent Advances in Unmanned Aerial Vehicles: A Review / F. Ahmed [et al.] // *Arabian Journal for Science and Engineering*. 2022. Vol. 47, iss. 7. Pp. 7963–7984. DOI: 10.1007/s13369-022-06738-0
8. UAV Path Planning Based on Improved A^* and DWA Algorithms / X. Bai [et al.] // *International Journal of Aerospace Engineering*. 2021. Vol. 2021. No. 4511252. 12 p. DOI: 10.1155/2021/4511252
9. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age / C. Cadena [et al.] // *IEEE Transactions on Robotics*. 2016. Vol. 32, iss. 6. Pp. 1309–1332. DOI: 10.1109/TRO.2016.2624754
10. Ubina N. A., Cheng S.-C. A Review of Unmanned System Technologies with Its Application to Aquaculture Farm Monitoring and Management // *Drones*. 2022, Vol. 6, iss. 1. Art. 12. 41 p. DOI: 10.3390/drones6010012
11. Visual SLAM for Unmanned Aerial Vehicles: Localization and Perception / L. Zhuang [et al.] // *Sensors*. 2024. Vol. 24, iss. 10. Art. 2980. 24 p. DOI: 10.3390/s24102980
12. Alatorcev D. V., Hamuhin A. V. Analiz effektivnyh metodov ocenki dal'nosti i algoritmov obrabotki videoinformacii na BPLA // *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki*. 2020. Vyp. 12. S. 255–261. (In Russian)
13. Ali B., Sadekov R. N., Codokova V. V. Algoritmy navigacii bespilotnyh letatel'nyh apparatov s ispol'zovaniem sistem tekhnicheskogo zreniya // *Giroskopiya i navigaciya*. 2022. T. 30, № 4 (119). С. 87–105. DOI: 10.17285/0869-7035.00105 (In Russian)

14. Khomonenko A. D., Yakovlev E. L. Obosnovanie arhitektury svertochnoj nejronnoj seti dlya avtonomnogo raspoznaniya ob"ektov na izobrazheniyah bortovoj vychislitel'noj sistemoj // Naukoemkie tekhnologii v kosmicheskikh issledovaniyah Zemli. 2018. T. 10, № 6. S. 86–93. DOI: 10.24411/2409-5419-2018-10190 (In Russian)
15. A Review on IoT Deep Learning UAV Systems for Autonomous Obstacle Detection and Collision Avoidance / P. Fraga-Lamas [et al.] // Remote Sensing. 2019. Vol. 11, iss. 18. 2144. 29 p. DOI: 10.3390/rs11182144
16. Novikov P. A., Khomonenko A. D., Yakovlev E. L. Kompleks programm dlya navigacii mobil'nyh ustrojstv vnutri pomeshchenij s pomoshch'yu nejronnyh setej // Informacionno-upravlyayushchie sistemy. 2016. № 1 (80). S. 32–39. DOI: 10.15217/issn1684-8853.2016.1.32 (In Russian)
17. Sposob postroeniya "suboptimal'nyh" marshrutov monitoringa raznotipnyh istochnikov bespilotnym letatel'nyim apparatom / A. V. Timoshenko [i dr.]. // Trudy MAI. 2020. № 111. 18 c. DOI: 10.34759/trd-2020-111-10 (In Russian)
18. Improved Artificial Bee Colony Algorithm-Based Path Planning of Unmanned Autonomous Helicopter Using Multi-Strategy Evolutionary Learning / Z. Han [et al.] // Aerospace Science and Technology. 2022. Vol. 122. Art. 7374. 17 p. DOI: 10.1016/j.ast.2022.107374
19. Energy-Efficient UAV-Assisted Mobile Edge Computing: Resource Allocation and Trajectory Optimization / M. Li [et al.] // IEEE Transactions on Vehicular Technology. 2020. Vol. 69, iss. 3. Pp. 3424–3438. DOI: 10.1109/TVT.2020.2968343
20. Energy-Efficient Trajectory Optimization for UAV-Assisted IoT Networks / L. Zhang [et al.] // IEEE Transactions on Mobile Computing. 2021. Vol. 21, iss. 12. Pp. 4323–4337. DOI: 10.1109/TMC.2021.3075083
21. Resource Allocation and Trajectory Optimization for QoE Provisioning in Energy-Efficient UAV-Enabled Wireless Networks / F. Zeng [et al.] // IEEE Transactions on Vehicular Technology. 2020. Vol. 69, iss. 7. Pp. 7634–7647. DOI: 10.1109/TVT.2020.2986776

Received: 25.08.2024

Accepted: 16.09.2024

УДК 519.876.5

Применение имитационного моделирования в процессах постройки судна

Соколов Сергей Сергеевич — докт. техн. наук, доцент, профессор кафедры «Комплексное обеспечение информационной безопасности». E-mail: sokolovss@gumrf.ru

Антонова Алена Евгеньевна — аспирант кафедры «Комплексное обеспечение информационной безопасности». Область научных интересов: автоматизация производственных процессов, имитационное моделирование. E-mail: antonovaae@gumrf.ru

Государственный университет морского и речного флота имени адмирала С. О. Макарова, Россия, 198035, Санкт-Петербург, ул. Двинская, 5/7

Для цитирования: Соколов С. С., Антонова А. Е. Применение имитационного моделирования в процессах постройки судна // Интеллектуальные технологии на транспорте. 2024. № 3 (39). С. 44–51. DOI: 10.20295/2413-2527-2024-339-44-51

***Аннотация.** В данной статье представлен пример применения метода имитационного моделирования процессов постройки судна в ПО AnyLogic. Кратко описывается процесс и характеризуется логистика постройки баржи, производимой на верфях, подготовлена карта производственного процесса. В заключение предложены три возможных варианта модернизации, которые заключаются в технологических, организационных улучшениях и сочетании обоих этих решений. ПО AnyLogic использовалось для сравнительного анализа предложенных вариантов применительно к процессу в базовом варианте. Этот метод позволяет получить точные данные для принятия обоснованных решений без необходимости проведения физических тестов. Применение имитационного моделирования является ключевым инструментом для оптимизации процессов судостроения и дальнейшего развития отрасли.*

***Ключевые слова:** имитационное моделирование, картографирование процессов, усовершенствование процессов, судостроительная промышленность*

Введение

Процесс постройки судна является сложным и многоэтапным производством, включающим различные операции, от проектирования и сборки до испытаний и сдачи готового судна. Вместе с тем эффективная производственная логистика играет ключевую роль в обеспечении оптимальной работы и экономической эффективности судостроительных предприятий [1]. Для улучшения производственной логистики и оптимизации процесса постройки судна можно использовать имитационное моделирование.

Применение имитационного моделирования в судостроительной отрасли позволяет оценить различные варианты производственных сценариев, их эффективность и потенциальные улучшения перед внедрением на практике. С помощью имитационного моделирования можно провести анализ процессов постройки судна, начиная с поставки материалов и заканчивая сдачей готового судна [2]. Моделирование позволяет определить оптимальные параметры для достижения максимальной производительности предприятия, минимизации

временных и финансовых затрат, а также обеспечения устойчивости и надежности производственных процессов.

Одним из основных преимуществ применения имитационного моделирования в судостроительной отрасли является возможность проведения симуляций в условиях, близких к реальным, без необходимости тратить время и ресурсы на проведение физических тестов и экспериментов [3]. Таким образом, предприятия могут принимать обоснованные решения на основе точных данных, полученных в результате моделирования.

Картирование процессов — это метод, который заключается в графическом отображении функционирования данного процесса или группы процессов, операций и представлении связей между ними. Однако в большинстве случаев необходимо понимать, из каких действий состоит такой процесс [4].

На технологической карте представлены конкретные действия, которые объединяются в материальные, информационные и финансовые потоки в данном процессе. Такие схемы часто дают менеджерам самое первое всеобъемлющее представление о том, как работает данный процесс. Эксперты разработали набор графических символов, которые представляют различные аспекты процес-

са [5]. На рис. 1 представлены наиболее распространенные символы.

Особенности постройки барж

Первый шаг по подготовке к постройке барж — это отправка списков материалов в отделы снабжения и контроля качества. Отдел снабжения размещает заказы на материалы в соответствии с перечнем материалов, за который отвечает отдел контроля качества, выполняющий их проверку на основе документации и сертификатов.

Следующим шагом является сбор заказанных материалов, которые хранятся на территории предприятия: часть из них — на складе, а большие металлические листы — на открытом воздухе. Затем они маркируются в зависимости от размера, толщины, типа и т. д. Транспортный отдел занимается их доставкой в производственный цех. Обработка материала осуществляется многоканальным способом [6]. Металлические листы различных форматов подаются на станок термической резки с ЧПУ (компьютеризированное числовое управление), где сначала маркируются определенные элементы листа, а затем ранее маркированные элементы вырезаются в соответствии с документацией, переданной в компьютер станка. Некоторые из них позже подвергают обработке (например, придают

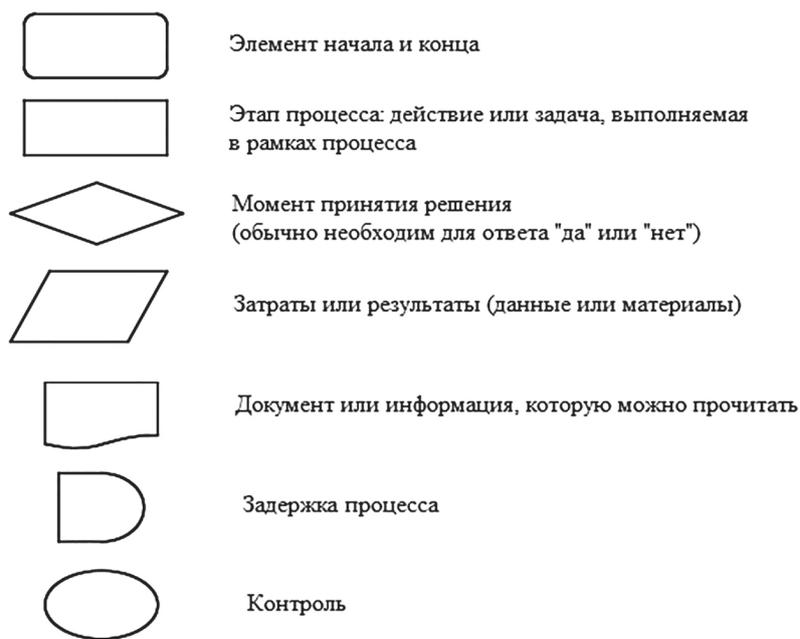


Рис. 1. Наиболее распространенные символы, используемые на картах процессов

форму) на других стойках. Созданные таким образом элементы затем передаются сборочным группам, состоящим из монтажников и сварщиков. Монтажники используют эти элементы для сборки конструкции (с контактными сварными швами), например, секции, корпуса, опоры моста и т. д., в то время как сварщики завершают всю работу.

На этапе изготовления отдельных секций барж происходит так называемая строительная правка, чтобы не возникло неправильного совмещения при сборке секций вместе в секции большего размера. Чтобы наилучшим образом использовать время, работы, связанные с усилением конструкции, выполняются по мере сварки блока [7]. Баржи оснащены всеми видами пропусков и патрубков для водяных, электрических, воздушных, канализационных и нефтяных установок. Однако если данный блок закончен, проводится контроль качества сварных швов и испытания на герметичность.

Следующими производственными этапами являются шлифовка, металлизация и покраска. Шлифование заключается в очистке поверхности материала от всех видов загрязнений с помощью пескоструйной обработки под высоким давлением благодаря сжатому воздуху. Предполагается, что металлизация защищает данную поверхность от коррозии путем напыления тонкого слоя металла (цинка). Покраска — это последний процесс, который заключается в нанесении последующих слоев краски напылением на окрашиваемую поверхность. Следующим процессом является оснащение установки надлежащими устройствами, установками, системами и т. д. Частично это происходит во время покраски, когда конструкция все еще находится на причале, а остальное — после спуска на воду.

Далее баржа проходит окончательное испытание корпуса на герметичность и проверяются все его устройства и системы. Если все пройдет успешно, контракт будет окончательно сдан в эксплуатацию в присутствии ответственного лица.

Составление карты производственного процесса

На основе проведенного анализа процессов, происходящих на предприятии, была составлена

карта логистического процесса, используемого при производстве баржи (рис. 2).

На основе подготовленной карты процесса производства барж предлагаются три вида улучшений:

- 1) технологические усовершенствования;
- 2) организационные улучшения;
- 3) альтернативный процесс, объединяющий оба вышеупомянутых подхода, карта которого представлена на рис. 3.

Технологическое совершенствование направлено на изменения, которые способствовали бы повышению эффективности производства. Большинство машин, используемых в базовом технологическом процессе, можно заменить станками с ЧПУ. Внедрение таких устройств должно сократить продолжительность определенных производственных операций и улучшить их качество.

Применение автоматизированных машин — это долгосрочная инвестиция, которая окупается. Это также позволяет ускорить все производство, благодаря чему на производственном предприятии может приниматься больше заказов для реализации [8]. К сожалению, такая инвестиция является довольно дорогостоящей, но она увеличит прибыль компании в будущем. Внедрение станков с ЧПУ в производственный процесс позволяет быстрее изменять заданные параметры обработки элементов. Такие машины могут запоминать введенные настройки, что позволяет быстро изменять их без необходимости переопределения.

Другая форма совершенствования влечет за собой организационные изменения производства. Это предполагает реализацию метода «точно в срок» [9]. Рекомендуемые изменения также заключаются в ликвидации центрального склада снабжения и межблочного склада для операций резки. Еще одним рекомендуемым изменением является сокращение количества операций по оснащению баржи с двух до одной, которые могут быть выполнены непосредственно после спуска баржи на воду. Для этого все необходимые элементы оборудования и т. д. будут собраны в одном месте, что позволит улучшить организацию работы.

Третье усовершенствование (рис. 3) включает в себя интеграцию технологических

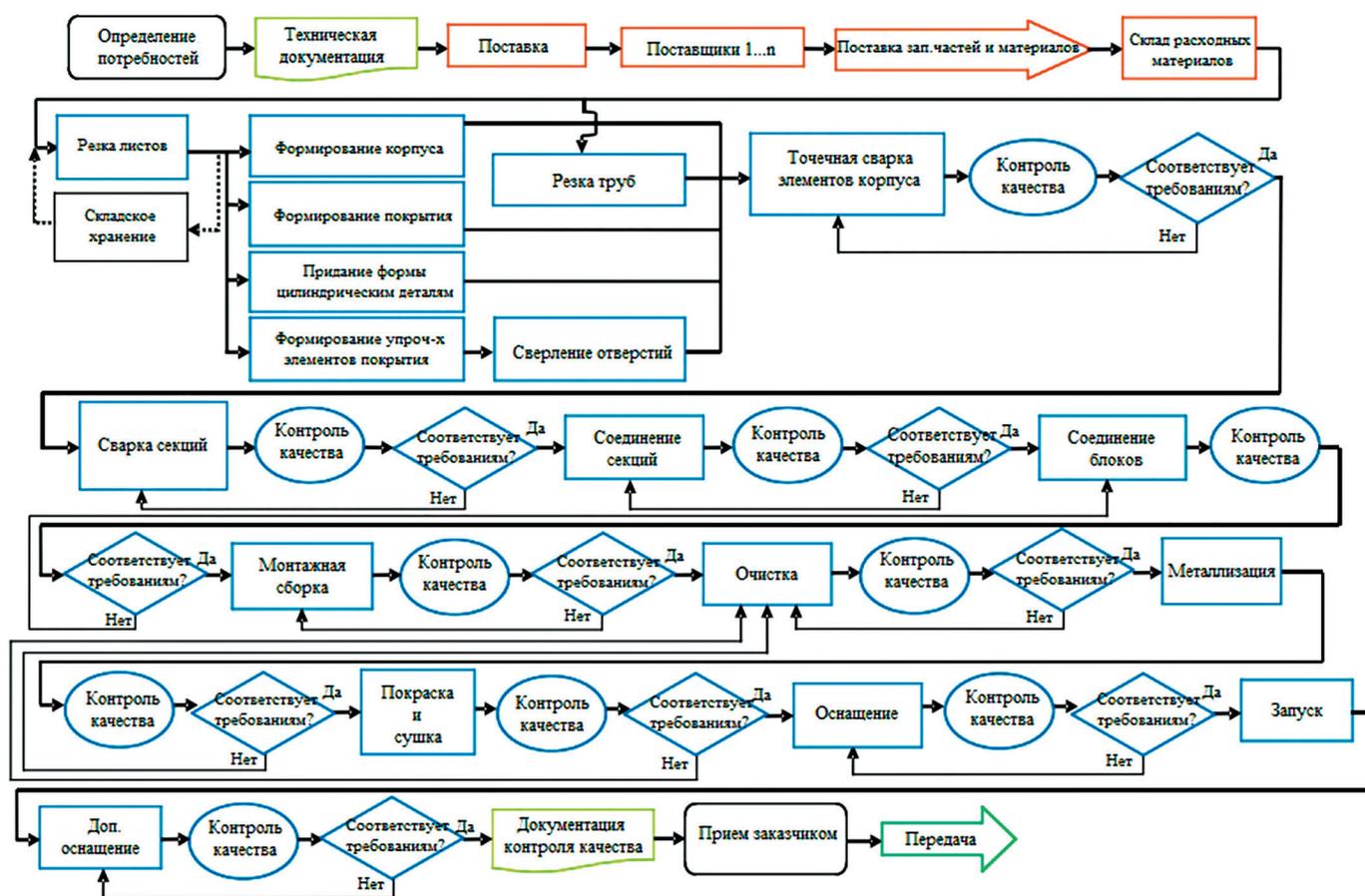


Рис. 2. Карта производственного процесса (состояние до улучшения)

и организационных улучшений, которые должны привести к наибольшему сокращению времени завершения процесса производства одной баржи.

