қазақстан-британ техникалық университетінің **ХАБАРШЫСЫ**

HERALD of the kazakh-british technical university

ВЕСТНИК казахстанско-британского технического университета

Volume 19, Issue 1 January-March 2022

ҚАЗАҚСТАН-БРИТАН ТЕХНИКАЛЫҚ УНИВЕРСИТЕТІНІҢ **ХАБАРШЫСЫ**

HERALD OF THE KAZAKH-BRITISH TECHNICAL UNIVERSITY

ВЕСТНИК КАЗАХСТАНСКО-БРИТАНСКОГО ТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Алматы

№ 1 (60)

2022

Главный редактор – Кулпешов Б.Ш.

члены редколлегии:

Акжалова А.Ж., Асилбеков Б.К., Ахметжанов А.А., Баженов Н.А., Байжанов Б.С., Бейсенханов Н.Б., Бисембаев А.С., Буркитбаев М.М., Зазыбин А.Г., Ивахненко А.П., Исахов А.А., Исмаилов А.А., Kang Wanli, Кожабеков С.С., Колесников А.В., Коробкин В.В., Курбатов А.П., Нусупов К.Х., Пак А.А., Сарсенбекулы Б., Судоплатов С.В., Тургазинов И.К., Умаров Ф.Ф., Шамои П.С.

Издание зарегистрировано Министерством культуры и информации Республики Казахстан. Свидетельство о постановке на учет СМИ № 9757 – Ж от 03.12.2008 г.

Журнал зарегистрирован в Международном центре по регистрации сериальных изданий ISSN (ЮНЕСКО, г. Париж, Франция)

Подписной индекс – 74206

Издается с 2004 года. Выходит 4 раза в год.

УЧРЕДИТЕЛЬ Казахстанско-Британский технический университет

ISSN 1998-6688

© Казахстанско-Британский технический университет, 2022.

МҰНАЙ ГАЗ ИНЖЕНЕРИЯСЫ ЖӘНЕ ГЕОЛОГИЯ

Исмагулова Д.М.

Жасанды нейрондық желісін қолдана отырып, шикі мұнай	өңдеу зауытында фракциялық
айдау бағанын оңтайландыру	

ФИЗИКА-МАТЕМАТИКА ЖӘНЕ ТЕХНИКАЛЫҚ ҒЫЛЫМДАР

Айтжанова С.А.
Өлшемдер орындау әдістемелері, шолу және қолдану17
Амирбек Г.С., Кулпешов Б.Ш.
Омега-категориялық дерлік әлсіз о-минималды теорияларында бинарлық дөңестік рангісі23
Асубаева Е.М., Абдиахметова З.М.
Trade-in авто бағаларын болжау кезінде жіктеу әдістерін салыстырмалы талдау
Мархабатов Н.Д., Судоплатов С.В.
Тұрақты графтардың аппроксимациялары44

СОДЕРЖАНИЕ

НЕФТЕГАЗОВАЯ ИНЖЕНЕРИЯ И ГЕОЛОГИЯ

Исмагулова Д.М.

Оптимизация ректификационной колонны на нефтеперерабатывающем заводе с использованием	
искусственной нейронной сети	.6

ФИЗИКО-МАТЕМАТИЧЕСКИЕ И ТЕХНИЧЕСКИЕ НАУКИ

Айтжанова С.А.

Методики выполнения измерений, обзор и применение1	7
Амирбек Г.С., Кулпешов Б.Ш.	
Бинарный ранг выпуклости в почти омега-категоричных слабо о-минимальных	
сеориях	3
Асубаева Е.М., Абдиахметова З.М.	
Сравнительный анализ методов классификации данных при прогнозировании	
цен trade-in авто	0
Мархабатов Н.Д., Судоплатов С.В.	
Аппроксимации регулярных графов44	4

OIL AND GAS ENGINEERING, GEOLOGY

Ismagulova D.M.

PHYSICAL, MATHEMATICAL AND TECHNICAL SCIENCES

Aitzhanova S.A. Measurement procedures, reference and application17
Amirbek G.S., Kulpeshov B.Sh.
Binary convexity rank in almost omega-categorical weakly o-minimal theories23
Assubayeva Y.M., Abdiakhmetova Z. M.,
Comparative analysis of data classification methods for prediction of trade-in auto prices
Markhabatov N.D., Sudoplatov S.V.
Approximations of regular graphs

UDC 62-551.4 IRSTI 50.47.29

https://doi.org/10.55452/1998-6688-2022-19-1-6-16

OPTIMIZATION OF FRACTIONAL DISTILLATION COLUMN IN CRUDE OIL REFINERY USING ARTIFICIAL NEURAL NETWORK

ISMAGULOVA D.M.

JSC "Wood KSS", 060011, Atyrau, Kazakhstan

Abstract. The paper outlines the methods, which improve the controlling process of separating methanol from water in the distillation column to produce crude oil products. Nowadays, many industries use PID controllers to control process variables like temperature, flow, pressure, level, which helps maintain good performances. However, PID controllers can have slightly bad performances in complicated control systems, such as in Multiple Input and Multiple Output (MIMO) systems; due to this, optimization methods of improving PID are considered. A tremendous amount of work has been done refining, studyingandimproving the PID controlling techniquesand methods. However, PID still faces challenges in a variety of common control problems. This article represents NeuralNetworkAlgoritmbased PID controller, whichisusedtocontrolthe separating process of methanol from water in the distillation column, due to Neuralnetwork's good generalization results. The Wood and Berry mathematical Model was chosen as the main control object.

Keywords: distillation column, Neural network, PID controllers, Artificial intelligence, MIMO system.

ЖАСАНДЫ НЕЙРОНДЫҚ ЖЕЛІСІН ҚОЛДАНА ОТЫРЫП, ШИКІ МҰНАЙ ӨҢДЕУ ЗАУЫТЫНДА ФРАКЦИЯЛЫҚ АЙДАУ БАҒАНЫН ОҢТАЙЛАНДЫРУ

ИСМАГУЛОВА Д.М.

«Wood KSS» АҚ, 060011, Атырау қ., Қазақстан

Аңдатпа. Мақалада метанолды судан дистилляциялау бағанында шикі мұнай өнімдерін алу үшін бөлуді бақылау процесін жақсартатын әдістер көрсетілген. Қазіргі уақытта көптеген өнеркәсіптер PID контроллерлерін температура, ағын, қысым, деңгей сияқты айнымалыларды басқару үшін пайдаланады, бұл жақсы көрсеткіштерді сақтауға көмектеседі. Алайда PID контроллерлері күрделі басқару жүйелерінде, мысалы, бірнеше енгізу және бірнеше шығару (MIMO) жүйелерінде сәл нашар көрсеткіштерге ие болуы мүмкін, сондықтан PID-ді жақсартудың оңтайландыру әдісі қарастырылады. PID техникасын зерттеуге, жетілдіруге, басқарудың жетілдірілген әдістерін жасауға көптеген жылдар жұмсалды. Дегенмен әлі де PID контроллері жауап бере алмайтын бірқатар жалпы басқару қиындықтары бар. Бұл жұмыста, нейрондық желісінің жақсы жалпылау нәтижелеріне байланысты, жасанды нейрондық желісіне негізделген PID контроллері метанолды дистилляциялық бағандағы судан бөлу процесін бақылау үшін қолданылады. Басқарудың негізгі объектісі ретінде Вуд және Берриматематикалық моделі таңдалды.

Түйінді сөздер: дистилляция бағаны, нейрондық желі, PID контроллері, жасанды интеллект, *MIMO* жүйесі.

ОПТИМИЗАЦИЯ РЕКТИФИКАЦИОННОЙ КОЛОННЫ НА НЕФТЕПЕРЕРАБАТЫВАЮЩЕМ ЗАВОДЕ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ

ИСМАГУЛОВА Д.М.

АО «Wood KSS», 060011, г. Атырау, Казахстан

Аннотация. В статье изложены методы, позволяющие улучшить управление процессом отделения метанола от воды в ректификационной колонне с целью получения сырых нефтепродуктов. В текущее время почти во всех отраслях индустрии используются ПИД-регуляторы для управления такими переменными процессами, как температура, расход, давление, уровень, что действительно помогает поддерживать хорошие характеристики. Однако ПИД-контроллеры могут иметь несколько плохую производительность в сложных системах управления, например в системах с многоканальным выходом (МІМО), из-за чего рассматривается метод оптимизации для улучшения ПИД. Было потрачено много лет на изучение, уточнение и совершенствование техники ПИД-регулирования, а также на разработку улучшенных методов управления. Тем не менее все еще существует ряд общих проблем управления, при которых ПИД-регулятор сталкивается с трудностями. В этой работе ПИД-регулятор на основе нейронной сети используется для управления процессом отделения метадов в дистилляционной колонне благодаря хорошим результатам нейронной сети. Математическая модель Вуда и Берри была выбрана в качестве основного объекта управления.

Ключевые слова: ректификационная колонна, нейронная сеть, ПИД-регуляторы, искусственный интеллект, система МІМО.