Сравнительный анализ разработанных улучшений в ПО AnyLogic

Был проведен сравнительный анализ с целью определения возможных преимуществ внедрения процесса улучшения. Инструментом, который использовался для этой цели, была программа для компьютерного моделирования и симуляции процессов под названием AnyLogic.

Данное программное обеспечение является единственным инструментом моделирования, который поддерживает все наиболее распространенные существующие на сегодняшний день методологии моделирования: системную динамику, процессно-ориентированное (дискретное событие) и агентно-ориентированное моделирование [10].

На рис. 4 представлена модель базового (до улучшений) процесса производства барж, разработанная с использованием программного обеспечения AnyLogic.

Модель использовалась для проведения имитационного моделирования процесса производства серии из 50 барж. Результаты моделирования, полученные для производственного процесса до внедрения улучшений, показали, что время изготовления одной баржи составило в среднем ~880 часов.

Было замечено, что 70 % произведенных барж были изготовлены в течение 800–850 часов. Однако для оставшихся 30 % барж потребовалось больше времени для постройки (максимальное время изготовления составило ~1295 часов, минимальное — ~812 часов). Это означает, что некоторые изделия не прошли контроль качества с первого раза и потребовались исправления, что значительно увеличило продолжительность производства.

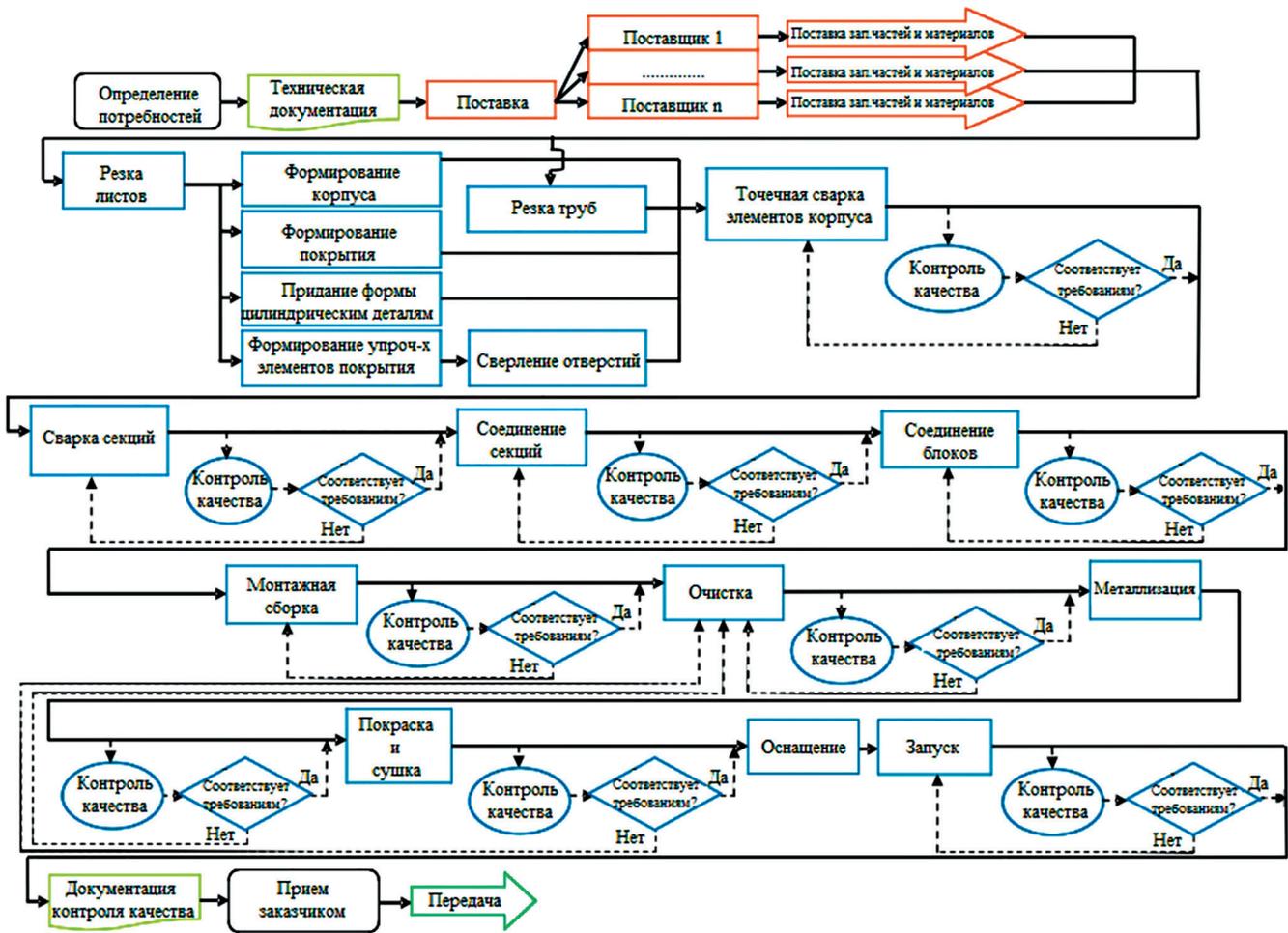


Рис. 3. Карта производственного процесса с технологическими и организационными улучшениями

В случае технологического усовершенствования процесс производства одной баржи длился в среднем ~ 805 часов, что означает, что он был на $\sim 8,5\%$ короче, чем базовый процесс. Целых 78% работ было произведено за $750\text{--}800$ часов. Самое длительное время производства баржи составило ~ 1017 часов, а самое короткое — ~ 679 часов. В результате этого улучшения временной диапазон более ограничен.

Второе улучшение повлекло за собой применение метода «точно в срок». Это позволило ликвидировать склады, сократить количество операций по оснащению до одной, использовать выборочный контроль качества каждого второго или третьего элемента в зависимости от операции. Кроме того, качество продукции было увеличено с 95 до 99% . Благодаря этим изменениям производственный процесс для одной баржи длился в среднем ~ 808 часов. По сравне-

нию с базовым процессом он на $\sim 8,2\%$ короче. В этом случае большая часть барж (32%) была изготовлена в течение $800\text{--}810$ часов. Остальные баржи были изготовлены в более короткие или более продолжительные сроки. Самое продолжительное время производства баржи составило ~ 829 часов, в то время как самое короткое — ~ 774 часа.

Вышеупомянутые типы модернизации сократили время изготовления одной баржи в среднем до ~ 760 часов. По сравнению с базовым процессом он на $\sim 13,5\%$ короче. Моделирование показало, что 22% барж были изготовлены за $765\text{--}770$ часов. Самое продолжительное время производства баржи составило ~ 777 часов, в то время как самое короткое — ~ 737 часов.

Таким образом, исходя из вышеперечисленных показателей можно сделать вывод, что наибольшая разница между минимальным и максимальным

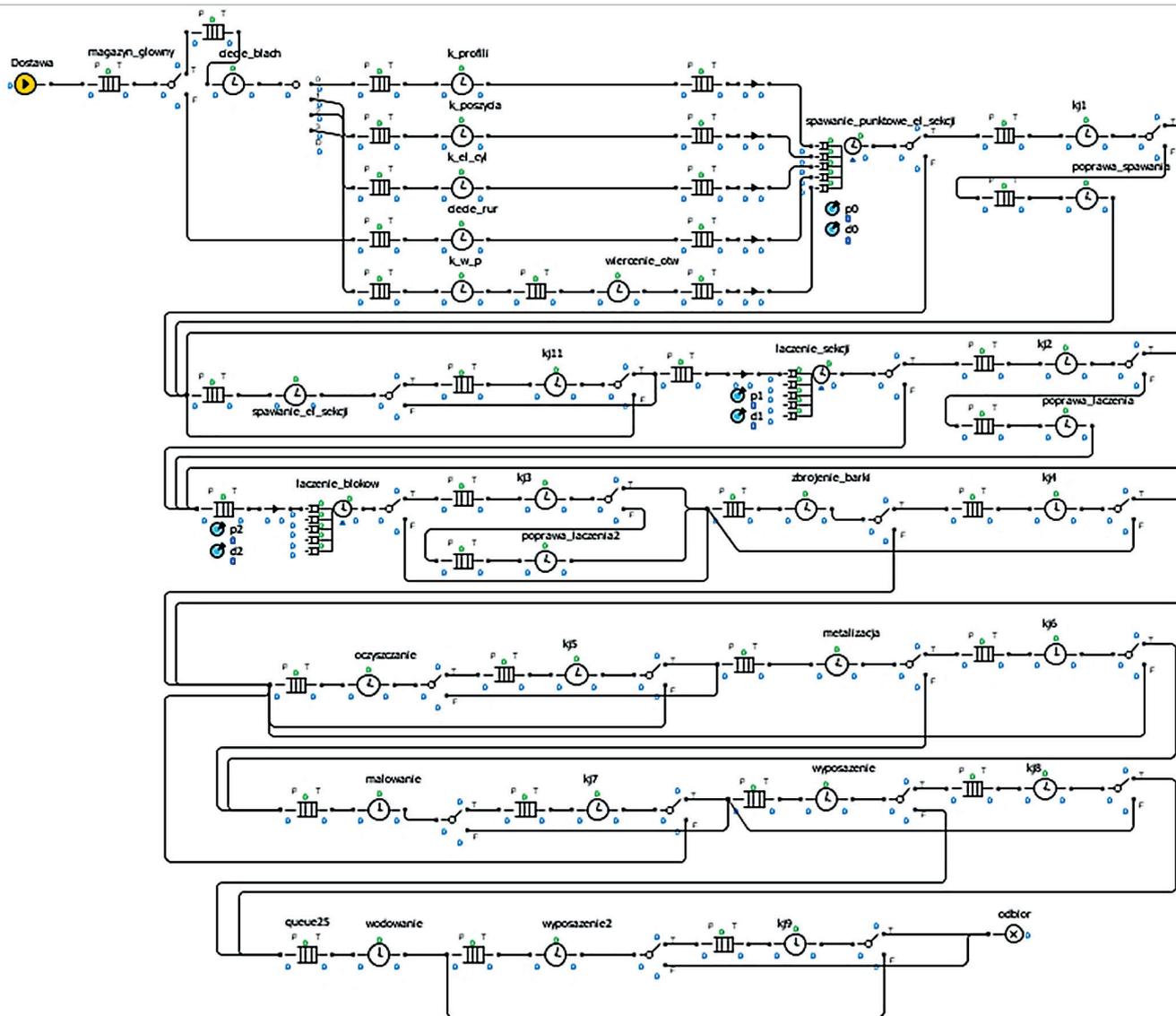


Рис. 4. Модель производственного процесса постройки баржи (состояние до улучшения), составленная в ПО AnyLogic

временем была получена в случае базового процесса и первого улучшения. Это результат проведения контроля качества после большинства операций в этих процессах. Как в случае применения выборочного контроля качества, так и во время второго и третьего улучшений время изготовления баржи сокращается.

Заклучение

Строительство судна — это сложный и многокомпонентный процесс, требующий тщательного планирования, проектирования и реализации. Картография процесса постройки судов играет ключевую роль в обеспечении эффективности, безопасности и качества каждого этапа производства.

Применение имитационного моделирования процессов постройки судна позволяет судостроительным предприятиям повысить эффективность производственной логистики, сократить временные и финансовые затраты, улучшить качество производимой продукции и повысить конкурентоспособность на рынке. Таким образом, имитационное моделирование является эффективным инструментом для оптимизации процессов в судостроительной отрасли и их дальнейшего развития [11].

Данная работа показала, что составление карты производственного процесса с использованием ПО AnyLogic позволяет представить процесс так же, как в методе составления карты, но при этом

провести параметрический анализ потоков и зависимостей между заданными этапами процесса. В результате проведенного имитационного моделирования были сделаны следующие выводы:

- внедрение технологического усовершенствования процесса производства барж может сократить время их изготовления на 75 часов (что составляет 8,5 % от общего времени производства);

- внедрение в процесс производства баржи организационных улучшений может сократить время ее изготовления на 72 часа (что составляет 8,2 % от общего времени производства);

- объединение обоих этих усовершенствований в одно позволяет сократить время производства на 120 часов (что составляет 13,5 % от общего времени изготовления) и соответствует 15 рабочим дням.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Савченко О. В., Половинкин В. Н. Современное состояние, проблемы и перспективы развития отечественного гражданского судостроения // Труды Крыловского государственного научного центра. 2022. Т. 3, № 401. С. 152–164. DOI: 10.24937/2542–2324-2022-3-401-152-164
2. Акопов А. С. Имитационное моделирование: учебник и практикум для академического бакалавриата. М.: Юрайт, 2017. 389 с.
3. Langford J. W. Logistics: Principles and Applications. Second Edition. New York: McGraw Hill, 2007. 578 p.
4. Conger S. A. Process Mapping and Management. New York: Business Expert Press, 2011. 488 p.
5. Evolving Toolbox for Complex Project Management / A. Gorod [et al.] (eds.). Boca Raton (FL): CRC Press, 2019. 544 p.
6. Качанов И. В. Проектирование судов: пособие по курсовому и дипломному проектированию: в 3 ч. Ч. 1. Минск: Белорусский национальный технический ун-т, 2017. 60 с.
7. Народецкий Я. З. Технология постройки барж. М.: Речной транспорт, 1956. 26 с.
8. Лимановская О. В. Имитационное моделирование в AnyLogic 7: учебное пособие: в 2 ч. Ч. 1. Екатеринбург: Изд-во Уральского ун-та, 2017. 152 с.
9. Just-in-Time Systems / R. Z. Ríos-Mercado, Y. A. Ríos-Solís (eds.). New York: Springer New York, 2011. 320 p. DOI: 10.1007/978-1-4614-1123-9
10. AnyLogic: Simulation Modeling Software Tools & Solutions for Business. URL: <http://www.anylogic.com> (assessed: 29.03.2024).
11. Darwish A. Business Process Mapping: A Guide to Best Practice. Patterson Lakes: Writescope Publishers, 2011. 238 p.

Дата поступления: 11.06.2024

Решение о публикации: 17.07.2024

Application of Simulation Modeling in Ship Construction Processes

Sergey S. Sokolov — Doctor of Technical Sciences, Associate Professor, Professor of the Department of Integrated Information Security. E-mail: sokolovss@gumrf.ru

Alyona E. Antonova — Graduate student of the Department of Integrated Information Security. Research interests: automation of production processes, simulation modeling. E-mail: antonovaae@gumrf.ru

Admiral Makarov State University of Marine and River Fleet, 5/7 Dvinskaya str., St. Petersburg, 198035, Russia

For citation: Sokolov S. S., Antonova A. E. Application of Simulation Modeling in Ship Construction Processes // Intellectual Technologies on Transport. 2024. No. 3 (39). Pp. 44–51. DOI: 10.20295/2413-2527-2024-339-44-51. (In Russian)

Abstract. *This article provides an example of the application of the method of simulation modeling of ship construction processes in AnyLogic software. The process is briefly described and the logistics of building a barge produced at shipyards is characterized, and a map of the production process is prepared. In conclusion, three possible modernization options were proposed, which consist in technological, organizational improvements and a variant combining both of these solutions. AnyLogic software was used for a comparative analysis of the proposed options in relation to the process in the basic version. This method allows you to obtain accurate data for making informed decisions without the need for physical tests. The use of simulation modeling is a key tool for optimizing shipbuilding processes and further development of the industry.*

Keywords: *simulation modeling, mapping of processes, process improvements, shipbuilding industry*

REFERENCES

1. Savchenko O. V., Polovinkin V. N. Sovremennoe sostoyanie, problemy i perspektivy razvitiya otechestvennogo grazhdanskogo sudostroeniya // Trudy Krylovskogo gosudarstvennogo nauchnogo centra. 2022. T. 3, № 401. S. 152–164. DOI: 10.24937/2542–2324-2022-3-401-152-164 (In Russian)
2. Akopov A. S. Imitacionnoe modelirovanie: uchebnik i praktikum dlya akademicheskogo bakalavriata. M.: Yurajt, 2017. 389 s. (In Russian)
3. Langford J. W. Logistics: Principles and Applications. Second Edition. New York: McGraw Hill, 2007. 578 p.
4. Conger S. A. Process Mapping and Management. New York: Business Expert Press, 2011. 488 p.
5. Evolving Toolbox for Complex Project Management / A. Gorod [et al.] (eds.). Boca Raton (FL): CRC Press, 2019. 544 p.
6. Kachanov I. V. Proektirovanie sudov: posobie po kursovomu i diplomnomu proektirovaniyu: v 3 ch. Ch. 1. Minsk: Belorusskij nacional'nyj tekhnicheskij un-t, 2017. 60 s. (In Russian)
7. Narodeckij Ya. Z. Tekhnologiya postrojki barzh. M.: Rechnoj transport, 1956. 26 s. (In Russian)
8. Limanovskaya O. V. Imitacionnoe modelirovanie v AnyLogic 7: uchebnoe posobie: v 2 ch. Ch. 1. Ekaterinburg: Izd-vo Ural'skogo un-ta, 2017. 152 s. (In Russian)
9. Just-in-Time Systems / R. Z. Ríos-Mercado, Y. A. Ríos-Solís (eds.). New York: Springer New York, 2011. 320 p. DOI: 10.1007/978-1-4614-1123-9.
10. AnyLogic: Simulation Modeling Software Tools & Solutions for Business. URL: <http://www.anylogic.com> (assessed: 29.03.2024).
11. Darwish A. Business Process Mapping: A Guide to Best Practice. Patterson Lakes: Writescop Publishers, 2011. 238 p.

Received: 11.06.2024

Accepted: 17.07.2024

УДК 681.3

Разработка и внедрение системы управления информационными киосками в университете: подходы, технологии

Забродин Андрей Владимирович	— канд. ист. наук, доцент кафедры «Информационные и вычислительные системы». E-mail: teach-case@yandex.ru
Добровольская Марина Андреевна	— студент, бакалавр, кафедра «Информационные и вычислительные системы». E-mail: dobro-7878@mail.ru
Петрова Елизавета Юрьевна	— студент, бакалавр, кафедра «Информационные и вычислительные системы». E-mail: e.u.petrova@yandex.ru
Федоров Даниил Юрьевич	— студент, бакалавр, кафедра «Информационные и вычислительные системы». E-mail: elfbrus@gmail.com
Хонина Ольга Игоревна	— студент, бакалавр, кафедра «Информационные и вычислительные системы». E-mail: olga.khonina.0311@mail.ru

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: Забродин А. В., Добровольская М. А., Петрова Е. Ю., Федоров Д. Ю., Хонина О. И. Разработка и внедрение системы управления информационными киосками в университете: подходы, технологии // Интеллектуальные технологии на транспорте. 2024. № 3 (39). С. 52–64. DOI: 10.20295/2413-2527-2024-339-52-64

Аннотация. Исследуется возможность и необходимость разработки специализированной системы информирования путем внедрения информационных киосков в образовательную среду. Представлен вариант архитектуры приложения, включающий несколько важных аспектов. Особое внимание уделяется процессу поиска необходимой студенту информации, такой как, например, расписание и информация о наличии библиотечной литературы, что способствует более эффективному обучению в университете. Результаты исследования могут быть полезными для образовательных учреждений, стремящихся улучшить качество образования.

Ключевые слова: информационный киоск, университет, информация, архитектура, разработка, библиотека, карта, расписание

Работа выполнена при поддержке федерального государственного бюджетного образовательного учреждения высшего образования «Петербургский государственный университет путей сообщения Императора Александра I».

Введение

В последние годы университеты стремятся улучшить доступность информации для своих студентов, преподавателей и посетителей путем внедрения современных информационных техно-

логий. Одним из таких решений являются информационные киоски, которые позволяют быстро и удобно получать актуальную информацию о расписании занятий, университетских мероприятиях,

библиографических ресурсах и многом другом. Однако эффективное управление сетью таких киосков требует создания единой системы, обеспечивающей удобное администрирование, актуализацию данных и интеграцию с другими университетскими системами.

В данной статье рассматриваются подходы, технологии и результаты разработки системы управления информационными киосками для университетского корпуса. Описаны этапы анализа требований, проектирования, разработки и внедрения системы, а также представлены результаты ее тестирования и оценки эффективности.

Обзор существующих решений

Перед разработкой системы управления информационными киосками был проведен анализ существующих на рынке решений [1]. Он позволил оценить доступные варианты, выявить их преимущества и недостатки, а также определить наиболее подходящие подходы и технологии для разработки собственной системы, отвечающей специфическим требованиям университета.

Существующие решения можно условно разделить на две категории:

1. Готовые коммерческие системы, такие как 22 Miles, Visix AxisTV Signage Suite и Scala, предлагают широкий набор функций, включая управление контентом, планирование показа, мониторинг работы киосков и интеграцию с другими системами. Они обычно предоставляются по подписке и обладают высокой степенью гибкости и масштабируемости. Однако стоимость таких решений может быть значительной, особенно для крупных университетов с большим количеством киосков [2].

2. Системы с открытым исходным кодом, например, Screenly OSE и Xibo, предоставляют базовый функционал для управления информационными киосками и могут быть настроены под специфические потребности университета. Они обладают преимуществом в виде бесплатного использования и возможности модификации исходного кода. Однако внедрение и поддержка таких систем могут требовать дополнительных ресурсов и квалификации специалистов [3].

Ключевые факторы, которые были учтены при анализе существующих решений:

1) функциональность: наличие необходимых функций, таких как управление контентом, планирование показа, мониторинг работы киосков, интеграция с другими университетскими системами (LMS, библиотечные системы), сбор статистики использования;

2) масштабируемость: возможность расширения системы в будущем при увеличении числа киосков и объема информации;

3) стоимость: соотношение цены и функциональности, учет стоимости лицензий, внедрения и поддержки;

4) интеграция: возможность объединения с существующими университетскими системами для обмена данными и автоматизации процессов;

5) безопасность: наличие механизмов защиты информации от несанкционированного доступа и вандализма;

6) простота использования: удобство интерфейса для администраторов и пользователей киосков.

Анализ существующих решений показал, что ни одно из них полностью не удовлетворяет потребностям университета. Коммерческие системы обладают высокой стоимостью, кроме того, они не соответствуют требованию импортозамещения программного обеспечения в государственной организации. А системы с открытым исходным кодом требуют значительных усилий по настройке и поддержке.

Принято решение разработать собственную систему управления информационными киосками, учитывающую специфику университета и интегрированную с существующими информационными системами. Это снизит затраты на внедрение и поддержку, обеспечит функциональность, безопасность и удобный интерфейс.

Разработка собственной системы управления киосками — оптимальное решение для университета, позволяющее эффективно использовать ресурсы и обеспечить высокий уровень обслуживания.

Известны следующие примеры реализации информационных киосков в других университетах:

- Московский государственный университет имени М. В. Ломоносова. В МГУ информационные киоски используются для предоставления информации о расписании занятий, новостях университета, плане мероприятий, а также для навигации по кампусу.

- Санкт-Петербургский государственный университет. В СПбГУ информационные киоски интегрированы с системой электронной библиотеки и позволяют студентам искать и заказывать книги.

- Национальный исследовательский университет «Высшая школа экономики». В НИУ ВШЭ информационные киоски используются для проведения опросов и сбора обратной связи от студентов и преподавателей.

Опыт внедрения информационных киосков в других университетах подтверждает эффективность этого инструмента для повышения доступности информации и улучшения качества обслуживания студентов, преподавателей и посетителей.

Основания для разработки

Университет проходит цифровую трансформацию. За счет внедрения специализированных программных решений должны быть автоматизированы различные процессы: от расчета учебной нагрузки до создания рабочих учебных планов. Для этого было предложено внедрение технологии «Информационный киоск». Информационные киоски обеспечивают студентов, преподавателей и посетителей университета легким доступом к информации, например, объявлениям, новостям и событиям, а также к образовательным материалам, онлайн-курсам и библиотечным ресурсам, что способствует активной образовательной деятельности студентов [4, 5]. Система управления киосками позволяет оперативно обновлять контент, что важно для своевременного информирования студентов и сотрудников о любых изменениях, будь то расписание занятий или новые мероприятия. Таким образом, информационные киоски играют важную роль в интеллектуальной трансформации учебного заведения, способствуя улучшению доступности информации, коммуникации, инноваций и образовательного процесса в университете [6, 7].

Требования к программному обеспечению

В процессе анализа информационных потребностей университета были выявлены следующие функциональные и технические требования.

Во-первых, система должна предоставлять полную и актуальную информацию об университете, включая его структуру с факультетами, кафедрами и другими подразделениями. Это включает распространение информации о контактах и местоположении всех учебных аудиторий, обеспечивая пользователям легкий доступ к необходимым ресурсам.

Во-вторых, функциональные требования включают инструменты поиска. Студенты и сотрудники должны иметь возможность выполнять поиск по ключевым словам, категориям и фильтрам для быстрого доступа к нужной информации.

В-третьих, очень важны интерактивные возможности системы. Например, расписание студентов и преподавателей должно обновляться автоматически через сайт с помощью API. Это гарантирует, что информация всегда будет актуальной и удобной для использования. Интерактивная карта университета с подробной схемой, номерами аудиторий, фотографиями и графиком работы отделов помогает студентам и гостям легко ориентироваться в университете и находить нужные места.

В-четвертых, чтобы обучение было еще эффективнее, важно подключить к системе информацию из библиотеки. Студенты смогут прямо в киоске проверить, есть ли нужные им книги и учебные материалы в библиотеках университета. Это сильно упростит поиск информации и сделает учебу удобнее.

В-пятых, функциональные требования включают организацию событий и управление новостным потоком через карусель новостей. Это позволяет эффективно информировать студентов и сотрудников об актуальных событиях, достижениях и планах университета [8, 9].

Проектирование системы

Существует несколько подходов к построению веб-приложений. В данной работе представлен подход на основе «толстого клиента» —

приложения, обеспечивающего расширенную функциональность независимо от центрального сервера. Модель приложения состоит из серверных и клиентских компонентов [10]. К серверным компонентам относится связь Python с фреймворком FastAPI, а к клиентским — React, который используется для разработки приложения, обеспечивая интерактивность и динамическое обновление данных на веб-странице [11]. Архитектура приложения представлена на рис. 1.

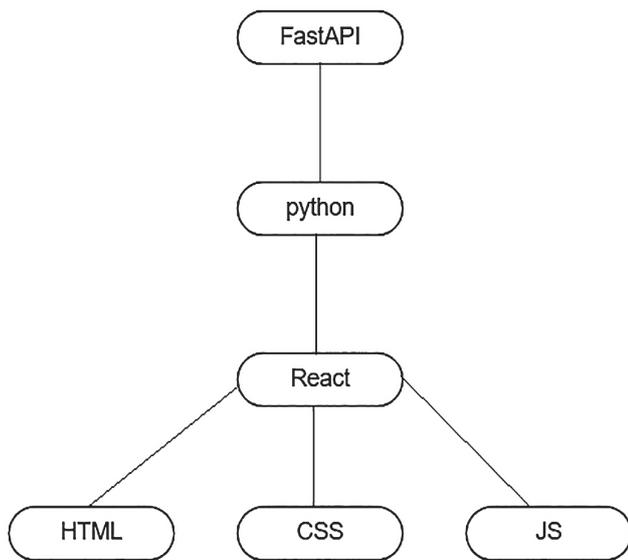


Рис. 1. Архитектура приложения

Описание компонентов

В качестве языка программирования для разработки серверной части приложения выбран Python, известный своей простотой, удобочитаемостью и широким использованием в веб-разработке [4]. Для создания серверной части веб-приложения применяется FastAPI — современный веб-фреймворк Python. FastAPI использует стандарт ASGI, что обеспечивает асинхронное, конкурентное соединение с клиентами и высокую производительность. При необходимости FastAPI может работать и с WSGI. Важной особенностью FastAPI является возможность использования асинхронных функций для маршрутов и конечных точек, что повышает эффективность обработки запросов. Кроме того, FastAPI позволяет писать чистый и современный код с использова-

нием подсказок типов, что улучшает читаемость и упрощает отладку.

Для разработки внешних пользовательских интерфейсов выбрана библиотека JavaScript с открытым кодом — React. Библиотека React позволяет создавать как одностраничные, так и многостраничные приложения, обеспечивая высокую скорость разработки, простоту и масштабируемость.

В React для определения структуры и содержания компонентов используется JSX, позволяющий встраивать HTML (HyperText Markup Language — стандартный язык разметки для создания веб-страниц) непосредственно в код JavaScript. Для описания внешнего вида элементов HTML используется CSS (Cascading Style Sheets — язык стилей). React предоставляет несколько способов подключения CSS: через внешние файлы, CSS-модули, styled-components или inline-стили.

JavaScript является основным языком программирования, на котором пишутся компоненты и логика приложения в React. Он позволяет создавать интерактивный и динамичный контент на веб-страницах, обеспечивая отзывчивость и удобство взаимодействия с пользователем.

Выбранный стек технологий (Python, FastAPI, React, HTML, CSS, JavaScript) представляет собой мощный и гибкий инструмент для разработки современных и эффективных веб-приложений, что делает его оптимальным выбором для создания системы управления информационными киосками.

Взаимодействие компонентов

Архитектура взаимодействия частей программного комплекса «Информационный киоск», внутренней сети университета и внутренней сети библиотеки показана на рис. 2.

Разработка пользовательского интерфейса

Перед началом разработки интерфейса приложения были изучены этапы разработки UI-дизайна [12] и проведено исследование, в ходе которого выявлено, что целевой аудиторией являются студенты, преподаватели и абитуриенты. Для пользователей приложения необходимо, чтобы программа выполняла ряд функций:

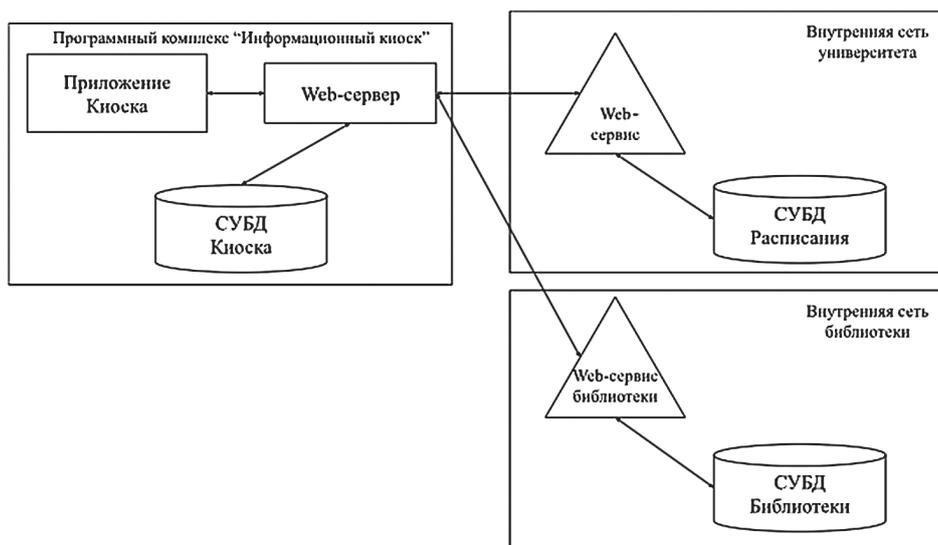


Рис. 2. Архитектура взаимодействия компонентов

- показывала карту, где можно посмотреть местонахождение аудиторий;
- расписание, в котором можно сделать поиск по группе, преподавателю или аудитории;
- содержала информацию о мероприятиях, на которые можно зарегистрироваться;
- бота-помощника, который помогает с вопросами по организации университета; дает информацию о книгах в библиотеке [5].

Примерами приложений с похожей тематикой могут служить информационные киоски в торговых центрах или других университетах. Для выявления основных композиций элементов, принципов работы, лучших практик, удобных интерфейсов и привлекательных визуальных решений были выбраны планшетные приложения и версии сайтов.

Ключевыми элементами являются:

- хедер с текущими датой и временем, названием страницы, временем до сессии;
- навигационные иконки;
- два вида календарей — на месяц с расписанием событий и на неделю с индивидуальным расписанием на неделю;
- всплывающие окна с подробной информацией;
- карточки книг;
- карта вуза по корпусам и этажам.

Следующим этапом разработки было создание варфреймов — макетов с примерным расположением

элементов на экране. На этом этапе приложение выглядит как набор экранов с абстрактными фигурами, но такие макеты помогают увидеть общую картину и расположить блоки логически на экране, а также продумать взаимосвязь экранов между собой.

После идет этап воплощения идеи в жизнь, а именно продумывание дизайна. Подбираются иконки, графические элементы, цветовая палитра, логические блоки и прочее. Было принято решение сделать приложение в минималистичном стиле. Для этого подбирались иконки с тонкой обводкой, за основу цветовой палитры были взяты фирменные цвета университета и их оттенки. Фоны для основных экранов сделаны неброскими, чтобы не отвлекать пользователя от информации на экране.

Во время разработки дизайна для киоска учитывались все эвристики Нильсена [13]:

- соответствие системы реальному миру (match between the system and the real world);
- согласованность и стандарты (consistency and standards);
- предотвращение ошибок (error prevention);
- на виду, а не в памяти (recognizing rather than recall);
- эстетичный и минималистичный дизайн (aesthetic and minimalist design);
- понимание проблем и их решение (help users recognize, diagnose, and recover from errors).