Introduction

The crude oil refining industry has a huge effect oneveryone's life. Numerous products can be produced using raw petroleum since it is a soup of various sorts of hydrocarbon molecules, each of which has its own set of unique chemical and physical properties. Furthermore, these attributes make every particular hydrocarbon in crude oil either a good fuel, a useful fluid or solid. This work will consider methods for controlling the fractional distillation column for producing crude oil products. The objective of the work is to control the process of separating methanol from water in the distillation column using a Neural Network based PID controller in the atmospheric distillation column to send it to the storage tank properly. Artificial neural networks have been used in a variety of applications by several manufacturers for fault detection, control management and pattern recognition. The most significant advantage of ANN is learning from historical data and being utilized for many industrial applications. Furthermore, compared to conventional PID control algorithms, neural-based PID improves the system's realtime characteristics and complexity. [1] In turn, Mostafa MJAHED [2] shows good performance and benefits from decreased values of rising time, overshootingand settling times and lesser oscillatory response using the Genetic Algorithm. Compared to traditional tuning methods, Genetic Algorithms outperformed them in terms of steadystate response and output performances. Compared to the research above, in [3] the presented neural PID model, the PID coefficients are considered Gaussian potential function networks (GPFN) weights. Furthermore, they are fine-tuned using an online learning algorithm. So, the presented model outperforms the basic PID controller with fixed gains in terms of capability and flexibility. The PID neural network performs admirably in terms of position control and behavior [4]. Genetic algorithm with particle swarm optimization results demonstrates that the settling time, overshoot percentage, rise time are better than conventional PID controllers [5]. Ibtissem Chiha, Noureddine Liouane, and Pierre Borne propose Multiobjective ant colony optimization to tune PID controllers. The results show that the proposed tuning technique outperforms the genetic algorithms and traditional approach and control system performance [6]. Another concept of artificial neural networks (ANN) was proposed to establish new setpoints after system disturbances and proved to have a much better speed and feasible solution [7].

The presented new model improves the performance of the traditional controller; this new control approach is conceptually simple and can be easily implemented in oil and other industries. Furthermore, additional energy costs and costs associated with product specifications can be avoided. In work [8], Muravyova and Mustaev, 2017 solved a problem associated with the large error in the amount of cement at the outlet relative to a given capacity, as well as to increase the speed of the control system and increase its fault tolerance by using an artificial neural network in the Matlab environment. M.M. Gouda, S. Danaher, C. P. in the research [9] designed a more robust and efficient fuzzy logic controller. Itdecreases the sensitiveness of the systemand improvesfast changes of theparameters, and has lower energy consumption. The research paper [10] represents that the performance of PID with the Ziegler -Nichols method is better than the conventional Ziegler-Nichols technique. Experiments show minimum overshoot, settling time for rate demand utilities of DC motor. Different types of artificial intelligence were compared; among them, Neural Network was chosen as the main controller for our distillation column.

Problem statement

Many industries use Proportional-Integral-Derivative (PID) controllers to maintain and regulate process variables. In in industrial process loops, PID is the most common feedback control systems. They are simple to comprehend and put into practice. Many scientists have spent time studying, refining, and improving the PID controlling methods, as well as designing workarounds for the flaws they've discovered.

However, there are a variety of common control issues that PID can't solve, some of which can be solved with appropriate augmentations. Due to the aggressiveness of control processes, a conventional PID controller would have issues regulating them. Also, advanced control is necessary for MIMO systems, where the controller must coordinate the initiatives of several actuators to manipulateseveral control variables simultaneously.

For example, the distillation column, as it is a MIMO system, requires advanced control. Moreover, they can face quite bad performances like overshooting, steady-state error, response time increase, etc. Those problems can lead to hazardous situations and loss of money. Due to this, new methods of process control are being developed. In this research, a new method of improving PID control is presented.

Relevance of the project

Proportional-integral-derivative (PID) controllers are commonly used to regulate the process variables of many different types of dynamic systems. Due to itssimple structure and ability to provide an excellent closed-loop response characteristic, these controllers are significant in control process. Nonetheless, selecting a suitable PID controller might be challenging. Various methods for tuning controllers have been developed in response to this issue. The Ziegler-Nichols method is the most frequently used tuning technique, but sometimes it can be difficult to establish optimal controller coefficients. As a result, many artificial intelligence algorithms, such as neural network (ANN), fuzzy logic, swarm, ant colony optimizations and others, have been created to tune PID parameters.

The main drawbacks of the PID controller include the complexity of controlling the three parameters and the fact that it does not work well for systems with time-varying, nonlinear systems, linear systems with a time delay and complex systems. AI-based controllers have more benefits than traditional PID controllers, such as independence, better reliability, lower load, smarter control loops, higher speedand adaptability in the enterprise, regardless of human intervention.

Atmospheric Distillation Column

An atmospheric distillation column with a tray contacting device is mainly used to separate the crude oil components into its fractions: valuable products like gasoline, LPG, kerosene, Diesel fuel, naphtha, and heavy gas oil. Tray or plates enable good separation of the fractions of crude oil. The working principle of ADU is fractional distillation or distillation on boiling ranges. An atmospheric distillation column is demonstrated in figure 1.

Trays located inside the ADU collect various fractions as they cool to their boiling value and vaporize. Forthe oil to be vaporized at the bottom of the column, the reboiler heats the crude oil to $350 \degree C$. Using a condenser, each fraction of crude oil is cooled and condensed at various temperature values at the top of the column.

As each fraction of condensation, the liquid is collected in the trays of the column. Higher boiling fractions condense on the column's lower trays, and lowersteaming point fractions condense on the higher trays. A reflux drum is used to keep the condensed vapor, resulting in which reflux can be sent back from the top of the distillation column. One input stream and two product streams make up the distillation column.

Mathematical model

There was chosen a mathematical model of distillation column created by Wood and Berry. They established a mathematical model of an 8-tray binary fractional distillation column with a complete condenser and a basket-style reboiler for disunion of methanol from water.

The model has been frequently utilized in recent decades in several researches to evaluate the efficacy of various control methods since it has been proven useful. Equations (1) and (2) were used to express it when it was discovered experimentally:

$$\begin{bmatrix} x_D(s) \\ x_B(s) \end{bmatrix} = G(s) \begin{bmatrix} R(s) \\ S(s) \end{bmatrix}$$
(1)

$$\begin{bmatrix} x_D(s) \\ x_B(s) \end{bmatrix} = \begin{bmatrix} \frac{12.8}{16.7s+1}e^{-s} & \frac{-18.9}{21s+1}e^{-3s} \\ \frac{6.6}{10.9s+1}e^{-7s} & \frac{-19.4}{14.4s+1}e^{-3s} \end{bmatrix} \begin{bmatrix} R(s) \\ S(s) \end{bmatrix}$$
(2)

Here, outputs (controlled variables) are:

 $x_D(s)$ – methanol proportion in the distillate,

 $x_B(s)$ – the amount of methanol of lowerproducts;

Inputs (manipulated variables) are:

R(s) – reflux flow speed,

S(s) – flow speed of steam in the reboiler. The P&ID diagram of the distillation column for this process is illustrated in Fig. 1. P&ID diagram was developed in the online "Visual Paradigm" tool.



Figure 1 – P&ID diagram of fractional distillation column

The control structure of the distillation column

The steam and reflux flow rates regulate the top and bottom product concentrations, while the feed flow rate acts as a disturbance. A distillation column must have at least four feedback control loops to monitor and control distillate concentration, bottom concentration, reboiler level, and reflux rate level. As a result, it's classified as a MIMO control system. This paper proposes a controller for controlling distillate and bottom concentrations. MIMO control can be accomplished by using two controllers, one for each component (two outputs). Figure 2 shows a block scheme of a PID controller-based MIMO system for a distillation column.



Figure 2- PID controller-based MIMO controller

Here: $G_{c1}(s), G_{c2}(s)$ – PID Controllers, $G_{11}(s)G_{12}(s)G_{21}(s)G_{22}(s)$ – transfer functions of control objects.

Due to multiple variables, the control structure of the MIMO system reqresdecoupler elements. The block diagram of the control system with decoupler is illustrated in Fig. 3. Decouplers helps to remove process interactions to turn the MIMO system into interacting, allowing for smooth tuning of control loops. Equation (3) shows decoupler elements $D_{12}(s)$ an $D_{21}(s)$:

$$D_{12}(s) = -\frac{G_{12}(s)}{G_{11}(s)}$$
(3)

$$D_{21}(s) = -\frac{G_{21}(s)}{G_{22}(s)}$$



Figure 3- MIMO with simplified decoupling method

The diagonal elements Q(s) of the product G(s)D(s) as a diagonal matrix are the independent SISO systems that should be found.

Simplified decoupling is presented in Equations 4 and 5:

$$D(s) = \begin{bmatrix} 1 & D_{12}(s) \\ D_{21}(s) & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{G_{12}(s)}{G_{11}(s)} \\ -\frac{G_{21}(s)}{G_{22}(s)} & 1 \end{bmatrix}$$
(4)

$$Q = \begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} \begin{bmatrix} 1 & D_{12}(s) \\ D_{21}(s) & 1 \end{bmatrix} = \begin{bmatrix} G_{11} - \frac{G_{12}G_{21}}{G_{22}} & 0 \\ 0 & G_{22} - \frac{G_{12}G_{21}}{G_{11}} \end{bmatrix}$$
(5)

The controller design is shown in Equation 6:

$$K(s) = \begin{bmatrix} K_1(s) & 0\\ 0 & K_2(s) \end{bmatrix} = \begin{bmatrix} K_{p1} + \frac{K_{i1}}{s} + K_{d1}s & 0\\ 0 & K_{p2} + \frac{K_{i2}}{s} + K_{d2}s \end{bmatrix}$$
(6)

In such a way, the block diagram of the MIMO system shown in figure 3 can be simplified to two





Figure 4 - Independent SISO systems

So, the obtained Diagonal matrix (Q_1 and Q_2) of the system are shown in Equations 7, 8:

$$Q_1$$

$$=\frac{1.896 * 10^4 * s^3 + 4042 * s^2 + 291.5 * s + 6.37}{2.676 * 10^4 * s^5 + 3.591 * 10^4 * s^4 + 1.026 * 10^4 * s^3 + 1157 * s^2 + 56.6 * s + 1} (7)$$

 Q_2

$$= \frac{-3.294 * 10^4 \text{s}^3 - 6758 \text{s}^2 - 461.2 \text{s} - 9.655}{8.9 * 10^4 \text{s}^5 + 5.814 * 10^4 \text{s}^4 + 1.281 * 10^4 \text{s}^3 + 1271 \text{s}^2 + 58.3 \text{s} + 1}$$
(8)

Stability verification

The step response is the system's response to a unit step input; it helps analyze the system's stability and find characteristics such as overshoot, the steady-state error, and other dynamic characteristics of the system.

independent SISO systemsillustrated in figure 4.