Рассмотрим каждую эвристику подробнее на примерах интерфейсов [14].

Соответствие системы реальному миру

Первая эвристика Нильсена гласит: интерфейс должен говорить с пользователем на одном (знакомом ему) языке — просто и однозначно.

Примером соответствия данной эвристике служат иконки, которые используются для навигации по страницам. Для обозначения главной (домашней) страницы используется иконка дома, для страницы расписания — учебники, для интерактивной карты корпусов вуза — изображение геометки, для страницы библиотеки — иконка книги. А также на легенде карты используются понятные для человека иконки. Примеры показаны на рис. 3.

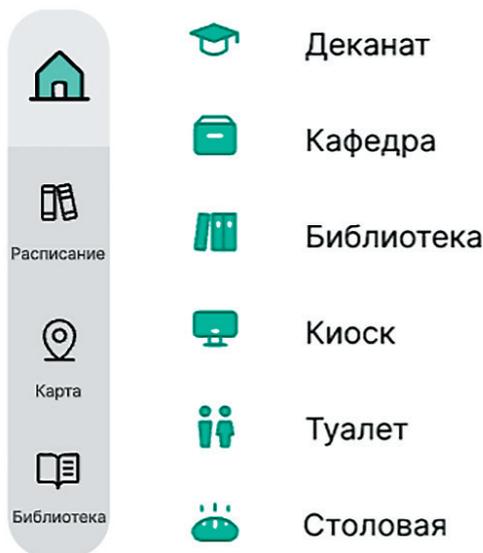


Рис. 3. Иконки, которые используются для навигации

Согласованность и стандарты

Эвристика вторая гласит: если у системы есть стандарты, их лучше придерживаться. Пользователи привыкли к определенным решениям, а неоправданное отклонение от них затрудняет взаимодействие с интерфейсом.

Также к этой эвристике относится единый дизайн для всего приложения. Для этого собирается UI-kit (рис. 4) — полный набор элементов и компонентов, необходимый для сборки большого однородного продукта. Он включает различные кнопки, иконки, поля для ввода данных и т. д., что позволяет сохранять узнаваемость продукта [15].

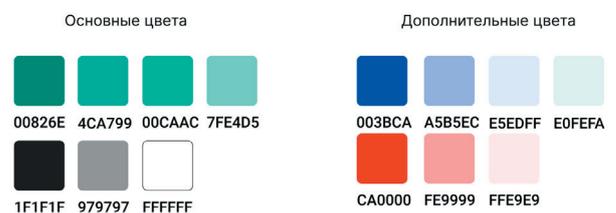
Шрифт

Использованные шрифты для реализации дизайна киоска



Цвета

Использованные цвета для реализации дизайна киоска



Иконки

Использованные иконки для реализации дизайна киоска



Кнопки

Использованные иконки для реализации дизайна киоска

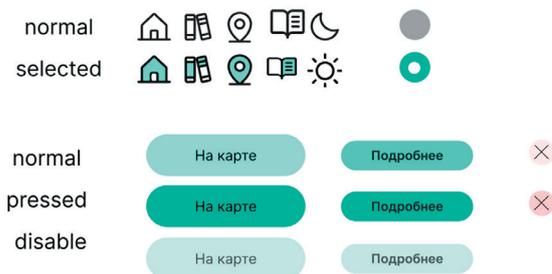


Рис. 4. UI-kit проекта

Пользователи привыкли, что строка поиска чаще всего сопровождается фильтрами, которые реализованы переключателями (радиокнопка) или чекбоксами (рис. 5).

Предотвращение ошибок

Эвристика третья: лучше предупредить о возможности ошибки, чем придумывать красивые и содержательные сообщения о ее последствиях. Явным примером является поисковая строка, в которой есть вариант запроса, чтобы пользователю было легче начать поиск (рис. 6).

На виду, а не в памяти

Эвристика четвертая гласит: задача дизайнера — помочь пользователю всегда оставаться в контексте и ничего не упустить.

Примером служат подписи к навигационным иконкам, которые помогают пользователю ориентироваться в страницах. Также примером является страница библиотеки, где кроме подробной информации есть также фото обложек, показанные на рис. 7.

Эстетичный и минималистичный дизайн

Эвристика номер пять: стремись к простоте там, где это возможно. Каждое лишнее слово и каждый

лишний элемент интерфейса затрудняют восприятие информации.

Например, в расписании основной информацией являются название предмета, время занятия и тип, а вот информацию о преподавателе и аудитории можно скрыть. Но совсем без этой информации не обойтись, поэтому она выводится в отдельном всплывающем окне (рис. 8).

Понимание проблем и их решение

Шестая эвристика гласит: если при использовании сайта возникает проблема, нужно как можно яснее ее описать и предложить варианты решения.

Соответствие этой эвристике можно увидеть, если пользователь введет неправильный запрос. Например, студент не заметил, что установлен фильтр по названию группы, и ввел фамилию преподавателя. Ему будет выведена ошибка с указанием на неправильный запрос.



Рис. 5. Поисковая строка с фильтрами

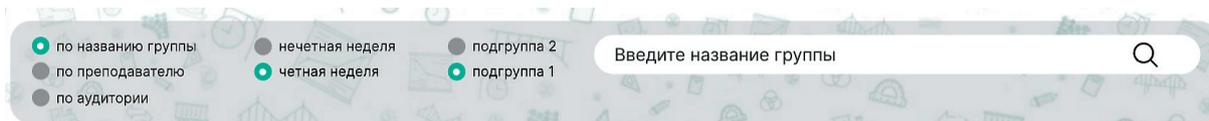


Рис. 6. Поисковая строка с подсказкой для пользователя

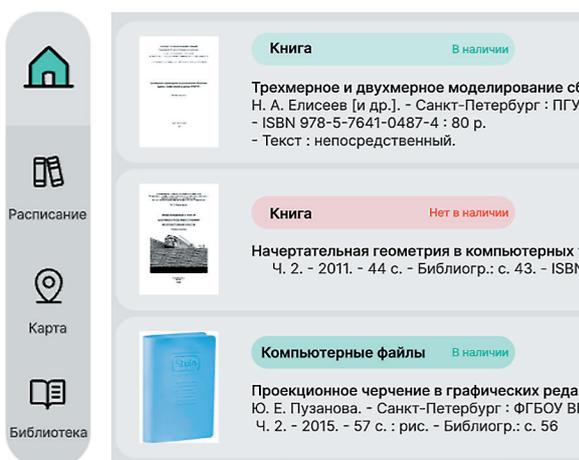


Рис. 7. Навигация по страницам

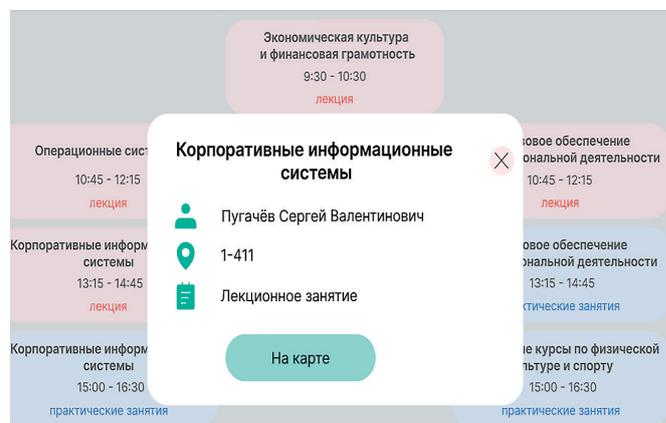


Рис. 8. Всплывающее окно с подробной информацией о занятии

Основные интерфейсы системы и их функциональные возможности

Основываясь на знаниях, полученных при изучении основных правил дизайна, был спроектирован дизайн интерфейса системы (рис. 9–12).

Реализация серверной части

В рамках проекта особое внимание уделяется разработке эффективной серверной части (бэкенда) системы. Бэкенд состоит из нескольких ключевых компонентов, обеспечивающих функциональность информационного киоска:

1. API для расписания: данный компонент отвечает за взаимодействие с сайтом ПГУПС с целью получения актуального расписания занятий.

Он использует библиотеку requests для отправки HTTP-запросов к страницам сайта, содержащим расписание, и получает HTML-документы в ответ. Далее библиотека BeautifulSoup4 используется для парсинга полученного HTML и преобразования данных в удобный формат JSON. Такой подход обеспечивает эффективную обработку и использование данных о расписании в других частях системы.

2. Главный экран: этот компонент отвечает за автоматизированный сбор новостей с сайта ПГУПС. Процесс сбора основан на использовании библиотеки Selenium, которая имитирует действия пользователя в браузере. Selenium в сочетании с драйвером ChromeDriver позволяет переходить



Рис. 9. Интерфейс главной страницы информационного киоска

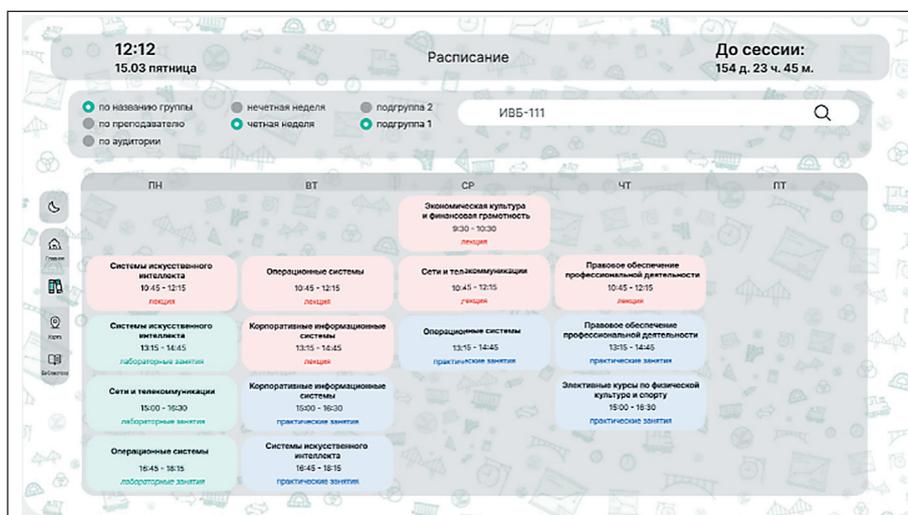


Рис. 10. Интерфейс страницы «Расписание»



Рис. 11. Интерфейс страницы «Интерактивная карта Университета»



Рис. 12. Интерфейс страницы «Библиотека»

по ссылкам на страницы с новостями, извлекая необходимую информацию: заголовки, даты публикации и описания новостей. Собранные данные, подобно данным о расписании, сохраняются в формате JSON для удобства дальнейшего использования.

3. Карта: компонент предоставляет информацию о расположении кабинетов, необходимую для отображения на интерактивной карте университета. Данные о кабинетах хранятся в формате JSON, что облегчает их чтение и отображение на странице.

4. Библиотека: данный компонент обеспечивает интеграцию с API библиотеки университета. Через программный интерфейс формируются запросы к сайту библиотеки (<https://library.pgups.ru>).

В ответ на запрос API возвращает данные о книгах, включая авторов, библиографические описания, штрих-коды и другую информацию.

Выбранная модульная структура бэкенда обеспечивает гибкость и масштабируемость системы, позволяя легко добавлять новые функции и интегрироваться с другими университетскими сервисами.

Механизмы обеспечения безопасности и надежности работы системы

Для обеспечения бесперебойной и безопасной работы информационных киосков, помимо разработки функционала, особое внимание уделяется вопросам надежности и защиты данных. Внедренная система включает в себя ряд механизмов,

направленных на предотвращение потери информации и обеспечение стабильности работы:

- Резервное копирование и восстановление данных.

Регулярное создание резервных копий данных и их хранение в защищенном месте — краеугольный камень стратегии обеспечения сохранности информации. Эта мера позволяет быстро восстановить функциональность системы в случае сбоя, будь то аппаратные или программные ошибки, а также минимизировать последствия возможных злоумышленных действий.

- Обновление программного обеспечения.

Постоянное обновление всех компонентов программного обеспечения, включая операционные системы, прикладное ПО и антивирусные программы, является критически важным для поддержания безопасности системы. Своевременная установка обновлений позволяет закрыть известные уязвимости, предотвращая потенциальные атаки и обеспечивая устойчивость к новым угрозам.

- Регулярное обслуживание и техническая поддержка.

Плановое техническое обслуживание оборудования и оперативная техническая поддержка играют ключевую роль в обеспечении стабильности работы системы. Регулярные проверки позволяют выявить и устранить потенциальные проблемы на ранней стадии, предотвращая возникновение серьезных сбоев. Оперативное реагирование на запросы пользователей и быстрое решение возникающих проблем гарантируют комфортную и эффективную работу с информационными киосками.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Максимов И. И. Технология использования информационных терминалов при построении информационной системы / науч. рук. О. Н. Моргунова // Актуальные проблемы авиации и космонавтики: сборник материалов XIII Международной научно-практической конференции, посвященной Дню космонавтики (Красноярск, 10–14 апреля 2017 г.): в 3 т. Красноярск: Сиб. гос. аэрокосмич. ун-т, 2017. Т. 2. С. 116–118.

2. Разработка многомодального информационного киоска / А. Л. Ронжин [и др.] // Труды СПИИРАН. 2007. Вып. 5. С. 227–245. DOI: 10.15622/sp.5.12.

3. Филиппов Д. С. Методы адаптивной трансформации контента в форме гипертекста на корпоративном ТВ и информационных киосках вуза // Открытое образование. 2024. Т. 28, № 3. С. 56–64. DOI: 10.21686/1818-4243-2024-3-56-64.

4. Информационные технологии в образовании: учебник для вузов / Е. В. Баранова [и др.]. СПб.: Лань, 2022. 296 с.

Заключение

В работе продемонстрировано, что внедрение информационных киосков может играть важную роль в сокращении эксплуатационных расходов, поскольку они оптимизируют процессы и снижают потребность в трудоемких услугах. Кроме того, они предоставляют ценную информацию о предпочтениях, поведении студентов и работников, позволяя университету адаптировать свои сервисы. Это существенно облегчит работу учебных заведений и взаимодействие студентов с информационной средой.

Тем не менее рассмотренная система не является законченной и имеет потенциал для дальнейшего совершенствования. Оно включает в себя расширение возможностей, таких как распознавание лиц, информационные помощники, возможность авторизации и онлайн-регистрации на конференции, что позволит сделать систему более функциональной и удобной для использования в реальных образовательных процессах.

Внедрение информационных киосков в экосистему университета является перспективной разработкой. Таким образом, проведенное в рамках данной работы исследование улучшает процесс работы университета. Несмотря на обширную функциональность, работа над программой продолжается и включает в себя механизм авторизации и регистрации на конференции онлайн, а также дальнейшую интеграцию приложения в структуру системы обучения. Данная технология активно внедряется в подобные структуры и в скором времени станет массовой заменой бумажным информационным стендам.

5. Ковтунов Н. А. Особенности разработки веб-приложений для информационных киосков // Вопросы технических и физико-математических наук в свете современных исследований: сборник статей по материалам V–VI Международной научно-практической конференции. 2018. № 5–6 (4). С. 5–9.
6. Буслова Н. С., Клименко Е. В. Информационный киоск как компонент информационно-образовательной среды учебного заведения // Информатизация образования: теория и практика: сборник материалов международной научно-практической конференции (Омск, 21–22 ноября 2014 г.) / под общ. ред. М. П. Лапчика. Омск: Омский гос. пед. ун-т, 2014. С. 111–113.
7. Фомин С. И., Малахов А. В., Эсаулов К. А. Повышение доступности информации в образовательных учреждениях с использованием информационных киосков // Вопросы образования и науки: теоретический и методический аспекты: сборник научных трудов по материалам международной заочной научно-практической конференции (Тамбов, 30 апреля 2012 г.): в 7 ч. Ч. 5. Тамбов: Юком, 2012. С. 134–135.
8. Карасев Д. Н., Кравченко С. В. Проектирование интерактивного путеводителя по университету // Современные технологии в науке и образовании (СТНО-2018): сборник трудов международного научно-технического форума (Рязань, 28 февраля — 2 марта 2018 г.): в 11 т. / под общ. ред. О. В. Миловзорова. Т. 4. Рязань: Рязанский гос. радиотехн. ун-т, 2018. С. 58–61.
9. Нагих С. С., Лунев В. К. Разработка и проектирование информационного терминала / науч. рук. Л. А. Попова // Проблемы социального и научно-технического развития современном мире: материалы XIX Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых (с международным участием) (Рубцовск, 20–21 апреля 2017 г.). Рубцовск: Рубцовский индустриальный ин-т, 2017. С. 44–46.
10. Томашевская В. С., Яковлев Д. А. Стандарты, на которые можно ориентироваться при реализации информационно-справочного киоска // Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем «Радиоинфоком-2019»: сборник научных статей IV Международной научно-практической конференции (Москва, 11–15 ноября 2019 г.). М.: МИРЭА — Российский технологический университет, 2019. С. 452–456.
11. Музыченко Е. Н., Папанова А. С. Информационный киоск в наше время // Наука и образование: отечественный и зарубежный опыт: сборник трудов 26-й Международной научно-практической конференции (Белгород, 23 декабря 2019 г.). Белгород: ГиК, 2019. С. 89–91.
12. Никоненко С. Что такое пользовательский интерфейс и как происходит разработка UI // Блог студии Purrweb. URL: <http://www.purrweb.com/ru/blog/chto-takoe-polzovatelskij-interfejs-i-kak-proiskhodit-razrabotka-ui> (дата обращения 15.07.2024).
13. Nielsen J. 10 Usability Heuristics for User Interface Design // Nielsen Norman Group. URL: <http://www.nngroup.com/articles/ten-usability-heuristics> (дата обращения 20.07.2024).
14. Заяц А. М., Ульянов А. В., Яловка Л. А. Информационный киоск кафедры ИСиТ // Информационные системы и технологии: теория и практика: сборник научных трудов. Вып. 7 / отв. ред. А. М. Заяц. СПб.: Санкт-Петербургский гос. лесотехн. ун-т, 2015. С. 3–4.
15. Усикова Е. Что такое UI-кит // Точка зрения от Bang Bang Education. URL: <https://bangbangeducation.ru/point/ux-ui-dizain/chto-takoe-ui-kit> (дата обращения 02.07.2024).

Дата поступления: 28.08.2024

Решение о публикации: 18.09.2024

Development and Implementation of a Management System for Information Kiosks at the University: Approaches, Technologies

- Andrey V. Zabrodin** — PhD in Historical Sciences, Associate Professor of the Department of Information and Computing Systems. E-mail: teach-case@yandex.ru
- Marina A. Dobrovolskaya** — student, Undergraduate student, Department of 'Information and Computing Systems. E-mail: dobro-7878@mail.ru
- Elizaveta Yu. Petrova** — student, Undergraduate student, Department of 'Information and Computing Systems. E-mail: e.u.petrova@yandex.ru
- Daniil Yu. Fedorov** — student, Undergraduate student, Department of 'Information and Computing Systems. E-mail: elfbrus@gmail.com
- Olga I. Khonina** — student, Undergraduate student, Department of 'Information and Computing Systems. E-mail: olga.khonina.0311@mail.ru

Emperor Alexander St. Petersburg State Transport University, 9, Moskovsky pr., Saint Petersburg, 190031, Russia

For citation: Zabrodin A. V., Dobrovolskaya M. A., Petrova E. Yu., Fedorov D. Yu., Khonina O. I. Development and Implementation of a Management System for Information Kiosks at the University: Approaches, Technologies // *Intellectual Technologies on Transport*. 2024. No. 3 (39). Pp. 52–64. DOI: 10.20295/2413-2527-2024-339-52-64 (In Russian)

Abstract. *The article explores the possibility and necessity of developing a specialized information system by introducing information kiosks into the educational environment. A variant of the application architecture is presented, which includes several important aspects. Particular attention is paid to the process of finding the information the student needs, such as schedules and information about the availability of library literature, which contributes to more effective studying at the university. The results of the study may be useful for educational institutions seeking to improve the quality of education.*

Keywords: *information kiosk, university, information, architecture, development, library, map, schedule*

REFERENCES

1. Maksimov I. I. Tekhnologiya ispol'zovaniya informacionnyh terminalov pri postroenii informacionnoj sistemy / nauch. ruk. O. N. Morgunova // Aktual'nye problemy aviatsii i kosmonavtiki: Sbornik materialov XIII Mezhdunarodnoj nauchno-prakticheskoy konferencii, posvyashchennoj Dnyu kosmonavtiki (Krasnoyarsk, 10–14 aprelya 2017 g.): v 3 t. Krasnoyarsk: Sib. gos. aerokosmich. un-t, 2017. T. 2. S. 116–118. (In Russian)
2. Razrabotka mnogomodal'nogo informacionnogo kioska / A. L. Ronzhin [i dr.] // Trudy SPIIRAN. 2007. Vyp. 5. S. 227–245. DOI: 10.15622/sp.5.12. (In Russian)
3. Filippov D. S. Metody adaptivnoj transformacii kontenta v forme giperteksta na korporativnom TV i informacionnyh kioskah vuza // Otkrytoe obrazovanie. 2024. T. 28, № 3. S. 56–64. DOI: 10.21686/1818-4243-2024-3-56-64. (In Russian)
4. Informacionnye tekhnologii v obrazovanii: uchebnik dlya vuzov / E. V. Baranova [i dr.]. SPb.: Lan', 2022. 296 s. (In Russian)
5. Kovtunov N. A. Osobennosti razrabotki veb-prilozhenij dlya informacionnyh kioskov // Voprosy tekhnicheskikh i fiziko-matematicheskikh nauk v svete sovremennykh issledovanij: sbornik statej po materialam V–VI Mezhdunarodnoj

nauchno-prakticheskoy konferencii. 2018. № 5–6 (4). S. 5–9. (In Russian)

6. Buslova N. S., Klimenko E. V. Informacionnyj kiosk kak komponent informacionno-obrazovatel'noj sredy uchebnogo zavedeniya // Informatizaciya obrazovaniya: teoriya i praktika: sbornik materialov mezhdunarodnoj nauchno-prakticheskoy konferencii (Omsk, 21–22 noyabrya 2014 g.) / pod obshch. red. M. P. Lapchika. Omsk: Omskij gos. ped. un-t, 2014. S. 111–113. (In Russian)

7. Fomin S. I., Malahov A. V., Esaulov K. A. Povyshenie dostupnosti informacii v obrazovatel'nyh uchrezhdeniyah s ispol'zovaniem informacionnyh kioskov // Voprosy obrazovaniya i nauki: teoreticheskij i metodicheskij aspekty: sbornik nauchnyh trudov po materialam mezhdunarodnoj zaochnoj nauchno-prakticheskoy konferencii (Tambov, 30 aprelya 2012 g.): v 7 ch. CH. 5. Tambov: Yukom, 2012. S. 134–135. (In Russian)

8. Karasyov D. N., Kravchenko S. V. Proektirovanie interaktivnogo putevoditelya po universitetu // Sovremennye tekhnologii v nauke i obrazovanii (STNO-2018): sbornik trudov mezhdunarodnogo nauchno-tekhnicheskogo foruma (Ryazan', 28 fevralya — 2 marta 2018 g.): v 11 t. / pod obshch. red. O. V. Milovzorova. T. 4. Ryazan': Ryazanskij gos. radiotekhn. un-t, 2018. S. 58–61. (In Russian)

9. Nagih S. S., Lunev V. K. Razrabotka i proektirovanie informacionnogo terminala / nauch. ruk. L. A. Popova // Problemy social'nogo i nauchno-tekhnicheskogo razvitiya sovremennom mire: materialy XIX Vserossijskoj nauchno-tekhnicheskoy konferencii studentov, aspirantov i molodyh uchenyh (s mezhdunarodnym uchastiem) (Rubcovsk, 20–21 aprelya 2017 g.). Rubcovsk: Rubcovskij industrial'nyj in-t, 2017. S. 44–46. (In Russian)

10. Tomashevskaya V. S., Yakovlev D. A. Standarty, na kotorye možno orientirovat'sya pri realizacii informacionno-spravochnogo kioska // Aktual'nye problemy i perspektivy razvitiya radiotekhnicheskikh i infokommunikacionnyh sistem «Radioinfokom-2019»: sbornik nauchnyh statej IV Mezhdunarodnoj nauchno-prakticheskoy konferencii (Moskva, 11–15 noyabrya 2019 g.). M.: MIREA — Rossijskij tekhnologicheskij universitet, 2019. S. 452–456. (In Russian)

11. Muzychenko E. N., Papanova A. S. Informacionnyj kiosk v nashe vremya // Nauka i obrazovanie: otechestvennyj i zarubezhnyj opyt: sbornik trudov 26-j Mezhdunarodnoj nauchno-prakticheskoy konferencii (Belgorod, 23 dekabrya 2019 g.). Belgorod: GiK, 2019. S. 89–91. (In Russian)

12. Nikonenko S. Chto takoe pol'zovatel'skij interfejs i kak proiskhodit razrabotka UI // Blog studii Purweb. URL: <http://www.purweb.com/ru/blog/chto-takoe-polzovatel'skij-interfejs-i-kak-proiskhodit-razrabotka-ui> (data obrashcheniya 15.07.2024). (In Russian)

13. Nielsen J. 10 Usability Heuristics for User Interface Design // Nielsen Norman Group. URL: <http://www.nngroup.com/articles/ten-usability-heuristics> (data obrashcheniya 20.07.2024).

14. Zayac A. M., Ul'yanov A. V., Yalovka L. A. Informacionnyj kiosk kafedry ISiT // Informacionnye sistemy i tekhnologii: teoriya i praktika: sbornik nauchnyh trudov. Vyp. 7 / otv. red. A. M. Zayac. SPb.: Sankt-Peterburgskij gos. lesotekhn. un-t, 2015. S. 3–4. (In Russian)

15. Usikova E. Chto takoe UI-kit // Tochka zreniya ot Bang Bang Education. URL: <https://bangbangeducation.ru/point/ux-ui-dizain/chto-takoe-ui-kit> (data obrashcheniya 02.07.2024). (In Russian)

Received: 28.08.2024

Accepted: 18.09.2024

УДК 004.056

Разработка DLP-модуля для защиты персональных данных в распределенной автоматизированной системе

Ковтуненко Аркадий Алексеевич — специалист по защите информации, выпускник кафедры «Информатика и информационная безопасность». Область научных интересов: разработка программного обеспечения, информационная безопасность. E-mail: bernadod2000@mail.ru

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: Ковтуненко А. А. Разработка DLP-модуля для защиты персональных данных в распределенной автоматизированной системе // Интеллектуальные технологии на транспорте. 2024. № 3 (39). С. 65–72. DOI: 10.20295/2413-2527-2024-339-65-72

Аннотация. Рассматривается важность защиты персональных данных в автоматизированных системах. Представлено исследование о разработке программного модуля для защиты конфиденциальной информации, в том числе персональных данных, в распределенной автоматизированной системе. **Цель исследования:** создание программного средства, блокирующего утечки конфиденциальной информации. Разработанный программный модуль предназначен для блокировки утечки конфиденциальной информации по сетевым каналам и съемным носителям с помощью модели машинного обучения с централизованным управлением администратором безопасности. **Практическая значимость:** использование данного модуля в организации включает обеспечение централизованного управления безопасностью информации, снижение риска утечек конфиденциальных данных, а также поддержку в расследованиях инцидентов безопасности. Исследование может быть применимо как временная мера в период обновления и адаптации системы под новые реалии до внедрения полноценной системы защиты.

Ключевые слова: персональные данные, система управления информационной безопасностью, утечка конфиденциальной информации, машинное обучение, DLP (Data Leaks Protection)

Введение

В последние годы задача обеспечения безопасности конфиденциальной информации и в частности персональных данных (ПДн) стала как никогда актуальна. По данным экспертно-аналитического центра группы компаний InfoWatch [1], количество скомпрометированных записей ПДн в России за вторую половину 2022 года в 1,5 раза превысило их объем за первую половину того же года, а только в первой половине 2023 года было скомпрометировано почти столько же единиц ПДн, сколько за весь 2022 год — около 705 млн записей. И пусть количество утечек за последние годы несколько

снизилось, их объемы неуклонно растут в поражающих масштабах. Именно поэтому защита персональных данных в наше время — актуальная и крайне важная задача.

Огромные объемы персональных данных обрабатываются и хранятся в распределенных автоматизированных системах (АС). Распределенные АС используются в различных отраслях, включая здравоохранение, государственное управление и чрезвычайные службы. Именно это, а также их структура, в которой данные распределены между различными подразделениями и регионами, повышает сложность

защиты информации в таких системах и делает их уязвимыми для различных угроз информационной безопасности (ИБ). При этом утрата или компрометация ПДн может не только подорвать доверие общества к организации, но и привести к серьезным финансовым последствиям.

В период глобальных изменений в АС, таких как, например, переход на отечественную операционную систему Astra Linux, система уязвима и многие ее модули защиты могут находиться в нерабочем состоянии. В данный период вероятность утечек ПДн крайне высока, и, чтобы этого избежать, стоит использовать временные решения в период внедрения полноценной системы защиты, если ускорить данный процесс не получается. Такие решения характеризуются быстротой развертывания и малым потреблением ресурсов, но в то же время охватывают лишь основные направления защиты.

В данной работе будет предложен программный модуль на основе DLP-технологий, оптимизированный под систему Astra Linux, предотвращающий утечки ПДн, повышающий эффективность работы администратора безопасности и обеспечивающий централизованное управление безопасностью.

Выбор метода распознавания конфиденциальной информации в тексте

Программный алгоритм, отвечающий за обнаружение и категоризацию защищаемой информации, является основой функционирования любой DLP-системы. Данные алгоритмы чаще всего базируются на одной из двух групп методов анализа: лингвистических или статистических.

Лингвистические методы анализа текста основаны на знаниях о языке и его структуре. Эти методы учитывают языковые особенности и специфику текста, благодаря чему они могут распознавать термины и другие элементы, характерные для конкретного языка. Лингвистический анализ хоть и позволяет более точно выявлять конфиденциальную информацию благодаря распознаванию контекста и смысловой нагрузки, но в то же время требует больших объемов языковых данных для обучения моделей и правил. Другой недостаток

связан со сложностью четкой категоризации при использовании вероятностного подхода в подобных методах, из-за чего снижается точность срабатывания системы.

Статистические методы анализа, в свою очередь, демонстрируют точность, близкую к абсолютной. По сути, статистический анализ использует вероятностные методы для идентификации конфиденциальной информации на основе ее статистических характеристик. То есть статистические технологии относятся к текстам не как к связанной последовательности слов, а как к произвольной последовательности символов, поэтому одинаково хорошо работают с текстами на любых языках. Данные методы зачастую основаны на анализе частоты встречаемости определенных слов или выражений в тексте, определении закономерностей и выявлении отклонений от них, которые могут указывать на конфиденциальность информации. Хотя статистический анализ не всегда учитывает контекст и семантику текста, он все же является более эффективным для обработки больших объемов данных — требует меньше ресурсов для обучения и работы системы, а также быстрее обрабатывает текст в реальном времени.

По причине вышеперечисленных преимуществ было решено опираться на статистический анализ и использовать метод машинного обучения. Машинное обучение в контексте DLP используется для создания моделей, обнаруживающих конфиденциальную информацию на основе статистических закономерностей. Эти модели обучаются на больших объемах текстовых данных, по которым автоматически выявляют сигнатуры конфиденциальной информации. Также важным достоинством моделей машинного обучения является их доступность для улучшения в будущем благодаря модификации и добавлению новых данных, используемых для обучения.

Создание модели машинного обучения для обнаружения конфиденциальной информации в тексте

Первым этапом реализации машинного обучения является сбор данных, которые будут

использованы для обучения модели. Поскольку модель будет служить для распознавания конфиденциальной информации в тексте, то требуются данные, содержащие как конфиденциальную, так и неконфиденциальную информацию.

Объем текста отдельного примера влияет на качество обучения модели, поэтому стоит подбирать оптимальный объем текста для примеров — от нескольких десятков до двух сотен слов.

Но гораздо сильнее на качество обучения модели влияет количество самих примеров информации. Для модели будет использовано около тысячи обучающих примеров — такое количество подойдет для начального прототипа и хорошо сочетается с простыми алгоритмами, используемыми далее.

Каждый пример был помечен на предмет содержания им конфиденциальной информации и сохранен в табличном файле.

После сбора обучающие данные нужно обработать. Этот процесс включает несколько этапов:

- приведение текста к нижнему регистру;
- удаление ненужных символов, таких как знаки препинания;
- токенизация — разделение текста на слова;
- удаление стоп-слов — часто встречающихся, но малоинформативных слов;
- лемматизация — приведение слов к их леммам, базовым словарным формам.

Однако даже обработанные тексты все еще являются слишком сложными и многомерными для алгоритмов машинного обучения и не могут быть обработаны ими напрямую. Именно для этого перед обучением самой модели нужно извлечь из текстов признаки, то есть преобразовать их в числовые векторы, доступные для использования алгоритмами машинного обучения.

Для этой задачи был выбран метод TF-IDF. Данный метод проводит оценку важности отдельного слова в тексте относительно всего набора текстов, выделяет ключевые слова и определяет, какие из них наиболее точно характеризуют определенные тексты из набора. Если слово часто встречается в конкретном тексте, но не во многих, значит, оно наиболее точно описывает его содержание.

Последним шагом является выбор непосредственно алгоритма машинного обучения. Для разработки модуля был выбран метод опорных векторов (SVM).

Данный алгоритм является достаточно универсальным и используется для задач классификации. Метод опорных векторов относится к задаче обучения, то есть он основан на том, что имеется некоторая выборка значений, отношения которых к определенным классам заранее известны. Такой выборкой и являются обучающие примеры [2].