The step response for the first closed-loop system is illustrated in fig. 5 below.



Figure 5 – The step response of the 1st system

Figure 5 illustrates that system is underdamped and has overshoots and oscillations. Using the *step info()* command, it was found that the overshoot of the system was equal to $M_p = 6.64\%$, settling time t_s is 6.0801 seconds and rise time t_r is 2.1397 seconds. The steady-state error e_s . equal to 0.139. So, using PID controllers with an Artificial Neural Network algorithm, it is expected that the error will be eliminated, and the output response will be improved.

The step response for the second closed-loop system is illustrated in Figure 6 below.



Figure 6 illustrates that the system is not stable. The results show that the system doesn't have any overshoots $(%M_p)$, settling time (t_s) , rise time (t_r) or any other characteristics. The steady-

state error e_{ss} Was equal to 8.7242e^24. So, using PID controllers with an Artificial Intelligence algorithm, it is expected that the output response will be improved.

Artificial Neural Network

The purpose of an artificial neural network(ANN) is to use the target data to construct a system that accurately corresponds the input values to the output. The obtained model is used to get the intended output when the desired output is unrevealed. Neuralnetwork-based PID controllers aims to eliminate computing complexity and improve real-time performance comparing to a traditional PID controller. A control structure of ANN-basedPID controller is represented in fig. 7.



Figure 7 – Structure of PID controller based on neural network

Fig. 7 shows that the controller is made up of two parts: traditional PID control and a neural network. Here the conventional PID affects the control object directly. The PID controller's parameters are tuned using a neural network, which compares the target values with the input values to achieve performance optimization. Output neurons must match the parameters of PID.

Training neural network

To train a neural network, I used PID parameters obtained using a Genetic Algorithm and used them as target values. To get target values, a scheme was constructed where the output response of the closed-loopwas exported to the Workspace using simout1. For the input values of the NN, there were extracted output response values of the closed-loop without any PID controller.

There were used "*nnstart*" neural network fitting app for this article. A sigmoid like transfer function of a two-layer feedforward network is used for a hidden layer of our network in Matlab software. The output layer is based on the linear transfer function (TF). Thenetwork will be trained with the Levenberg-Marquardt backpropagation algorithm. The structure of the neural network, which has input, hidden, and output layers, as shown in figure 8.



Figure 8 – Neural network structure

10 neurons were selected for a hidden layer of neural network architecture and one neuron both for input, output layers as in the picture above. The training results of ANN for the first (on the left side) and second (on the right side) closedloop systems is illustrated in figure 9.

Neural Network	Neural Network							
Hidden Output	Output 1	Hidden Output Nutput 10 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0						
Algorithms		Algorithms						
Data Division: Random (dividerand) Training: Levenberg-Marquardt (trainlm) Performance: Mean Squared Error (mse) Calculations: MEX		Data Division: Random (dividerand) Training: Levenberg-Marquardt (trainIm) Performance: Mean Squared Error (mse) Calculations: MEX						
Progress		Progress						
Epoch: 0 106 iterations	1000	Epoch: 0 911 Iterations 1000						
Time: 0:00:02		Performance: 0.0988 0.00510 0.00						
Performance: 0.726 7.59e-06	0.00	Gradient: 0.332 4.11e-06 1.00e-07						
Gradient: 2.21 3.87e-05	1.00e-07	Mu: 0.00100 1.00e-10 1.00e+10						
Mu: 0.00100 1.00e-08	1.00e+10	Validation Checks: 0 6 6						
Validation Checks: 0 6	6	Plots						
Plots		Performance (plotperform)						
Performance (plotperform)		Training State (plottrainstate)						
Training State (plottrainstate)		Error Histogram (ploterrhist)						
Error Histogram (ploterrhist)		Regression (plotregression)						
Regression (plotregression)		Fit (plotfit)						
Fit (plotfit)		Plot Interval:						
Plot Interval:		Validation stop.						

Figure 9 – Neural network training results (First loop)

Overall, the training performance for the first system was slightly good compared to the second loop. The first loop was trained at 106 Epochs and the second at 911 Epochs, which is very long.

Finally, to check the output response of two

independent closed-loopsystems, a block diagram of the PID controller feedback system was constructed with an Artificial Neural Network, shown in figure 10.



Figure 10 - Neural network-based PID controller

The results of the first feedback system are illustrated in figure 11.



Figure 11 – Step response of 1st loop with NN based PID

Figure 11 shows that the output response of the first loop using NN based PID controller was improved. It was found that the overshoot of the system was equal to $M_p=0.0397\%$, settling time t_s is 0.0194 seconds and rise time t_r is 0.0150

seconds. So, by using PID controllers with the Artificial Neural Network algorithm, there was obtained better results.

The step response for the second closed-loop system with NN PID is illustrated in Figure 12.



Figure 12 – The step response of the 2nd system

The results show that the system hasan overshoot equal to $M_p=0.0397\%$, settling time $t_s=1.1777$, rise time $t_r - 0.9038$. So, NN based PID controller helped to make the system stable and improve overall performance.

Conclusion

In this article, the techniques of improving traditional PID controllers in crude oil refineries are considered. To sum up, there was studied the usage of artificial intelligence in different fields. As seen from the research done, the artificial intelligence-based PID controllers are one of the optimal algorithms that can be used for controlling the object and whole system. The artificial neural network helps to predict future plant behaviors. Neural networks are widely applied to solve technical problems. Based on my research, it was discovered that the majority of researchers use ANN for modeling and designing linear and nonlinear systems for industrial control systems.

Moreover, training a Neural network-basedPID controller feedback system to control the process of separating methanol from water in a distillation column in a crude oil refinery showed good performances compared to the system without the AI network. It was proven that by using a neural network, the system response improved overshoots eliminated, rise time and settling times decreased. Applying the NN based PID controllers and replacing the conventional PID controllers reduced the error between desired values and system output.

REFERENCES

1 Mohammadi A., Ryu J.-C. Neural network-based PID compensation for nonlinear systems: ballon-plate example // International Journal of Dynamics and Control, Springer, 2018, pp. 1–11.

2 Mostafa Mjahed. PID Controller Design using Genetic Algorithm. Technique Ecole Royale de l'Air, Mathematics and Systems Department, 40000 Marrakech, Morocco, pp. 312–318.

3 Yuan X., Xiang Y., Wang Y., Yan X. Neural Networks Based PID Control of Bidirectional Inductive Power Transfer System, Springer, 2015, vol. 43, issue 3, pp. 837–847.

4 Zhang, S., Liu, X., Sheng, Y. Analysis and System Simulation of Flight Vehicle Sliding Mode Control Algorithm Based on PID Neural Network. Lecture Notes in Real-Time Intelligent Systems, Springer, 2017, pp. 312–318.

5 Eng. A. Salem, Dr. M. Ammar, Prof. Dr. M. Moustafa. Tuning PID Controller Using Artificial Intelligence Techniques // 9th International Conference on Electrical Engineering (ICEENG 2014), MTC, 29-31 (May, 2014, Cairo, Egypt). 6 Ibtissem Chiha, Noureddine Liouane and Pierre Borne. Tuning PID Controller Using Multiobjective Ant Colony Optimization, vol. 2012, pp. 1–7.

7 Das Neves T.G., Ramos W.B., de Farias Neto G.W., Brito R. P. Intelligent control system for extractive distillation columns // Korean Journal of Chemical Engineering, Springer, 2015, vol. 35, Issue 4, pp. 826–834.

8 Muravyova E.A. and Mustaev R.R. Development of an Artificial Neural Network for Controlling Motor Speeds of Belt Weighers and Separator in Cement Production, Optical Memory and Neural Networks, Springer, 2017, vol. 26, no. 4, pp. 289–297.

9 Gouda M.M., Danaher S., Underwood C.P. Fuzzy logic control versus conventional PID control for controlling indoor temperature of a building space. IFAC Computer Aided Control Systems Design, Salford, UK, 2000.

10 Manoj Kushwah, Ashish Patra. PID Controller Tuning using Ziegler Nichols Method for Speed Control of DC Motor // International Journal of Scientific Engineering and Technology Research, vol.03, issue13, 2014, pp. 2924–2929.

Information about author

Ismagulova Dana Maksutkyzy

Master, Associate Automation and Control Engineer, JSC "Wood KSS", st. Kanysh Satbaev, 46, 060011, Atyrau, Kazakhstan; ORCID ID: 0000-0002-3423-9701; E-mail: danaismagulova7@gmail.com.

Автор туралы мәлімет

Исмагулова Дана Мақсұтқызы

Магистр, автоматтандыру және басқарма кіші инженері, «Wood KSS» АҚ, Қаныш Сәтбаев көшесі, 46, 060011, Атырау қ., Қазақстан; ORCID ID: 0000-0002-3423-9701; E-mail: danaismagulova7@gmail.com.

Сведения об авторе

Исмагулова Дана Максуткызы

Магистр, младший инженер по автоматизации и управлению, AO «Wood KSS», ул. Каныша Сатпаева, 46, 060011, г. Атырау, Казахстан; ORCID ID: 0000-0002-3423-9701; E-mail: danaismagulova7@gmail.com.

ФИЗИКО-МАТЕМАТИЧЕСКИЕ И ТЕХНИЧЕСКИЕ НАУКИ

УДК 389.14:006.354 МРНТИ 90.27.27:37

https://doi.org/10.55452/1998-6688-2022-19-1-17-22

МЕТОДИКИ ВЫПОЛНЕНИЯ ИЗМЕРЕНИЙ, ОБЗОР И ПРИМЕНЕНИЕ

АЙТЖАНОВА С.А.

Актюбинский филиал РГП «Казахстанский институт стандартизации и метрологии», 030000, г. Актобе, Казахстан

Аннотация. В настоящей статье изложены нормативные правовые акты, стандарты и применение методик выполнения измерений. Описание нормы характерных показателей (требований) на каждую продукцию, будь то хлеб, молоко, нефтепродукты, изделия легкой или тяжелой промышленности и т.д., регламентируется в технических регламентах, а далее в нормативных документах на продукцию, где для каждого показателя обычно бывают ссылки на методы испытаний этой характеристики продукции. И для испытания продукции всегда необходим метод проведения испытаний. В большинстве случаев для данных целей уже разработаны стандартные методы проведения испытаний и измерений, которые описаны в отечественных или межгосударственных стандартах: СТ РК или ГОСТ, а также в международных – IEC, ISO и др. Однако для проведения испытаний в отдельных случаях стандартные методы отсутствуют. Решением задачи в данном случае является разработка собственной методики выполнения измерений, так называемой нестандартизированной, методики, где погрешности измерений при условии соблюдения всех требований документа гарантированы. В статье описаны стадии разработки, аттестации, утверждения и применения как обычных методик, так и референтных методик выполнения измерений. В статье приведены ссылки из законодательства Республики Казахстан в области обеспечения единства измерений на стандарты, нормативные правовые акты и Закона РК «Об обеспечении единства измерений».

Ключевые слова: методика, методика выполнения измерений, МВИ, государственная система обеспечения единства измерений, ГСИ

ӨЛШЕМДЕР ОРЫНДАУ ӘДІСТЕМЕЛЕРІ, ШОЛУ ЖӘНЕ ҚОЛДАНУ

АЙТЖАНОВА С.А.

«Қазақстан стандарттау және метрология институты» РМК, «ҚазСтандарт» РМК Ақтөбе филиалы, 030000, Ақтөбе қ., Қазақстан

Аңдатпа. Осы мақалада нормативтік құқықтық актілер, стандарттар және өлшемдер орындау әдістемелерінің қолданылуы мазмұндалған. Нан, сүт, мұнай өнімдері, жеңіл немесе ауыр өнеркәсіп бұйымдары және т.б. сияқты әрбір өнімдерге тән көрсеткіштердің (талаптардың) нормаларының сипаты техникалық регламенттерде, кейін әдетте әрбір көрсеткіш үшін өнімнің осы сипаттамасын сынау әдістеріне сілтемелер келтірілген өнімге арналған нормативтік құжаттарда реттеледі. Өнімді сынау үшін сынақтарды өткізу әдісі қажет. Көптеген жағдайларда осы мақсаттар үшін отандық немесе мемлекетаралық стандарттарда – ҚР СТ немесе ГОСТ, сондай-ақ халықаралық IEC, ISO және т.б. стандарттарда сипатталған сынақтар мен өлшемдерді орындаудың стандартты әдістері әзірленген. Алайда жекелеген жағдайларда сынақтарды өткізу үшін стандартты әдістер жоқ. Осы жағдайда мәселенің шешімі стандартталмаған деп аталатын өз өлшемдер орындау

ВЕСТНИК КАЗАХСТАНСКО-БРИТАНСКОГО ТЕХНИЧЕСКОГО УНИВЕРСИТЕТА, №1 (60), 2022

әдістемелерін әзірлеу болып табылады, мұнда құжаттың барлық талаптарын сақтаған кезде өлшемдердің қателіктері кепілдендірілген. Мақалада қарапайым, сондай-ақ референттік өлшемдер орындау әдістемелерін әзірлеу, аттестаттау, бекіту және қолдану сатылары сипатталған. Мақалада өлшем бірлігін қамтамасыз ету саласындағы Қазақстан Республикасы заңнамасынан стандарттар, нормативтік құқықтық актілер мен «Өлшем бірлігін қамтамасыз ету туралы» Заңына сілтемелер келтірілген.

Түйінді сөздер: әдістеме, өлшемдер орындау әдістемесі, ӨОӘ, мемлекеттік өлшем бірлігін қамтамасыз ету жүйесі, МӨЖ

MEASUREMENT PROCEDURES, REFERENCE AND APPLICATION

AITZHANOVA S.A.

Aktobe branch of RSE "KazStandart"

RSE "Kazakhstan Institute of Standardization and Metrology", 030000, Aktobe, Kazakhstan

Abstract. This article outlines regulatory legal acts, standards and the application of measurement procedures. The description of the norm of characteristic indicators (requirements) for each product, be it bread, milk, petroleum products, light or heavy industry products, etc. is regulated in technical regulations, and further in regulatory documents for products, where for each indicator there are usually references to test methods of this product characteristic. For testing products, a test method is always necessary. In most cases, standard methods of testing and measurements have already been developed for these purposes, which are described in domestic or interstate standards: ST RK or GOST, as well as in international IEC, ISO, etc. However, there are no standard methods for testing in some cases. The solution to the problem in this case is the development of its own measurement methodology, the so-called non-standardized methodology, where measurement errors, subject to compliance with all the requirements of the document, are guaranteed. The article describes the stages of development, certification, approval and application of both conventional methods and reference measurement procedures. The article contains references from the legislation of the Republic of Kazakhstan in the field of ensuring the unity of measurements to standards, regulatory legal acts and the Law «On ensuring the unity of measurements».

Keywords: Procedure, measurement procedure, MP, state system for ensuring the unity of measurements, SSM

Введение

Методики, или методики выполнения измерений (МВИ), в соответствии с терминологией статьи 1 Закона Республики Казахстан [1] и Правил [2] определены как комплекс правил и норм, обеспечивающих получение результатов измерений с определенной погрешностью.

Основные стандарты, устанавливающие порядок разработки, метрологической аттестации и применения методик, – это СТ РК 2.18 [3], ГОСТ 8.010 [4].

Основная часть

В связи с гармонизацией международных и государственных требований к объектам оценки соответствия значительно расширился перечень регламентируемых характеристик для разных

групп промышленной и продовольственной продукции, где качество и безопасность продукции характеризуются техническими, технологическими и метрологическими нормами. Характеристики метрологических и технических норм должны гарантироваться достоверностью получаемых результатов, что достигается в результате обеспечения единства измерений. В этой связи особое внимание уделяется методикам, методам и средствам измерений, испытательному и вспомогательному оборудованию, стандартным образцам и аттестованным смесям, применяемым для контроля качества, а также для обеспечения безопасности продукции.

МВИ используются для обеспечения норм точности в различных отраслях экономики, а также для выполнения установленных требований и норм технических регламентов Таможенного союза (далее – ТР ТС) в рамках ЕАЭС, обеспечивающих качество и безопасность продукции и оказываемых услуг в республике. Поэтому возникает необходимость пересмотра и совершенствования действующих МВИ и разработки новых методик.

Согласно СТ РК 2.18 [3], ГОСТ 8.010 [4] МВИ должны содержать как минимум следующие разделы, описывающие требования к:

- описанию области применения методик;

- погрешности измерений (точности измерений, приписанным характеристикам погрешности);

- техническим и метрологическим характеристикам, применяемым к средствам измерений, испытательному и вспомогательному оборудованию, стандартным образцам;

- условиям выполнения измерений, безопасности рабочих мест операторов и экологии, уровню квалификации специалистов;

- описанию подготовки к выполнению измерений;

- описанию отбора и подготовки образца пробы;

- описанию выполнения измерений;

- обработке, оформлению полученных результатов измерений, контролю погрешности результатов измерений по данной методике.

Лаборатория, применяющая МВИ, где показатели точности измерений, такие как повторяемость, внутрилабораторная прецизионность, воспроизводимость, правильность и точность анализа, установлены согласно РМГ 61 [5] или погрешности по ГОСТ 8.207 [6], контролирует данные приписанные характеристики по РМГ 76 [7]. Нормы для внутрилабораторного контроля устанавливаются с учетом приписанных характеристик погрешности или ее составляющих.

Разработанные и метрологически аттестованные методики подлежат регистрации в реестре государственной системы обеспечения единства измерений Республики Казахстан (ГСИ РК). Как правило, это МВИ, к которым установлены метрологические требования в нормативных правовых актах, в перечнях измерений ст. 22 [1]. Данную процедуру выполняет Государственный научный метрологический центр (ГНМЦ) при РГП «КазСтандарт».