Работа SVM продемонстрирована рис. 1, а. Пусть обучающие примеры представляют собой набор точек на плоскости и разбиты на заранее известные два класса: конфиденциальная и неконфиденциальная информация. Задача метода заключается в построении границы между этими двумя множествами. Причем все новые значения будут классифицироваться в зависимости от положения относительно этой границы: выше прямой — информация конфиденциальна, ниже — неконфиденциальна. Однако этот пример достаточно тривиален. В реальных задачах речь идет о пространствах высоких размерностей и граница в виде прямой не сможет в них существовать, так же как и значения в виде точек, которые в реальных вычислениях являются числовыми векторами. Поэтому в методе SVM границей является не просто прямая, а гиперплоскость — подпространство с размерностью, на единицу меньшей, чем исходное пространство.

Однако прямых, разделяющих два множества точек, может быть много. Возникает вопрос: какая же именно должна являться границей? В методе опорных векторов выбирается граница, расстояние от которой до каждого класса максимально. Такая граница называется оптимальной разделяющей гиперплоскостью. Чем больше расстояние между гиперплоскостью и ближайшими векторами обоих классов, тем лучше будет обобщающая способность у обучаемой модели. А сами ближайшие к границе векторы каждого класса, расстояние от которых до границы является величиной зазора между классами и разделяющей гиперплоскостью, называются опорными векторами.

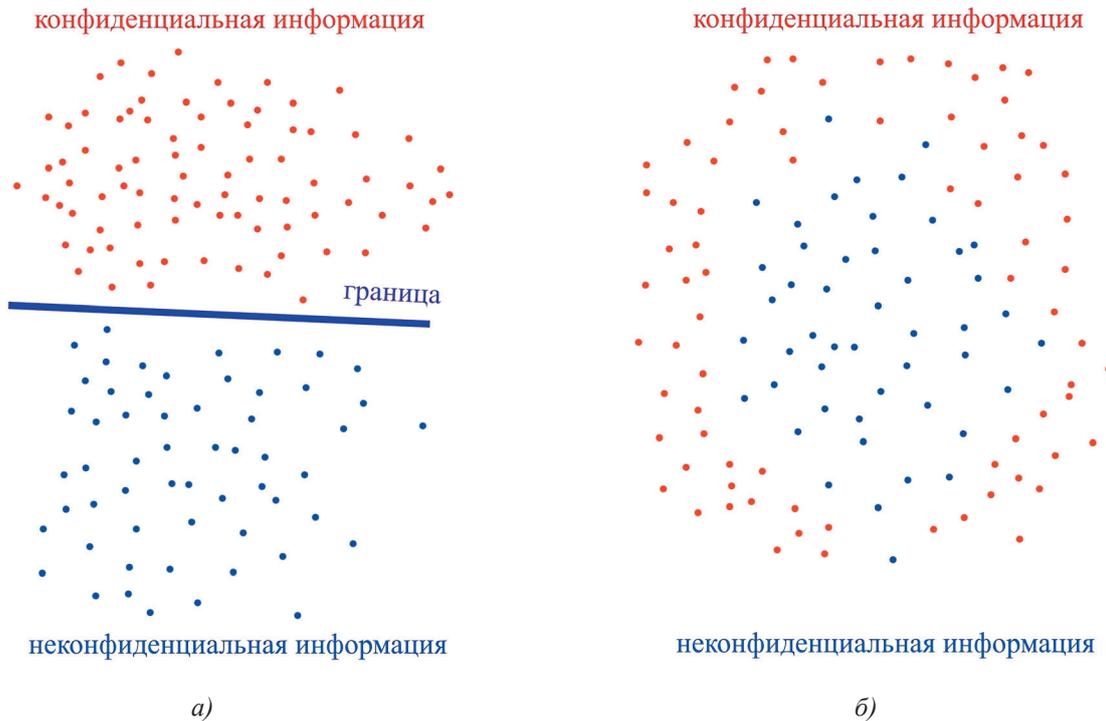


Рис. 1. Работа метода опорных векторов

(а — линейная разделимость данных, б — линейная неразделимость данных)

Таким образом, основная задача метода SVM заключается в построении оптимальной разделяющей гиперплоскости. Однако в реальных задачах, особенно таких сложных, как распознавание конфиденциальной информации, данные очень редко бывают линейно разделимыми (рис. 1, б). Поэтому линейное ядро метода SVM хоть и можно использовать для простого и наглядного тестирования ввиду его большой скорости, но для реальных задач оно покажет себя не лучшим образом. Для таких задач использование нелинейных ядер поможет значительно улучшить результаты обучения модели. Среди таких ядер стоит выделить радиальную базисную функцию RBF. Ядро RBF обеспечивает лучшую производительность за счет своей гибкости и способности работать с нелинейными границами, хотя обучение модели с таким ядром занимает несколько больше времени.

Распознавание и блокирование утечек конфиденциальной информации

Требуется определить, по каким каналам могут возникать утечки конфиденциальной информации. В первую очередь стоит обратить внимание

на сетевые каналы утечки. Около половины всех утечек конфиденциальной информации за последние годы происходит именно с использованием сетевых технологий, таких как интернет, локальные сети, электронная почта, облачные сервисы и прочее. На втором месте среди каналов утечек, которые можно контролировать на программном уровне, находятся утечки конфиденциальной информации, связанные с носителями данных [3]. К ним относятся кража, потеря или намеренная передача устройств и съемных носителей в руки злоумышленников.

Для того чтобы блокировать утечки конфиденциальной информации по Сети, потребуется анализировать сетевой трафик на рабочем месте. Будет осуществляться проверка на то, содержит ли анализируемый пакет слои TCP и IP, то есть является ли IP-пакетом с TCP-содержимым. Если пакет подходит под требования, то из него будет извлекаться полезная нагрузка в переменную строки. Анализ полезной нагрузки на конфиденциальность будет производиться с использованием TF-IDF векторизатора и SVM-модели. После определения конфиденциальности информации в пакете

нужно блокировать его передачу. Для этого будет использоваться подсистема ядра Linux NFQUEUE, позволяющая перехватывать сетевые пакеты и передавать их в пользовательское пространство для будущей обработки.

Для блокирования утечек конфиденциальной информации с помощью съемных носителей потребуется отслеживать все записываемые на носители файлы, анализировать их содержимое и в случае обнаружения конфиденциальных данных блокировать запись, производя откат изменений или удаление файла. Данное решение программа будет принимать в зависимости от ответа администратора на запрос о содержании конфиденциальной информации на носителе.

Однако данная программа все еще представляет собой единый скрипт, выполняемый локально на одном устройстве. Одной из главных целей разработки данного модуля является повышение эффективности работы администратора безопасности и централизованное управление безопасностью. Для достижения этой цели модуль необходимо разделить на клиентскую и серверную части.

Серверная часть будет располагаться на АРМ администратора безопасности. На ней будут находиться файлы модели машинного обучения и ее векторизатора, а также скрипт для повторного обучения модели в случае обновления ее базы данных. Данная часть программы будет отвечать за обработку сообщений и запросов на удаление конфиденциальной информации от клиентских частей.

Клиентская часть располагается на АРМ сотрудника и реализует основные функции по мониторингу сетевого трафика и информации на съемных носителях. Скрипт начинает свою работу одновременно с запуском системы — запрашивает у серверной части файлы SVM-модели и векторизатора, после чего запускает мониторинг сетевого трафика и цикл анализа файлов в каталогах съемных носителей. В процессе работы клиентская часть модуля обращается к серверной с запросами на удаление или откат изменений файлов на съемных носителях, передает сообщения о перехваченных и заблокированных пакетах, а также о завершении работы модуля.

Дополнительные механизмы безопасности модуля

Чтобы модуль в случае отказа сервера или соединения мог продолжать выполнять свои основные задачи, функция отправки серверу сообщения об отброшенном пакете в случае разрыва соединения будет сохранять сообщение в специальный локальный файл. После восстановления соединения все содержимое файла будет отправлено на сервер, после чего файл будет очищен. Функция отправки запроса на удаление или откат изменения файла на съемном носителе в случае разрыва соединения будет автоматически принимать решения, будто запрос одобрен администратором. Это усложняет работу сотрудника на АРМ, но зато наверняка предотвращает утечки, блокируя любую передачу подозрительной информации на съемный носитель, пока сетевое соединение не будет восстановлено. Таким образом, модуль сможет выполнять свои основные функции даже без соединения с сервером.

На данном этапе проектирования сообщения и запросы, передаваемые на серверную часть к администратору от клиентских модулей программы, представляют собой просто текстовые сообщения, выводимые на экран АРМ администратора. Соответственно, после перезапуска программы, например на следующий рабочий день, все предыдущие сообщения и запросы будут отсутствовать. Для DLP-технологий возможность ведения аудита крайне важна для расследований и анализа произошедших или предотвращенных утечек данных и прочих инцидентов безопасности. Поэтому дополним разрабатываемый модуль функцией журналирования событий, которая будет сохранять все сообщения и запросы от клиентов в базе данных (БД) на сервере.

Разработка будет направлена на библиотеки и инструменты для системы управления базами данных (СУБД) PostgreSQL, так как на ней базируются основные отечественные СУБД [4]. В БД будут использоваться две основные таблицы: для сообщений сетевого мониторинга и для запросов администратору. Каждая таблица имеет столбец с первичным ключом в виде автоматически увеличивающегося идентификатора, столбец

с идентификатором клиентского устройства, столбец с временем произошедшего события и столбец с текстом сообщения или запроса. Помимо этого, таблица запросов имеет столбец, содержащий ответ администратора на запрос, — разрешение или запрет на действие клиентского модуля.

Сообщения от клиентских частей включают данные о пакетах или файлах, вероятно содержащих конфиденциальные данные. На данном этапе сообщения передаются практически в открытом виде — в виде байтовой последовательности, при перехвате которой можно восстановить из нее изначальный объект. Для обеспечения безопасности необходимо шифровать сообщения перед передачей.

Шифрование будет производиться на основе криптографии на эллиптических кривых ECC. Безопасность шифрования ECC обусловлена сложностью решения задачи дискретного логарифмирования на основе эллиптических кривых [5]. Для обмена ключами будем использовать протокол Диффи — Хеллмана, основанный на криптографии эллиптических кривых. Обе стороны генерируют пары ключей, где закрытый ключ — это случайное целое число, а открытый ключ — произведение закрытого ключа на базовую точку на эллиптической кривой, являющуюся общей для обеих сторон. По открытому каналу стороны обмениваются открытыми ключами. Каждая сторона, используя полученный открытый ключ, вычисляет из него общий секрет — произведение открытого ключа другой стороны и своего закрытого ключа. Общий секрет получается одинаковым у обеих сторон, несмотря на разные множители. На основе общего секрета уже можно сгенерировать криптографический ключ нужной длины и использовать его для шифрования сообщений. Сделать это можно с помощью функции HKDF, которая работает в два этапа: сначала создает из общего секрета псевдослучайный ключ, а после на его основе создает конечный ключ нужной длины.

Для самого шифрования будем использовать симметричный алгоритм блочного шифрования AES. Данный алгоритм блочного шифрования разбивает исходный текст на блоки по 16 байт и, пред-

ставляя их в виде квадратной матрицы текста, шифрует каждый по очереди. Шифрование включает замену исходных байтов на другие по специальной таблице S-box, циклический сдвиг строк матрицы текста влево, умножение каждого столбца матрицы текста на соответствующий столбец матрицы байтов одного блока, поэлементное добавление к матрице текста раундовых ключей [6]. А дешифрование представляет собой последовательность инвертированных операций шифрования, выполняемых в обратном порядке. AES будет реализован в режиме обратной связи по зашифрованному тексту. В этом варианте перед шифрованием каждого следующего блока текста он складывается по модулю два с зашифрованным результатом шифрования предыдущего блока. Предыдущим блоком для первого блока данных считается вектор инициализации — случайно сгенерированный блок данных, и благодаря этому шифрование одного и того же текста по тому же ключу каждый раз будет давать уникальный результат, поскольку вектор инициализации генерируется снова для каждой операции шифрования.

Разработка программного модуля

Программа была написана на языке Python, поскольку операционная система Astra Linux, на работу с которой она направлена, полностью поддерживает его. Использовались как стандартные механизмы языка, так и сторонние библиотеки, такие как re, NLTK, pymorphy2, sklearn, Scapy, netfilterqueue, Watchdog, socket, pickle, cryptography.

Разработанное клиент-серверное приложение прошло тестирование на испытательном стенде, представляющем собой две виртуальные машины на ОС Astra Linux Common Edition 2.12 с установленным между ними локальным соединением. На виртуальной машине администратора также дополнительно была установлена СУБД PostgreSQL Pro Standard 11 и создана БД с двумя таблицами. После определения переменных окружения с адресом сервера, клиента и БД были развернуты модули. Клиентский модуль был установлен на автозапуск с системой и правами суперпользователя с помощью файла systemd.

По результатам тестирования можно сделать вывод, что разработанный программный модуль полностью выполняет требуемые функции по блокированию утечек конфиденциальной информации и централизованному контролю работы со стороны администратора безопасности.

Заключение

В результате разработки был создан программный модуль для защиты конфиденциальной информации от утечек с АРМ сотрудников. Данный

модуль представляет собой клиент-серверное приложение с ведением аудита и криптографической защитой. Данный модуль может быть использован в организациях с распределенными системами для повышения уровня централизации управления безопасностью и блокирования утечек в период перевода организаций на иные серийные средства защиты. Модуль является базовой версией, которая может дополняться со временем и по требованиям конкретной организации к фильтруемой информации и структуре программы.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Утечки информации ограниченного доступа в мире и России, первое полугодие 2023 года. Аналитический отчет // Экспертно-аналитический центр InfoWatch, 2023. 17 с. URL: <http://www.infowatch.ru/sites/default/files/analytics/files/utechki-informatsii-ogranichennogo-dostupa-v-mire-i-rossii-za-pervoe-polugodie-2023-goda.pdf> (дата обращения: 01.04.2024)
2. Полетаева Н. Г. Классификация систем машинного обучения // Вестник Балтийского федерального университета им. И. Канта. Серия: Физико-математические и технические науки. 2020. № 1. С. 5–22.
3. Утечки информации // Anti-Malware.ru. URL: <http://www.anti-malware.ru/threats/leaks> (дата обращения: 01.04.2024).
4. Обзор отечественных СУБД // 1С-MSSoft.Ru. 2023. 4 декабря. URL: <http://www.mssoft.ru/news/40506> (дата обращения: 09.04.2024).
5. Обухов В. А. Криптография на основе эллиптических кривых (ЕСС) // Потомки Аль-Фаргани. 2023. Т. 1, вып. 4. С. 182–188. DOI: 10.5281/zenodo.10337673
6. Батыркаев А. Г., Глотина И. М. Что такое шифрование AES и как оно работает // Агротехнологии XXI века: стратегия развития, технологии и инновации: материалы Всероссийской научно-практической конференции (Пермь, 16–18 ноября 2021 года). Пермь: Прокрость, 2021. С. 254–258.