Этапы разработки методик, затем аттестации и регистрации в реестре ГСИ РК:

- разработка Технического задания (исход-

ных данных) и программы на МВИ, выбор метода или методов испытаний, измерений, исследование характеристик, параметров методики (проведение экспериментальных испытаний, измерений и составление технического отчета), согласование МВИ с надзорными органами и заинтересованными предприятиями, организациями и утверждение, оформление МВИ по Приложению Б ГОСТ 8.010 [4], СТ РК 1.5 [8], СТ РК 2.505 [9].

- метрологическая аттестация и утверждение МВИ.

Самые актуальные области применения методик выполнения измерений – это охрана окружающей среды, защита жизни и здоровья граждан, безопасность труда и движения транспорта, оценка соответствия, государственные учетные операции, торговые операции между покупателем и продавцом, измерения в сфере коммунальных услуг и услуг связи.

Приведем примеры одних из немногих МВИ, разработанных в области нефти и газа, а также экологии:

- «МВИ установками измерительными «ОЗНА-МАССОМЕР» по определению массы нефти и объема нефтяного газа различными методами»;

- «МВИ СИКН (система измерений количества и показателей качества нефти)» для разных предприятий при месторождениях, согласованных с АО «КазТрансОйл»;

- «МВИ СИРГ (система измерений расхода газа)» для разных предприятий при месторождениях, согласованных с АО «КазТрансОйл»;

- «Методика выполнения измерений. Определение натрия хлористого, натрия сернокислого, сухого остатка, влаги и нерастворимого остатка в отходах производства титриметрическим и гравиметрическим методами»;

- «Методика выполнения измерений содержания урана в технологических урансодержащих растворах и твердых полупродуктах потенциометрическим методом с использованием титратора»;

- «Количество извлекаемых из недр сырой нефти, нефти без учета воды, нефтяного газа. Методика выполнения измерений многофазными расходомерами Vx для компании «Шлюмберже Лоджелко Инк.» в PK»;

- «МВИ массовых концентраций действующих веществ пестицидов (гербицидов) в препаративных формах, высокочистых пестицидах в твердых и жидких матрицах методом газожидкостной хроматографии».

Возможно также использование аттестованных МВИ, разработанных в странах СНГ в соответствии с требованиями межгосударственного документа ПМГ 44 [10].

Подробную информацию о зарегистрированных методиках выполнения измерений можно найти на сайте https://techreg.qoldau.kz [11], а также информация о МВИ, зарегистрированных до 11 апреля 2019 г., приведена в архиве Реестра ГСИ РК на сайте https:// ksm.kz [12].

Безусловно, МВИ состоят из основных объектов ГСИ РК: это измеряемая единица величины, методы измерений, метрологические характеристики стандартных образцов и средств измерений. Использование новейших технологий при разработке методик обуславливает развитие системы передачи размеров единиц измерений от государственных эталонов рабочим эталонам и от последних – всем средствам измерений. При этом важно соблюдать обеспечение прослеживаемости единицы измерения исследуемого объекта МВИ к государственным или рабочим эталонам, государственным стандартным образцам. Также необходимо учитывать, что, согласно статье 9 [1], в республике допускается применять единицы величин Международной системы единиц, принятой Генеральной конференцией по мерам и весам в соответствии с документами Международной организации законодательной метрологии (МОЗМ), а также в установленном порядке по ГОСТ 8.417 [13].

При отсутствии государственных или рабочих эталонов, отсутствии стандартных образцов возможно применение референтных методик выполнения измерений. Статья 1 [1] и Правила [2] регламентируют, что референтную методику используют для оценки правильности результатов измерений, полученных по другим МВИ. При этом величины единиц результатов измерений референтной МВИ должны быть того же рода, что и в исследуемой МВИ. Также референтные методики применяются для калибровки средств измерений или аттестации метрологических характеристик стандартных образцов.

Для применения методики в качестве референтной МВИ лаборатории необходимо проведение опробования данной методики и описание прослеживаемости ее результатов до государственного стандартного образца либо до процедур, связанных с государственными или эталонами высших разрядов с высоким уровнем неопределенности. Кроме того, лаборатория должна быть аккредитована, что является показателем правильности выполнения методики референтного измерения и соответствующего применения используемых средств измерений, оборудования, реактивов и реагентов.

Разработанные лабораториями референтные МВИ подтверждаются и затем публикуются национальными метрологическими институтами (или международными организациями) в сотрудничестве с Международным комитетом по весам и мерам (МКМВ; International Committee for Weightsand Measures; CIPM).

Вывод

Таким образом, повышение точности результатов измерений МВИ, установление и подтверждение соответствия методики предъявляемым к ней метрологическим требованиям, обеспечение прослеживаемости результатов до эталонов, или стандартного образца, или референтной методики с требуемым для измерений уровнем неопределенности являются важным условием для обеспечения единства измерений.

ЛИТЕРАТУРА

1 Об обеспечении единства измерений. Закон РК от 7 июня 2000 г. № 53-II. URL: https://adilet.zan. kz/rus/docs/Z00000053.

2 Правила разработки, метрологической аттестации, утверждения и регистрации в реестре государственной системы обеспечения единства измерений методик выполнения измерений и референтных методик выполнения измерений. НПА. Приказ министра по инвестициям и развитию РК от 27 дек. 2018 г. № 932. URL: https://adilet.zan.kz/rus/docs/V1800018108/history.

3 ГСИ РК. Методики выполнения измерений. Порядок разработки, метрологической аттестации, регистрации и применения. СТ РК 2.18-2009. URL: http://shop.ksm.kz.

4 ГСИ. Методики выполнения измерений. Основные положения. ГОСТ 8.010-2013. URL: http:// shop.ksm.kz.

5 ГСИ. Показатели точности, правильности, прецизионности методик количественного химического анализа. Методы оценки. РМГ 61-2010. URL: http://shop.ksm.kz.

6 ГСИ. Прямые измерения с многократными наблюдениями. Методы обработки результатов наблюдений. Основные положения. ГОСТ 8.207-76. URL: http://shop.ksm.kz.

7 ГСИ. Внутренний контроль качества результатов количественного химического анализа. РМГ 76-2014. URL: http://shop.ksm.kz.

8 Общие требования к построению, изложению, оформлению и содержанию стандартов. СТ РК 1.5-2019. URL: http://shop.ksm.kz.

9 ГСИ РК. Построение, изложение, оформление и содержание документов на методики выполнения измерений. СТ РК 2.505-2018. URL: http://shop.ksm.kz.

10 ГСИ. Порядок признания методик выполнения измерений. ПМГ 44-2001. URL: http://shop.ksm.kz.

11 URL: https://techreg.qoldau.kz.

12 URL: https:// ksm.kz.

13 ГСИ. Единицы величин. ГОСТ 8.417-2002. URL: http://shop.ksm.kz.

REFERENCES

1 Ob obespechenii edinstva izmerenii. Zakon RK ot 7 iyunya 2000 g. № 53-II. URL: https://adilet.zan. kz/rus/docs/Z000000053.

2 Pravila razrabotki, metrologicheskoi attestatsii, utverzhdeniya i registratsii v reestre gosudarstvennoi sistemy obespecheniya edinstva izmerenii metodik vypolneniya izmerenii i referentnykh metodik vypolneniya izmerenii. NPA. Prikaz ministra po investitsiyam i razvitiyu RK ot 27 dek. 2018 g. № 932. URL: https://adilet.zan.kz/rus/docs/V1800018108/history.

3 GSI RK. Metodiki vypolneniya izmerenii. Poryadok razrabotki, metrologicheskoi attestatsii, registratsii i primeneniya. ST RK 2.18-2009. URL: http://shop.ksm.kz.

4 GSI. Metodiki vypolneniya izmerenii. Osnovnye polozheniya. GOST 8.010-2013. URL: http://shop. ksm.kz.

5 GSI. Pokazateli tochnosti, pravil'nosti, pretsizionnosti metodik kolichestvennogo khimicheskogo analiza. Metody otsenki. RMG 61-2010. URL: http://shop.ksm.kz.

6 GSI. Pryamye izmereniya s mnogokratnymi nablyudeniyami. Metody obrabotki rezul'tatov nablyudenii. Osnovnye polozheniya. GOST 8.207-76. URL: http://shop.ksm.kz.

7 GSI. Vnutrennii kontrol' kachestva rezul'tatov kolichestvennogo khimicheskogo analiza. RMG 76-2014. URL: http://shop.ksm.kz.

8 Obshchie trebovaniya k postroeniyu, izlozheniyu, oformleniyu i soderzhaniyu standartov. ST RK 1.5-2019. URL: http://shop.ksm.kz.

9 GSI RK. Postroenie, izlozhenie, oformlenie i soderzhanie dokumentov na metodiki vypolneniya izmerenii. ST RK 2.505-2018. URL: http://shop.ksm.kz.

10 GSI. Poryadok priznaniya metodik vypolneniya izmerenii. PMG 44-2001. URL: http://shop.ksm.kz. 11 URL: https://techreg.goldau.kz.

12 URL: https:// ksm.kz.

13 GSI. Edinitsy velichin. GOST 8.417-2002. URL: http://shop.ksm.kz.

Сведения об авторе

Айтжанова Саулеш Амангельдыевна

Главный специалист РГП «Казахстанский институт стандартизации и метрологии», Актюбинский филиал РГП «КазСтандарт», ул. С. Нурмагамбетова, 16, 030000, г. Актобе, Казахстан; ORCID ID: 0000-0002-9726-6184;

E-mail: s.aytzhanova@ksm.kz.