Дата поступления: 24.08.2024

Решение о публикации: 20.09.2024

Development of a DLP Module to Protect Personal Data in a Distributed Automated System

Arkady A. Kovtunenکو — information security specialist, graduate of the Department “Informatics and Information Security”. Research interests: software development, information security. E-mail: bernadod2000@mail.ru

Emperor Alexander I Petersburg State Transport University, 9, Moskovsky pr., Saint Petersburg, 190031, Russia

For citation: Kovtunenکو A. A. Development of a DLP module to protect personal data in a distributed automated system // Intellectual Technologies on Transport. 2024. No. 3 (39). P. 65–72. DOI:10.20295/2413-2527-2024-339-65-72 (In Russian)

Abstract. *The importance of personal data protection in automated systems is being considered. A study is provided on the development of a software module for protecting confidential information, including personal data, in a distributed automated system. **The primary goal of the study:** to create is to create software that prevents leaks of confidential information. The developed software module is designed to block the leakage of confidential information through network channels and removable media using a machine learning model with centralized security administrator management. **The practical significance:** the use of this module in organizations includes ensuring centralized information security management, reducing the risk of confidential data leaks, and supporting the investigation of security incidents. The research can be used as a temporary measure during the period of updating and adapting the system to new realities until a full-fledged protection system is achieved.*

Keywords: *personal data, information security management system, confidential information leaks, machine learning, DLP (Data Leaks Protection)*

REFERENCES

1. Utechki informacii ogranichenogo dostupa v mire i Rossii, pervoe polugodie 2023 goda. Analiticheskij otchet // Ekspertno-analiticheskij centr InfoWatch, 2023. 17 c. URL: <http://www.infowatch.ru/sites/default/files/analytics/files/utechki-informatsii-ogranichenogo-dostupa-v-mire-i-rossii-za-pervoe-polugodie-2023-goda.pdf> (data obrashcheniya: 01.04.2024). (In Russian)
2. Poletaeva N. G. Klassifikaciya sistem mashinnogo obucheniya // Vestnik Baltijskogo federal'nogo universiteta im. I. Kanta. Seriya: Fiziko-matematicheskie i tekhnicheskie nauki. 2020. № 1. S. 5–22. (In Russian)
3. Utechki informacii // Anti-Malware.ru. URL: <http://www.anti-malware.ru/threats/leaks> (data obrashcheniya: 01.04.2024). (In Russian)
4. Obzor otechestvennyh SUBD // 1C-MSSoft.Ru. 2023. 4 dekabrya. URL: <http://www.mssoft.ru/news/40506> (data obrashcheniya: 09.04.2024). (In Russian)
5. Obuhov V. A. Kriptografiya na osnove ellipticheskikh krivyh (ECC) // Potomki Al'-Fargani. 2023. T. 1, vyp. 4. S. 182–188. DOI: 10.5281/zenodo.10337673 (In Russian)
6. Batyrkaev A. G., Glotina I. M. CHto takoe shifrovanie AES i kak ono rabotaet // Agrotekhnologii XXI veka: strategiya razvitiya, tekhnologii i innovacii: materialy Vserossijskoj nauchno-prakticheskoy konferencii (Perm', 16–18 noyabrya 2021 goda). Perm': Prokrost", 2021. S. 254–258. (In Russian)

Received: 24.08.2024

Accepted: 20.09.2024

Показатели и алгоритмы оценивания качества результатов идентификации состояния прилегающих территорий

Зеленцов Вячеслав Алексеевич¹ — докт. техн. наук, профессор, главный научный сотрудник лаборатории информационных технологий в системном анализе и моделировании Санкт-Петербургского федерального исследовательского центра РАН. Область научных интересов: системный анализ, методы и информационные технологии поддержки принятия решений, космические информационные технологии, методы обработки аэрокосмической информации в интегрированных системах наземно-космического мониторинга. E-mail: v.a.zelentsov@gmail.com

Мочалов Виктор Федорович² — старший научный сотрудник Военного института (научно-исследовательского) Военно-космической академии имени А. Ф. Можайского. Область научных интересов: обработка и анализ материалов мультиспектральной и гиперспектральной аэрокосмической съемки, в том числе высокого пространственного разрешения для определения параметров растительности, устойчивости экосистем с учетом данных наземных измерений

¹ Санкт-Петербургский федеральный исследовательский центр РАН, Россия, 199178, Санкт-Петербург, 14-я линия В.О., 39

² Военно-космическая академия имени А. Ф. Можайского, Россия, 197198, Санкт-Петербург, ул. Ждановская, 13

Для цитирования: Зеленцов В. А., Мочалов В. Ф. Показатели и алгоритмы оценивания качества результатов идентификации состояния прилегающих территорий // Интеллектуальные технологии на транспорте. 2024. № 3 (39). С. 73–82. DOI: 10.20295/2413-2527-2024-339-73-82

Аннотация. Рассматриваются вопросы применения методов машинного обучения для оценивания состояния территорий в непосредственной близости к инфраструктурным объектам железной дороги. Представлены основные источники исходных данных, а также алгоритмы оценивания качества результатов автоматизированной идентификации элементов ландшафта на примере лесной растительности. Идентификация состояния лесной растительности осуществляется на базе спектрально-яркостных признаков, определяемых с помощью материалов мультиспектральной аэрокосмической съемки. Предложены показатели качества результатов автоматизированной обработки материалов съемки и алгоритмы расчета рассмотренных показателей. Приведены примеры оценивания качества результатов применения методов машинного обучения.

Ключевые слова: мультиспектральная аэрокосмическая съемка, спектрально-яркостные признаки, методы машинного обучения, качество идентификации, показатели и алгоритмы оценивания

Благодарности. Исследования выполнены в рамках бюджетной темы FFZF-2022-0004. При проведении разработок использованы ресурсы Центра коллективного пользования «Северо-Западный центр мониторинга и прогнозирования развития территорий». URL: <https://ckp-rf.ru/catalog/ckp/2079622/>

Введение

Идентификация состояния лесных насаждений вблизи объектов железнодорожного транспорта с точки зрения пожарной безопасности является важной практической задачей. Требуется своевременно реагировать на факты нарушений порядка содержания лесных массивов, особенно в пожароопасный период. Для этого в лесных массивах необходимо выявлять, в частности, следующие нарушения, способствующие самопроизвольному возникновению и распространению пожаров: несвоевременный вывоз мусора, порубочных остатков; возникновение заболеваний лесных насаждений, приводящих к усыханию растительности; несвоевременное устранение последствий ветровала. Лесные насаждения являются одним из наиболее распространенных элементов ландшафта на прилегающих территориях. Предложенные показатели качества могут характеризовать также качество идентификации состояния водных объектов, полей, горных массивов, других природных и антропогенных объектов. Представление достоверных данных о выявленных нарушениях можно обеспечить в ходе проведения наземных обследований, но эти работы, как правило, сопровождаются большими материальными и временными затратами. Одним из способов, позволяющих сократить влияние указанных выше недостатков, является проведение мультиспектральной аэрокосмической съемки [1] и ее углубленная автоматизированная обработка с помощью интеллектуально насыщенных алгоритмов.

Современные алгоритмы обработки материалов съемки часто основаны на методах машинного обучения [2–5]. При идентификации состояния прилегающих территорий обработка материалов съемки предполагает решение задачи так называемой семантической сегментации (semantic segmentation) элементов ландшафта в пределах рассматриваемой сцены. При решении такой задачи необходимо заранее иметь углубленное семантическое описание известных элементов ландшафта. Для них на основе материалов съемки выполняется построение обучающих и контрольных выборок. Для железнодорожных объ-

ектов должен быть составлен перечень типовых нарушений, оказывающих влияние на безопасность железной дороги. Географически привязанные участки растительного покрова должны быть зарегистрированы и снабжены семантическим описанием. На каждом участке выполняются измерения спектральных отражательных характеристик элементов ландшафта с учетом сезонной изменчивости. То есть должна быть создана специализированная система исходных данных.

В дальнейшем можно будет применять методы обработки материалов съемки, основанные на учете спектрально-яркостных признаков идентификации элементов ландшафта [1]. Спектрально-яркостные признаки опираются на результаты измерений отражательных характеристик растительности в нескольких спектральных каналах видимого и ближнего инфракрасного диапазонов спектра. По результатам проведения атмосферной коррекции осуществляется определение абсолютных физических величин — коэффициентов спектральной яркости. Знание абсолютных физических величин обеспечивает возможность обработки материалов съемки от мультиспектральной аппаратуры, размещенной на борту различных космических аппаратов и авиационных носителей, включая беспилотные летательные аппараты. При этом также могут рассчитываться и учитываться в ходе обработки различные специализированные индексы, характеризующие состояние растительности в условиях известных антропогенных воздействий [1]. Сложности формирования исходных данных и дальнейшее применение методов автоматизированной обработки подтверждают необходимость оценивания качества результатов обработки.

Одной из типовых задач обработки результатов дистанционных измерений является отнесение каждого пикселя изображения анализируемого участка территории к типу поверхности, обладающему заранее известными свойствами. Глубина семантического описания определяется содержанием практической задачи и может представлять собой, например, описание видового состава растительности, возраст и другие количественные

характеристики, имеющие значение для безопасности инфраструктурных объектов железной дороги. Кроме того, должны учитываться и отражаться в семантическом описании различные виды антропогенного воздействия на элементы ландшафта, в частности, последствия пожаров, ветровалов, загрязнения бытовыми и промышленными отходами, нефтепродуктами и др.

Оценивание качества обработки материалов съемки может выполняться на основе различных показателей, и от их выбора зачастую зависят выводы о достоверности полученных результатов. Для демонстрации роли показателей оценивания качества результатов обработки материалов съемки рассмотрим пример семантической сегментации трех типов поверхностей после возникновения нарушения в состоянии лесного массива вблизи одного из условных объектов железной дороги [6]. В качестве исходных данных приняты результаты измерения спектральных отражательных характеристик известных элементов ландшафта. Измерения выполнены с помощью мультиспектральной аппаратуры, размещенной на борту космического аппарата Sentinel-2.

Проведен анализ изменения состояния растительности и спектральных отражательных характе-

ристик элементов ландшафта до неблагоприятного воздействия на основе материалов съемки от 17 августа 2019 года и после воздействия на основе материалов съемки от 2 июня 2020 года. Семантическое описание территории выполнено в ходе проведения наземных обследований. Изображения фрагмента сцены в псевдоцветах приведены на рис. 1.

Семантическое описание содержало данные о трех основных элементах ландшафта с условными метками класса 1, 2 и 3. С помощью семи методов машинного обучения осуществлена семантическая сегментация указанных элементов ландшафта. Верификация результатов идентификации выполнялась в ходе наземных обследований в выбранных заранее контрольных точках. Отдельные точки контроля отмечены на рис. 1 в виде значков пикселей.

При идентификации нарушения в состоянии лесной растительности получены результаты, приведенные в табл. 1. Размер контрольной выборки составлял 102 пикселя. Было идентифицировано три типа поверхностей, обозначенных соответствующими метками класса. Номера меток отражают состояние территорий с высокой, низкой степенью нарушения и ненарушенные территории соответственно.

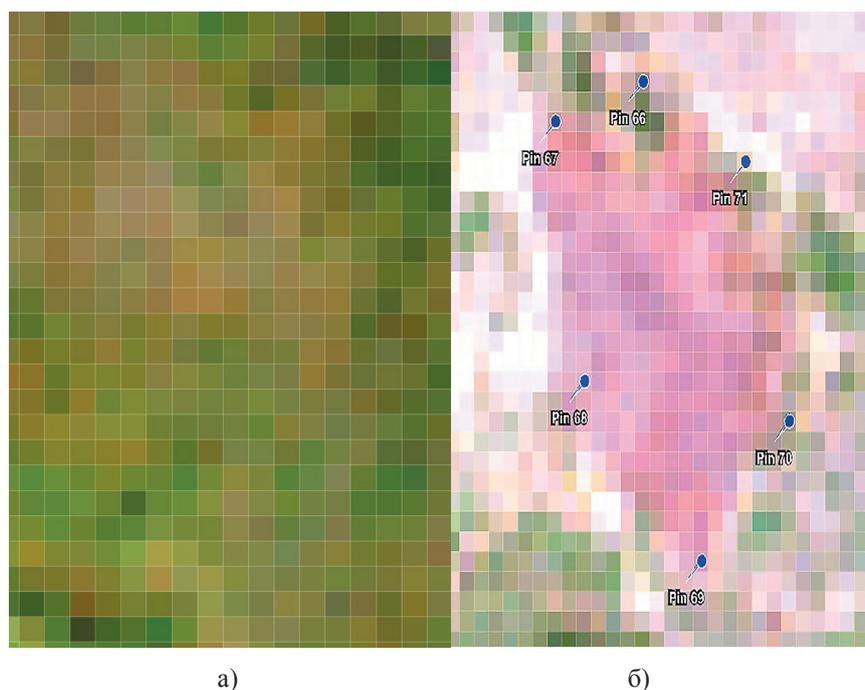


Рис. 1. Фрагменты сцены: а) до возникновения нарушения; б) после возникновения нарушения

Таблица 1

Результаты идентификации элементов ландшафта и данные о качестве идентификации

№ пикселя	Элемент ландшафта	Алгоритмы машинного обучения						
		<i>PPN</i>	<i>SVM</i>	<i>Tree</i>	<i>LR</i>	<i>Forest</i>	<i>KNN</i>	<i>Ensemble</i>
1	2	2	2	2	2	2	2	2
2	1	1	1	1	1	1	1	1
3	2	2	2	2	2	2	2	2
...								
100	1	1	1	1	1	1	1	1
101	1	1	1	1	1	1	1	1
102	3	3	3	3	3	3	3	3
Accuracy		1,0	1,0	0,97	1,0	0,98	0,98	1,0

Методы обработки материалов съемки или методы машинного обучения в табл. 1 имеют следующие обозначения:

- *PPN* — метод коррекции ошибки;
- *SVM* (support vector machine) — метод опорных векторов;
- *Tree* — дерево принятия решений;
- *LR* — метод логистической регрессии;
- *Forest* — метод случайного леса (random forest);
- *KNN* — метрический алгоритм k -ближайших соседей (k -nearest neighbors).

Комплексное сочетание алгоритмов (*Ensemble*) реализовано путем анализа результатов всех приведенных ранее алгоритмов машинного обучения. Для оценивания качества результатов обработки использовался традиционный обобщенный показатель — *Accuracy*. Значение показателя определяется путем анализа соотношений количества пикселей для верно идентифицированных элементов ландшафта к общему числу пикселей в пределах контрольной выборки для анализируемого элемента ландшафта.

Показатель *Accuracy* в таблице принимает значение равное единице или близкое к ней. Причиной такой ситуации может быть недостаточный размер контрольной выборки или низкая информативность привлекаемого показателя оценивания качества обработки.

Приведенный пример позволяет сделать вывод о том, что применяемый показатель (*Accuracy*) для

оценивания качества семантической сегментации элементов ландшафта носит неинформативный характер. Требуется обеспечение обоснованного выбора показателей качества тематической обработки, учитывающих семантическое описание идентифицируемых поверхностей, и разработка соответствующих алгоритмов их расчета.

Целью описываемых исследований является разработка предложений по возможному составу показателей и алгоритмов для оценивания качества обработки материалов съемки, основанных на анализе особенностей конкретных задач тематической обработки.

Основная часть

Источники исходных данных. В качестве источников исходных данных выступают материалы мультиспектральной космической съемки. Основные технические характеристики бортового съемочного оборудования приведены в табл. 2. Спектральная чувствительность оборудования и условное пространственное разрешение в наиболее информативных каналах сопоставимы.

Показатели и способы оценивания качества результатов обработки материалов съемки. При выборе показателей качества идентификации состояния элементов ландшафта предлагается предварительно провести анализ смыслового

Основные характеристики бортового съемочного оборудования

Принадлежность	Европейское космическое агентство, ESA	США, NASA	Россия	Китай
Носитель	Sentinel-2	Landsat 9	Канопус-В	БПЛА
Наименование или обозначение оборудования	MSI	OLI	MCC	Parrot Sequoia
Количество спектральных каналов, шт.	12	8	4	4
Условное пространственное разрешение, м	10–30	20	12	–
Границы спектрального диапазона, нм	432,2–2289,9	411,3–2389,6	460–840	530–810
Ширина полосы захвата, км	185	290	19	–
Периодичность съемки, сутки	5	5	3	–

содержания решаемой практической задачи управления прилегающей территории.

Подобным вопросам посвящена, в частности, работа [7], в которой приведен перечень часто применяемых показателей и порядок их расчета. Традиционные технологии обработки предусматривают формирование обучающей и контрольной выборок. Если обучающая выборка применяется для настройки алгоритмов обработки, то контрольная — представляет собой совокупность достоверных данных из независимых источников и служит для оценивания результатов обработки.

Недостатками большинства работ являются: попытка специалистов, занимающихся тематической обработкой материалов съемки, найти универсальные, обобщенные показатели для решаемых задач и всех идентифицируемых элементов ландшафта; низкая концентрация внимания на основных видах элементов ландшафта, требующих повышенной точности их идентификации; не всегда учитывается сбалансированность исходных данных, характеристики относительных размеров идентифицируемых поверхностей.

Приведем примеры различных практических задач, от содержания которых зависит выбор того или иного показателя качества обработки.

Обязательное выделение на снимках элементов ландшафта, требующих повышенного приоритетного внимания. Например, при выявлении пожароопасных участков леса такими элементами яв-

ляются места высыхания растительности, участки с ветровалом, участки, загрязненные мазутом или другими нефтепродуктами.

В других часто решаемых задачах в ходе семантической сегментации различных видов поверхностей в пределах анализируемого контура требуется провести идентификацию двух или более видов лесной растительности, имеющих сопоставимое значение.

Очевидно, что качество обработки материалов съемки в этих двух задачах должно оцениваться по-разному. Однако и в том, и в другом случае основу оценивания составляют значения первичных показателей, отражающих количественные значения правильно и ошибочно идентифицированных пикселей изображения. На основе значений первичных показателей в дальнейшем вычисляются комплексные показатели.

Предлагаемый общий подход к выбору показателей для оценивания качества обработки состоит в следующем. Идентифицируемые элементы ландшафта разделяются на два класса элементов, обладающих условно положительными и отрицательными свойствами с точки зрения решаемой задачи. Например, в задаче выявления пожароопасных участков леса к элементам ландшафта с отрицательными свойствами относятся места высыхания растительности и участки с ветровалом. Тогда результаты идентификации можно представить в виде значений следующих первичных показателей качества обработки [3]:

TP (True Positive) — количество пикселей, представляющих территории с истинно неповрежденными участками растительности;

TN (True Negative) — количество пикселей, представляющих территории с истинно поврежденными участками растительности;

FP (False Positive) — количество пикселей, представляющих территории с ошибочно неповрежденными участками растительности;

FN (False Negative) — количество пикселей, представляющих территории с ошибочно поврежденными участками растительности.

Первичные показатели определяются на основе анализа данных контрольной выборки, которая, как и обучающая выборка, формируется на основе исходных данных – выбранных фрагментов материалов съемок с известным семантическим описанием. Но в отличие от обучающей выборки, данные контрольной выборки не используются в обучении модели автоматизированной обработки материалов съемки. Значения данных контрольной и обучающей выборок считаются достоверными.

Иллюстрация порядка определения значений первичных показателей качества идентификации двух поверхностей приведена на рис. 2.

На рис. 2 по оси абсцисс отложены условные пороговые численные значения, отражающие положение гиперпараметра, или разделяющей поверхности (Threshold), на идентифицируемые элементы ландшафта в рассматриваемом примере.

По оси ординат в виде гистограммы представлены значения количества пикселей, отражающих соответствующие элементы ландшафта. Условно положительными свойствами (P) обладает ряд 2. Условно отрицательными свойствами обладает ряд 1. Красной линией обозначено текущее расположение разделяющей поверхности. Слева от красной линии голубая поверхность, ограниченная голубым контуром, соответствует значению *TP*, а ограниченная зеленым контуром — *FP*. Справа от разделяющей поверхности зеленая фигура, ограниченная зеленым контуром, соответствует значению *TN*, а ограниченная голубым контуром — *FN*.

В зависимости от выбора порогового значения виртуальной плоскости определяется пространственное расположение красной линии и рассчитываются значения промежуточных показателей *TP*, *TN*, *FP*, *FN* путем суммирования числа пикселей в пределах выделенных контуров.

На основе приведенных первичных показателей определяются значения комплексных показателей, с помощью которых оценивается качество обработки [3]:

Полнота, чувствительность (Recall):

$$REC = \frac{TP}{TP + FN} \quad (1)$$

отражает долю элементов ландшафта, которые характеризуются условно положительными свойствами. С помощью показателя оценивается способность привлекаемого алгоритма

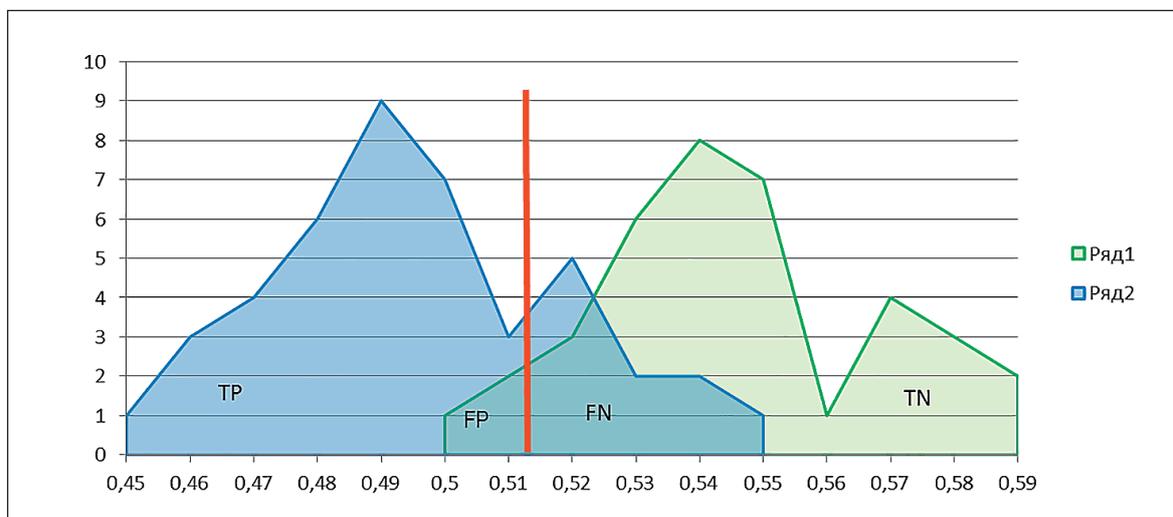


Рис. 2. Представление значений первичных показателей качества обработки

идентифицировать искомый, важный для потребителя вид ландшафта. Значение REC, рассчитанное по формуле (1), приближается к единице, когда количество пикселей с ошибочно отрицательными свойствами стремится к нулю. Таким образом показатель может свидетельствовать о вероятности успешной идентификации элементов ландшафта с положительными свойствами. На основе анализа значения REC можно принять решение о необходимости проведения дополнительных обследований или привлечения дополнительных алгоритмов обработки.

Точность (Precision):

$$PRE = \frac{TP}{TP + FP} \tag{2}$$

Точность — это отношение между истинно положительными и всеми положительными результатами идентификации. Значения REC и PRE характеризуют различные, конкурирующие свойства алгоритмов автоматизированной обработки. Поэтому целесообразно проводить их совместный анализ.

На рис. 3 приведен график изменения значений показателя PRE, рассчитанного по формуле (2), в зависимости от значений показателя REC, рассчитанного по формуле (1). Изменения отражают движение разделяющей поверхности, или красной линии, приведенной на рис. 2 от значения 0,5 до 0,55.

По оси абсцисс представлены значения показателя REC, по оси ординат — PRE, которые соответствуют числовым значениям, приведенным в табл. 3.

Таблица 3

Значения комплексных показателей качества обработки в зависимости от порогового значения

№	Порог	REC	PRE	OA
1	0,50	0,65	0,97	0,80
2	0,51	0,78	0,92	0,85
3	0,52	0,89	0,87	0,87
4	0,53	0,93	0,78	0,83
5	0,54	0,99	0,69	0,76
6	0,55	1,00	0,63	0,69

В табл. 3 приведен пример расчета значений комплексных показателей с учетом значений первичных показателей, приведенных на рис 2. При этом в столбце «Порог» отражаются значения пересечения красной линии с осью абсцисс.

В четвертом столбце табл. 3 представлены значения обобщенного показателя OA (Overall Accuracy), характеризующего долю правильно идентифицированных пикселей:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Показатель OA описывает общую точность привлекаемого метода обработки. Этот показатель особенно полезно использовать, когда каждый класс идентифицируемых поверхностей одинаково важен. Он рассчитывается как отношение количества пикселей с правильно определенными свойствами элементов ландшафта к их общему количеству. На рис. 4 приведен график изменения показателя OA и числовых значений порога разделяющей поверхности в зависимости от номера точек отсчета.

Анализ данных, приведенных на рис. 4, позволяет сделать вывод о возможности оптимизации выбора порогового значения разделяющей поверхности для достижения максимального значения показателя OA, рассчитываемого по формуле (3).

На базе рассмотренных показателей может быть сформирован комплексный показатель как объединение показателей REC и PRE в виде гармонического среднего:

$$F1 = 2 \cdot \frac{PRE \cdot REC}{PRE + REC} \tag{4}$$

Мера F1, рассчитанная в формуле (4), учитывает показатели PRE и REC как равноценные, а если более важным является один из этих показателей, применяется мера [7]:

$$F_{\beta} = \frac{(1 + \beta^2) \cdot PRE \cdot REC}{\beta^2 \cdot PRE + REC} \tag{5}$$

Порядок выбора значения β в формуле (5) с учетом важности одного из показателей PRE или REC определяется неравенством, которое отражает тот факт, что для оценивания качества обработки в задаче, для которой приоритетным является

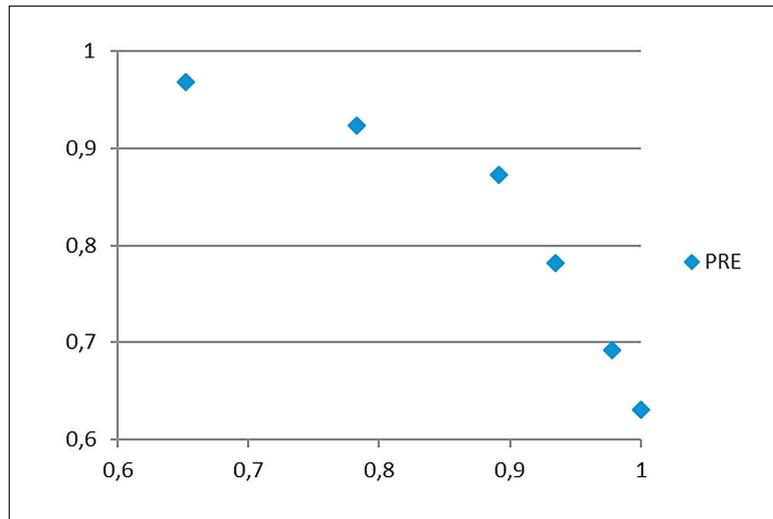


Рис. 3. Изменения значений показателя PRE в зависимости от REC

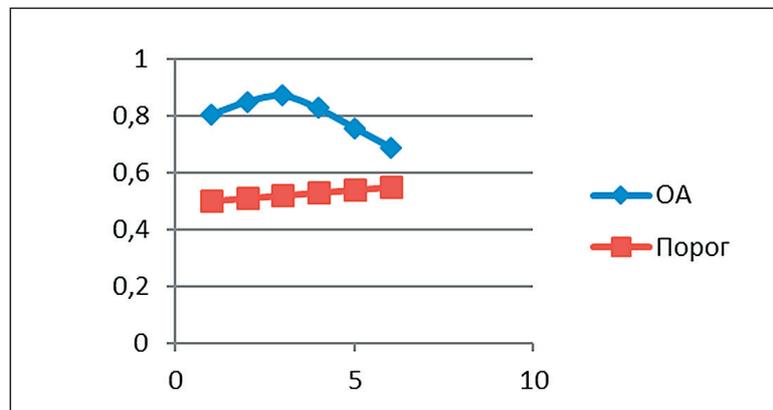


Рис. 4. Изменения значений показателя OA и порога разделяющей поверхности

достижение повышенного показателя PRE с заданным уровнем REC, выбирается значение $\beta < 1$.

$$\begin{cases} \beta < 1 \rightarrow PRE \\ \beta > 1 \rightarrow REC. \end{cases}$$

В табл. 4 приведены примеры выбора показателей для оценивания качества идентификации элементов ландшафта в зависимости от вида решаемой практической задачи.

Заключение

Представленные показатели для оценивания качества результатов обработки материалов съемки и порядок их расчета базируются на учете особенностей каждой конкретной задачи обработки данных. Показатели могут применяться как в случае применения традиционных алгоритмов, предусма-

тривающих расчет индексов, отражающих состояние элементов ландшафта, так и в случае применения алгоритмов машинного обучения, включая алгоритмы искусственных нейронных сетей.

Важным моментом при идентификации состояния прилегающих территорий является обоснованное формирование обучающих и контрольных выборок, обеспечение независимости этих выборок в алгоритмах машинного обучения и при оценивании качества выходных результатов.

Необходимо отметить, что, опираясь на предварительный анализ значений приведенных показателей качества, можно обосновывать требования к алгоритмам обработки в части оптимизации выбора значения гиперпараметра или порога разделяющей поверхности.

Примеры практических задач и показатели для оценивания их качества

Задача	Наименование задачи	Характеристика задачи, требования к качеству ее решения	Формула расчета качества идентификации
A	Идентификация двух видов леса: соснового, свойство (P) и лиственного, свойство (N)	Однородные исходные данные, требования к точности идентификации сопоставимы	$OA = \frac{TP + TN}{TP + TN + FP + FN}$
B	Идентификация участков леса, подверженных опасному заболеванию, свойство (N), на фоне здорового леса, свойство (P)	Преимущественное обязательное выявление элементов ландшафта, обладающих свойством (N)	$REC = \frac{TP}{TP + FN}$
C	Идентификация участков леса, характеризующихся повышенным возможным воздействием на безопасность эксплуатации объекта, свойство (P)	Выявление с максимальной достоверностью только тех элементов ландшафта, которые отличаются заданными свойствами (P)	$PRE = \frac{TP}{TP + FP}$
D	Идентификация двух видов леса: соснового, свойство (P) и лиственного, свойство (N)	Однородные исходные данные, требования по приоритетности идентификации поверхностей со свойствами (P) и (N) могут уточняться в зависимости от складывающейся ситуации	$F_{\beta} = \frac{(1 + \beta^2) \cdot PRE \cdot REC}{\beta^2 \cdot PRE + REC}$

Следует также подчеркнуть, что управление состоянием территорий вблизи железнодорожных объектов — это непрерывный плановый процесс, в котором в качестве регулярных операций должны быть предусмотрены систематическое выполнение съемочных работ, обработка материалов съемки, а также синхронные

наземные обследования на выбранных тестовых участках. Обязательным является выполнение этих операций при возникновении неблагоприятных воздействий на прилегающую территорию, а также по итогам реализации тех или иных мероприятий по управлению развитием территорий.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- Шовенгердт Р. А. Дистанционное зондирование. Методы и модели обработки изображений / пер. с англ. А. В. Кирюшина, А. И. Демьяникова, 3-е изд. М.: Техносфера, 2010. 560 с.
- Корабашев О. З. Анализ и перспективы применения методов машинного обучения для чрезвычайных ситуаций // Интеллектуальные технологии на транспорте. 2024. № 1 (37). С. 12–17. DOI: 10.20295/2413-2527-2024-137-12-17
- Рашка С. Python и машинное обучение / пер. с англ. А. В. Логунова. М.: ДМК Пресс, 2017. 418 с.
- Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Second Edition. Sebastopol (CA): O'Reilly Media, 2019. 848 p.
- Теребиж В. Ю. Введение в статистическую теорию обратных задач. М.: Физматлит, 2005. 376 с.
- Мочалов В. Ф., Хабаров Р. С. Обработка материалов мультиспектральной съемки на основе методов машинного обучения при управлении состоянием лесного массива // Сборник материалов IV Международной научной конференции по проблемам управления в технических системах (ПУТС-2021) (Санкт-Петербург, 21–23 сентября 2021). СПб.: ЛЭТИ, 2021. С. 269–272.
- Maxwell A. E., Warner T. A., Guillén L. A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studie. Part 1: Literature Review // Remote Sensing. 2021. Vol. 13, iss. 13. Art. 2450. 27 p. DOI: 10.3390/rs13132450

Дата поступления: 28.06.2024

Решение о публикации: 16.09.2024

Indicators and Algorithms for Assessing the Identification Results Quality of the Adjacent Territories State

Vyacheslav A. Zelentsov¹ — Dr. Sci. in Engineering, Professor, Chief Researcher at the Laboratory of Information Technologies in System Analysis and Modeling of the St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: system analysis, methods and information technologies of decision support, space information technologies, methods of processing aerospace information in integrated ground-space monitoring systems. E-mail: v.a.zelentsov@gmail.com

Viktor F. Mochalov² — Senior Researcher at the Military Research Institute. Research interests: processing and analysis of multispectral and hyperspectral aerospace survey materials, including high spatial resolution for determining vegetation parameters, ecosystem stability, taking into account ground-based measurement data.

¹ St. Petersburg Federal Research Center of the Russian Academy of Sciences, 14th line V.O., 39, St. Petersburg, 199178, Russia

² Mozhaisky Military Aerospace Academy, 13, Zhdanovskaya str., St. Petersburg, 197198, Russia

For citation: Zelentsov V. A., Mochalov V. F. Indicators and Algorithms for Assessing the Identification Results Quality of the Adjacent Territories State // Intellectual Technologies on Transport. 2024. № 3 (39). Pp. 73–82. DOI: 10.20295/2413-2527-2024-339-73-82 (In Russian)

Abstract. *Machine learning methods for identification of territories near railway infrastructure facilities are considered. The main sources of initial data are presented, as well as algorithms for assessing the quality of the results for landscape elements automated identification using the forest vegetation example. Forest vegetation state identification is carried out based on spectral-brightness characteristics determined using multispectral aerospace imaging materials. Quality indicators for the results of survey materials automated processing and algorithms for calculating the considered indicators are proposed. Examples are given to assess the results quality of using machine learning methods.*

Keywords: *multispectral remote sensing, spectral characteristics, machine learning methods, identification quality*

REFERENCES

1. Shovengerdt R. A. Distancionnoe zondirovanie. Metody i modeli obrabotki izobrazhenij / per. s angl. A. V. Kiryushina, A. I. Dem'yanikova, 3-e izd. M.: Tekhnosfera, 2010. 560 s. (In Russian)
2. Koraboshev O. Z. Analiz i perspektivy primeneniya metodov mashinnogo obucheniya dlya chrezvychajnyh situacij // Intellektual'nye tekhnologii na transporte. 2024. № 1 (37). S. 12–17. DOI: 10.20295/2413-2527-2024-137-12-17 (In Russian)
3. Rashka S. Python i mashinnoe obuchenie / per. s angl. A. V. Logunova. M.: DMK Press, 2017. 418 s. (In Russian)
4. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Second Edition. Sebastopol (CA): O'Reilly Media, 2019. 848 p.
5. Terebizh V. Yu. Vvedenie v statisticheskuyu teoriyu obratnyh zadach. M.: Fizmatlit, 2005. 376 s. (In Russian)
6. Mochalov V. F., Habarov R. S. Obrabotka materialov mul'tispektral'noj s'emki na osnove metodov mashinnogo obucheniya pri upravlenii sostoyaniem lesnogo massiva // Sbornik materialov IV Mezhdunarodnoj nauchnoj konferencii po problemam upravleniya v tekhnicheskikh sistemah (PUTS-2021) (Sankt-Peterburg, 21–23 sentyabrya 2021). SPb.: LETI, 2021. S. 269–272. (In Russian)
7. Maxwell A. E., Warner T. A., Guillén L. A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies. Part 1: Literature Review // Remote Sensing. 2021. Vol. 13, Is. 13. Art. № 2450. 27 p. DOI: 10.3390/rs13132450

Received: 28.06.2024

Accepted: 16.09.2024