Автор туралы мәлімет

Айтжанова Саулеш Амангельдықызы

«Қазақстан стандарттау және метрология институты» РМК Бас маманы, «ҚазСтандарт» РМК Ақтөбе филиалы, Нұрмағамбетов к-ші 1б, 030000, Ақтөбе қ., Қазақстан; ORCID ID: 0000-0002-9726-6184; E-mail: s.aytzhanova@ksm.kz.

Information about author

Aitzhanova Saulesh Amangeldyyevna

Chief Specialist of RSE "Kazakhstan Institute of Standardization and Metrology", Aktobe branch of RSE "KazStandart", Nurmagambetova str. 1b, 030000, Aktobe, Kazakhstan;

ORCID ID: 0000-0002-9726-6184;

E-mail: s.aytzhanova@ksm.kz.

UDC 510.67 IRSTI 27.03.66

https://doi.org/10.55452/1998-6688-2022-19-1-23-29

BINARY CONVEXITY RANK IN ALMOST OMEGA-CATEGORICAL WEAKLY O-MINIMAL THEORIES

AMIRBEK G.S., KULPESHOV B.SH.

Kazakh-British Technical University, 050000, Almaty, Kazakhstan

Abstract. The present paper concerns the notion of weak o-minimality that was initially deeply studied by D. Macpherson, D. Marker and C. Steinhorn. A subset A of a linearly ordered structure M is convex if for all $a, b \in A$ and $c \in M$ whenever a < c < b we have $c \in A$. A weakly o-minimal structure is a linearly ordered structure $M = \langle M, =, <, ... \rangle c$ such that any definable (with parameters) subset of M is a union of finitely many convex sets in M. A criterion for equality of the binary convexity ranks for non-weakly orthogonal non-algebraic 1-types in almost omega-categorical weakly o-minimal theories in case of existing an element of the set of realizations of one of these types the definable closure of which has a non-empty intersection with the set of realizations of another type is found.

Keywords: weak o-minimality, almost omega-categoricity, convexity rank, weak orthogonality, equivalence relation.

ОМЕГА-КАТЕГОРИЯЛЫҚ ДЕРЛІК ӘЛСІЗ О-МИНИМАЛДЫ ТЕОРИЯЛАРЫНДА БИНАРЛЫҚ ДӨҢЕСТІК РАНГІСІ

АМИРБЕК Г.С., КУЛПЕШОВ Б.Ш.

Қазақстан-Британ техникалық университеті, 050000, Алматы қ., Қазақстан

Аңдатпа. Мақала бастапқыда Д. Макферсон, Д. Маркер және Ч. Стайнхорн терең зерттеген әлсіз о-минималдылық түсінігіне қатысты. Сызықтық реттелген М құрылымының А ішкі жиыны дөңес болады, егер кез келген а, $b \in A$ және $c \in M$ кезінде a < c < b бізде $c \in b$ бізде $c \in A$ болса. Әлсіз о-минималды құрылым – бұл М құрылымының кез келген анықталатын (параметрлері бар) ішкі жиыны М-дегі дөңес жиындардың ақырлы санының бірігуі болатындай $M = \langle M, =, <, ... \rangle$ сызықты реттелген құрылым. Бинарлық дөңестік рангілері теңдігінің критерийі әлсіз ортогональды емес алгебралық емес 1-типтері үшін дерлік омега-категориялық әлсіз о-минималды теорияларда осы түрлердің біреуінің жүзеге асу жиынынан элемент болған жағдайда табылады, оның анықталатын жабылуы басқа түрдегі іске асыру жиынымен бос емес қиылысы бар.

Түйінді сөздер: әлсіз о-минималдық, дерлік омега-категориялық, дөңестік рангісі, әлсіз ортогоналдық, эквиваленттік қатынас.

БИНАРНЫЙ РАНГ ВЫПУКЛОСТИ В ПОЧТИ ОМЕГА-КАТЕГОРИЧНЫХ СЛАБО О-МИНИМАЛЬНЫХ ТЕОРИЯХ

АМИРБЕК Г.С., КУЛПЕШОВ Б.Ш.

Казахстанско-Британский технический университет, 050000, г. Алматы, Казахстан

Аннотация. Настоящая статья касается понятия слабой о-минимальности, первоначально глубоко исследованного Д. Макферсоном, Д. Маркером и Ч. Стайнхорном. Подмножество А линейно упорядоченной структуры М является выпуклым, если для любых а, $b \in A$ и $c \in M$ всякий раз, когда a < c < b, мы имеем с А. Слабо о-минимальной структурой называется линейно упорядоченная структура $M = \langle M, =, <, ... \rangle$ такая, что любое определимое (с параметрами) подмножество структуры М является объединением конечного числа выпуклых множеств в М. Найден критерий равенства бинарных рангов выпуклости для не слабо ортогональных неалгебраических 1-типов в почти омега-категоричных слабо о-минимальных теориях в случае существования элемента из множества реализаций одного из этих типов, определимое замыкание которого имеет непустое пересечение со множеством реализаций другого типа.

Ключевые слова: слабая о-минимальность, почти омега-категоричность, ранг выпуклости, слабая ортогональность, отношение эквивалентности.

Introduction

Let L be a countable first-order language. Throughout this paper we consider L-structures and suppose that L contains a binary relation symbol < which is interpreted as a linear order in these structures. The notion of weak o-minimality was originally studied in [1]. Real closed fields with a proper convex valuation ring provide an important example of weakly o-minimal structures [2, 3].

Let A and B be arbitrary subsets of a linearly ordered structure M. Then the expression A < B means that a < b whenever $a \in B$, and A < b means that $A < \{b\}$. For an arbitrary subset A of M we introduce the following notations: $A^+:=\{b \in M \mid A < b\}$ and $A^-:=\{b \in M \mid b < A\}$. For an arbitrary one-type p we denote by p(M) the set of realizations of p in M. If $B \subseteq M$ and E is an equivalence relation on M then we denote by B/E the set of equivalence classes (E-classes) which have representatives in B. If f is a function on M then we denote by Dom(f) the domain of f. A theory T is said to be binary if every formula of the theory T is equivalent in T to a boolean combination of formulas with at most two free variables.

Definition 1. Let T be a weakly o-minimal theory, $M \models T$, $A \subseteq M$, p, $q \in S_1(A)$ be non-algebraic. We say that p is not weakly orthogonal to q (denoting this byp $\mathscr{L}^w q$) if there exist an L_A -formula H(x, y), $\alpha \in p(M)$ and $\beta_1, \beta_2 \in q(M)$ such that $\beta_1 \in H(M, \alpha)$ and $\beta_2 \notin H(M, \alpha)$.

In other words, p is weakly orthogonal to q (denoting this by $p \perp^w q$) if $p(x) \cup q(y)$ has a unique extension to a complete 2-type over A.

Lemma 2. [4] Let T be a weakly o-minimal theory, $M \models T$, $A \subseteq M$. Then the relation of non-weak orthogonality \mathscr{L}^w is an equivalence relation on $S_1(A)$.

Definition 3 [5] Let T be a weakly o-minimal theory, M is a sufficiently saturated model of T, $A \subseteq M$. The rank of convexity of the set A (RC(A)) is defined as follows:

1) RC(A) = -1 if $A = \emptyset$

2) RC(A) = 0 if A is finite and non-empty.

3) $RC(A) \ge 1$ if A is infinite.

4) $RC(A) \ge \alpha + 1$ if there exist a parametrically definable equivalence relation E(x, y) and an infinite sequence of elements $b_i \in A$, $i \in \omega$ such that:

For every $i, j \in \omega$ whenever $i \neq j$ we have $M \models \neg E(b_i, b_i)$;

For every $i \in \omega$ $RC(E(M, b_i)) \ge \alpha$ and $E(M, b_i)$ is a convex subset of A.

5) $RC(A) \ge \delta$, if $RC(A) \ge \alpha$ for all $\alpha < \delta$, where δ is a limit ordinal.

If $RC(A) = \alpha$ for some α , we say that RC(A) is defined. Otherwise (i.e. if RC(A)) $\geq \alpha$ for all α), we put $RC(A) = \infty$.

The rank of convexity of a formula $\phi(x, \bar{a})$, where $\bar{a} \in M$, is defined as the rank of convexity of the set $\phi(M, \bar{a})$, i.e. $\text{RC}(\phi(x, \bar{a})) := \text{RC}(\phi(M, \bar{a}))$. The rank of convexity of an 1-type p is defined as the rank of convexity of the set p(M), i.e. RC(p) := RC(p(M)).

In particular, a theory has convexity rank 1 if there are no definable (with parameters) equivalence relations with infinitely many infinite convex classes.

We say that the convexity rank of an arbitrary set

A is binary and denote it by $RC_{bin}(A)$ if in Definition 3 parametrically definable equivalence relations are replaced by \emptyset -definable (i.e. binary) equivalence relations.

Definition 4. [6, 7] Let T be a complete theory, and $p_1(x_1)$, ..., $p_n(x_n) \in S_1(\emptyset)$. A type $q(x_1, ..., x_n) \in S_n(\emptyset)$ is said to be a $(p_1, ..., p_n)$ -type if

$$q(x_1, \ldots, x_n) \supseteq p_1(x_1) \cup p_2(x_2) \cup \ldots \cup p_n(x_n).$$

The set of all $(p_1, ..., p_n)$ -types of the theory T is denoted by $S_{p1,...,pn}(T)$. A countable theory T is said to be almost ω -categorical if for any types $p_1(x_1), ..., p_n(x_n) \in in S_1(\emptyset)$ there are only finitely many types $q(x_1, ..., x_n) \in S_{p1,...,pn}(T)$.

Almost ω -categoricity is closely connected with the notion of Ehrenfeuchtness of a theory. So in [6] it was proved that if T is an almost ω -categorical theory with I(T, ω) = 3 then a dense linear order is interpreted in \$T\$. Nonetheless there is an example (constructed by M.G. Peretyat'kin in [8]) of a theory with the condition I(T, ω) = 3 that is not almost ω -categorical.

In [9] the authors established almost ω -categoricity of Ehrenfeucht quite o-minimal theories and that the Exchange Principle for the algebraic closure holds in almost ω -categorical quite o-minimal theories. Recently in [10] orthogonality of any family of pairwise weakly orthogonal non-algebraic 1-types over \emptyset for such theories and binarity of almost ω -categorical quite o-minimal theories were proved. Also, in [11] binarity of almost omega-categorical weakly o-minimal theories of convexity rank 1 was established. At last, in the work [12] a criterion for binarity of almost omega-categorical weakly o-minimal theories in terms of convexity rank was found.

Theorem 5. [10] Let T be an almost omegacategorical weakly o-minimal theory, $p \in S_1(\emptyset)$ be non-algebraic. Then $RC_{bin}(p) < \omega$.

Recall some notions originally introduced in [1]. Let $Y \subset M^{n+1}$ be an \emptyset -definable subset, let π : $M^{n+1} \rightarrow M^n$ be the projection which drops the last coordinate, and let $Z := \pi(Y)$. For each $\overline{a} \in Z$ let $Y \overline{a} := \{y: (\overline{a}, y) \in Y\}$. Suppose that for every $\overline{a} \in Z$ the set $Y \overline{a}$ is convex and bounded above but does not have a supremum in M. We let ~ \emptyset -definable equivalence relation on M^n given by

 $\overline{a} \sim \overline{b}$ for all \overline{a} , $\overline{b} \in M^n \setminus Z$, and $\overline{a} \sim \overline{b} \Leftrightarrow$ sup Y $\overline{a} = \sup Y \overline{b}$ if \overline{a} , $\overline{b} \in Z$.

Let $\overline{Z} := Z / \sim$, and for each tuple $\overline{a} \in Z$ we denote by $[\overline{a}]$ the \sim -class of \overline{a} . There is a natural

 \emptyset -definable total order on $M \cup \overline{Z}$, defined as follows. Let $\overline{a} \in Z$ and $c \in M$. Then $[\overline{a}] < c$ if and only if w < c for all $w \in Y\overline{a}$. Also, we say $c < [\overline{a}]$ iff $\neg ([\overline{a}] < c)$, i.e. there exists $\in Y\overline{a}$ such that $c \le w$. If \overline{a} is not \sim -equivalent to \overline{b} then there is some $x \in M$ such that $[\overline{a}] < x < [\overline{b}]$ or $[\overline{b}] < x < [\overline{a}]$ and so <induces a total order on $M \cup \overline{Z}$ We call such a set \overline{Z} a sort (in this case, \emptyset -definable sort) in \overline{M} , where \overline{M} is the Dedekind completion of M, and view \overline{Z} as naturally embedded in \overline{M} . Similarly, we can obtain a sort in \overline{M} by considering infima instead of suprema.

Thus, we will consider definable functions from M to its Dedekind completion \overline{M} , more precisely in definable sorts of the structure \overline{M} , representing infima or suprema of definable sets.

Let A, D \subseteq M, D be infinite, Z $\subseteq \overline{M}$ be an A-definable sort and f: D \rightarrow Z be an A-definable function. We say f is locally increasing (locally decreasing, locally constant}) on D if for any a \in D there is an infinite interval J \subseteq D containing {a} so that f is strictly increasing (strictly decreasing, constant) on J; we also say f is locally monotonic on D if it is locally increasing or locally decreasing on D.

Let f be an A-definable function on $D \subseteq M$, E be an A-definable equivalence relation on D. We say f is strictly increasing (decreasing) on D/E if for any a, $b \in D$ with a < b and $\neg E(a, b)$ we have f(a) < f(b) (f(a) > f(b)).

Proposition 6. [13] Let M be a weakly o-minimal structure, $A \subseteq M$, $p \in S_1(A)$ be a nonalgebraic type. Then any A-definable function of which the domain contains the set p(M) is locally monotonic or locally constant on p(M).

Results

Definition 7 (Verbovskiy V.V., [14, 15]) Let M be a weakly o-minimal structure, B, D \subseteq M, A $\subseteq \overline{M}$ be a B-definable sort and f: D \rightarrow A be a B-definable function that is locally increasing (decreasing) on D. We say that the function f has depth n on the set D if there exist equivalence

relations $E_1(x, y), ..., E_n(x, y)$ partitioning D into infinitely many infinite convex classes so that for every $2 \le i \le n$ each E_i -class is partitioned

into infinitely many infinite convex E_{i-1} -subclasses and the following holds:

• f is strictly increasing (decreasing) on each E_1 -class;

• f is strictly decreasing (increasing) on D/E_k for every odd k \leq n (or the same, f is strictly decreasing (increasing) on each $E_{k+1}(a, M)/E_k$ for any $a \in D$);

• f is locally increasing (decreasing) on D/E_k for every even $k \le n$;

• f is strictly monotonic on D/E_{p} .

In this case, we say that the function f is locally increasing (decreasing) of depth n.

Obviously, a strictly increasing (decreasing) function is locally increasing (decreasing) of depth 0.

Theorem 8 (Verbovskiy V.V., [15]) Let T be a weakly o-minimal theory. Then any definable function into a definable sort has a finite depth.

Proposition 9 [4] Let T be a weakly o-minimal theory, $M \models T$, $A \subseteq M$, p, $q \in S_1(A)$, be non-algebraic, $p \measuredangle^w q$. Then the following holds:

(1) p is irrational \Leftrightarrow q is irrational;

(2) p is quasirational \Leftrightarrow q is quasirational.

Theorem 10. Let T be an almost ω -categorical weakly o-minimal theory, $M \models T$, $p, q \in S_1(\emptyset)$ be non-algebraic, $p \not\perp^w q$, $dcl(\{a\}) \cap q(M) \neq \emptyset$ for some $a \in p(M)$. Then the following conditions are equivalent:

(1) $RC_{bin}(p) > RC_{bin}(q);$

(2) there is no an \varnothing -definable function f: p(M) \rightarrow q(M) being a bijection of p(M) on q(M);

(3) dcl({b}) \cap p(M) = \emptyset for any b \in q(M);

(4) there exist an \varnothing -definable function f: p(M)

 \rightarrow q(M) being locally constant on p(M).

Proof of Theorem 10. By Proposition 9 the types p and q are either isolated or quasirational or irrational simultaneously. Without loss of generality, suppose that p and q are isolated. The remaining cases are considered similarly.

(1) \Rightarrow (2). Assume the contrary: there exists an $\square \emptyset$ -definable function f: p(M) \rightarrow (M) being a bijection of p(M) on q(M).

Let $RC_{bin}(p) = n$. Then there exist \emptyset -definable equivalence relations $E_1(x, y)$, $E_2(x, y)$, ..., $E_{n-1}(x, y)$ which partition p(M) into infinitely many infinite convex classes so that

$$E'_{1}(x, y) := \exists t_{1} \exists t_{2} [E_{1}(t_{1}, t_{2}) \land f(t_{1}) = x \land f(t_{2}) = y],$$

for some (any) $a \in p(M)$. Consider the following formulas:

$$\begin{split} E'_1(x,\,y) &:= \exists \ t_1 \ \exists \ t_2 \ [E_1(t_1,\,t_2) \land \ f(t_1) = x \land \ f(t_2) \\ = y], \end{split}$$

 $E'_{n-1}(x, y) := \exists t_1 \exists t_2 [E_{n-1}(t_1, t_2) \land f(t_1) = x \land f(t_2) = y].$

By Theorem 8 the function f is strictly monotonic on each E_1 -class and f is strictly monotonic on each $E_{k+1}(a, M)/E_k$ for any $a \in p(M)$, where $1 \le k \le n-2$. Therefore we have that $E'_1(x, y), \ldots, E'_{n-1}(x, y)$ are equivalence relations partitioning q(M) into infinitely many infinite convex classes so that

 $E'_1(b, M) \subset E'_2(b, M) \subset \ldots \subset E'_{n-1}(b, M),$

whence $RC_{bin}(q) \ge n$, that contradicts the hypothesis.

(2) \Rightarrow (3). Since dcl({a}) \cap q(M) $\neq \emptyset$ there exist b \in q(M) and an L-formula $\varphi(x, y)$ such that

$$M \vDash \exists ! y \varphi(a, y) \land \varphi(a, b).$$

Assume the contrary: $dcl(\{b\}) \cap p(M) \neq \emptyset$ Note that $a \in dcl(\{b\})$. Otherwise there exists $a_1 \in p(M)$ such that $a_1 \neq a$ and $a_1 \in dcl(\{b\})$. Since $b \in dcl(\{a\})$, we have that $a_1 \neq dcl(\{a\})$, and this implies an infinity of dcl($\{a\}$), contradicting the almost ω -categoricity of T. Thus, $a \in dcl(\{b\})$. Then there exists an L-formula $\varphi'(x, y)$

$$M \models \exists ! y \varphi'(a, y) \land \exists ! x \varphi'(x, b) \land \varphi'(a, b).$$

Define the function f as follows: $f(a) = b \Leftrightarrow \phi'(a, b)$. It is not difficult to see that f bijectively maps p(M) onto q(M), contradicting our assumption.

(3) ⇒ (4). Assume the contrary: f: p(M) → q(M) is an \emptyset -definable function and f is not locally constant on p(M). Then f must be locally monotonic on p(M), i.e. either locally increasing or locally decreasing by Proposition 6. But then f bijectively maps p(M) onto q(M). Then dcl({b}) \cap p(M) ≠ \emptyset for some (any) b ∈ q(M) which contradicts (3).

(4) \Rightarrow (1). Let f: p(M) \rightarrow q(M) be an $\square \emptyset$ -definable function being locally constant on p(M). Consider the following formula:

$$\begin{split} E(x, y) &:= [x < y \rightarrow \forall \ t \ (x < t < y \rightarrow f(x) = f(t)) \\ = f(y))] \land \end{split}$$

$$\wedge [x > y \rightarrow \forall t (x > t > y \rightarrow f(x) = f(t) = f(y))].$$

Clearly, E(x, y) is an equivalence relation partitioning p(M) into infinitely many infinite convex classes. Let $RC_{bin}(p) = n$. Then there exist \emptyset definable equivalence relations $E_1(x, y)$, $E_2(x, y)$, ..., $E_{n-1}(x, y)$ partitioning p(M) into infinitely many infinite convex classes so that

$$E_1(a, M) \subset E_2(a, M) \subset \ldots \subset E_{n-1}(a, M)$$

for some (any) $a \in p(M)$.

Obviously, for some $1 \le i \le n-1$ we have that $E(x, y) \equiv E_i(x, y)$. Then we assert that $RC_{bin}(q) = n - i$. Indeed, f is a constant on each E_i -class. Further, we consider the behaviour of the function f on each $E_{i+1}(a, M)/E_i$, where $a \in p(M)$. It must be strictly monotonic on each $E_{i+1}(a, M)/E_i$, since otherwise there exists an $\square \emptyset$ -definable equivalence relation $\overline{E}(x, y)$ such that

$$E_i(a, M) \subset E(a, M) \subset E_{i+1}(a, M)$$

which contradicts that the relation E_{i+1} is an immediate successor of the relation $E_i(x, y)$ among all $\square \varnothing$ -definable equivalence relations on p(M). Similarly, we can prove that f is strictly monotonic on each $E_{k+1}(a, M)/E_k$, where $i \le k \le n-2$ and f is strictly monotonic on $p(M)/E_{n-1}$.

Consider the following formulas:

$$\begin{split} E'_{i+1}(x,\,y) &:= \exists \ t_1 \ \exists \ t_2 \ [U_p(t_1) \land U_p(t_2) \land E_{i+1}(t_1,\\ t_2) \land \ f(t_1) &= x \land \ f(t_2) = y], \end{split}$$

...

$$\begin{split} E'_{n\text{-}1}(x,\,y) &:= \exists \ t_1 \ \exists \ t_2 \ [U_p(t_1) \land U_p(t_2) \land E_{n\text{-}1}(t_1,\\ t_2) \land \ f(t_1) &= x \land \ f(t_2) = y]. \end{split}$$

We can establish that $E'_{i+1}(x, y), ..., E'_{n-1}(x, y)$ are equivalence relations partitioning q(M) into infinitely many infinite convex classes so that

$$E'_{i+1}(b, M) \subset E'_{i+2}(b, M) \subset \ldots \subset E'_{n-1}(b, M),$$

whence $RC_{hin}(q) = \emptyset$ -definable equivalence

Conclusion

We have found necessary and sufficient conditions in order to the binary convexity ranks of non-weakly orthogonal non-algebraic 1-types in almost omegacategorical weakly o-minimal theories were equal in the case of existing some definable function between the sets of realizations of these 1-types. relation $E^q(x, y)$ partitioning q(M) into infinitely many infinite convex classes so that

$$E^{q}(b, M) \subset E'_{i+1}(b, M),$$

consider the following formula:

$$\hat{E}(\mathbf{x},\mathbf{y}) := \exists t_1 \exists t_2 [E^q(t_1,t_2) \land f(\mathbf{x}) = t_1 \land f(\mathbf{y})$$

= t_2].

Obviously,

$$E_i(a, M) \subset \hat{E}(a, M) \subset E_{i+1}(a, M)$$

contradicting also that the relation E_{i+1} is an immediate successor of the relation $E_i(x, y)$ among all \emptyset -definable equivalence relations on p(M). Similarly, we can prove that there is no an \emptyset -definable equivalence relation $E^q(x, y)$ partitioning q(M) into infinitely many infinite convex classes so that

$$\begin{split} & E'_{k}(b, M) \subset Eq~(b, M) \subset E'_{k+1}(b, M) \\ & \text{for every } i+1 \leq k \leq n-2 \text{ or} \\ & E'_{n-1}(b, M) \subset E^{q}(b, M). \\ & \text{Thus, } RC_{bin}(q) = n-i, i.e. \ RC_{bin}(p) > RC_{bin}(q). \end{split}$$

Corollary 11. Let T be an almost $\[mu]$ -categorical weakly o-minimal theory, p, q \in S₁($\[mu]$) be non-algebraic, $\[mu]^{W}$, dcl({a}) \cap q(M) $\[mu]^{\neq \emptyset}$ for some a \in p(M). Then the following conditions are equivalent:

- (1)} $\operatorname{RC}_{\operatorname{bin}}(p) = \operatorname{RC}_{\operatorname{bin}}(q);$
- (2)} there exists an \emptyset -definable function f: p(M) \rightarrow q(M) being a bijection of p(M) on q(M);
 - (3)} dcl({b}) \cap p(M) $\neq \emptyset$ for any b \in q(M); (4)} there exists an \emptyset -definable function f: p(M)
- \rightarrow q(M) being locally monotonic on p(M).

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP08855544).

REFERENCES

1 Macpherson H.D., Marker D. and Steinhorn C. Weakly o-minimal structures and real closed fields // Transactions of The American Mathematical Society, vol. 352, issue 12, 2000, pp. 5435–5483.

2 Dickmann M. Elimination of quantifiers for ordered valuation rings // The Journal of Symbolic Logic, vol. 52, 1987, pp. 116–128.

3 Van Den Dries L., Lewenberg A.H. T-convexity and tame extensions // The Journal of Symbolic Logic, vol. 60, issue 1, 1995, pp. 74-102.

4 Baizhanov B.S. Expansion of a model of a weakly o-minimal theory by a family of unary predicates // The Journal of Symbolic Logic, vol. 66, isuue 3, 2001, pp. 1382–1414.

5 Kulpeshov B.Sh. Weakly o-minimal structures and some of their properties // The Journal of Symbolic Logic, vol. 63, issue 4, 1998, pp. 1511–1528.

6 Ikeda K., Pillay A., Tsuboi A. On theories having three countable models // Mathematical Logic Quarterly, vol. 44, issue 2, 1998, pp. 161–166.

7 Sudoplatov S.V. Classification of countable models of complete theories, part 1. Novosibirsk: Novosibirsk State Technical University Publishing House, 2018, ISBN 978-5-7782-3527-4, 326 p.

8 Peretyat'kin M.G. A theory with three countable models // Algebra and Logic, vol. 19, issue 2, 1980, pp. 139–147.

9 Kulpeshov B.Sh., Sudoplatov S.V. Linearly ordered theories which are nearly countably categorical // Mathematical Notes, vol. 101, issue 3, 2017, pp. 475–483.

10 Altayeva A.B., Kulpeshov B.Sh. Binarity of almost w -categorical quite o-minimal theories // Siberian Mathematical Journal, vol. 61, issue 3, 2020, pp. 379–390.

11 Kulpeshov B.Sh., Mustafin T.S. Almost w -categorical weakly o-minimal theories of convexity rank 1 // Siberian Mathematical Journal, 2021, vol. 62, no. 1, pp. 52–65.

12 Kulpeshov B.Sh. A criterion for binarity of almost w categorical weakly o-minimal theories // Siberian Mathematical Journal, 2021, vol. 62, no. 6, pp. 1063–1075.

13 Kulpeshov B.Sh. Countably categorical quite o-minimal theories // Journal of Mathematical Sciences, vol. 188, issue 4 (2013), pp. 387–397.

14 Verbovskiy V.V. On depth of functions of weakly o-minimal structures and an example of a weakly o-minimal structure without a weakly o-minimal theory // Proceedings of Informatics and Control Problems Institute, 1996, pp. 207–216.

15 Verbovskiy V.V. On formula depth of weakly o-minimal structures // Algebra and Model Theory (A.G. Pinus and K.N. Ponomaryov, editors), Novosibirsk, 1997, pp. 209–223.

Information on authors

1. Amirbek Gaukhar Samatkyzy

Master Student, School of Mathematics and Cybernetics, Kazakh-British Technical University, 59, Tole bi street, 050000, Almaty, Kazakhstan; ORCID ID: 0000-0003-1442-4691; E-mail: ga amirbek@kbtu.kz.

2. Kulpeshov Beibut Shaiykovich (corresponding author)

Doctor of Physical and Mathematical Sciences, Professor, School of Mathematics and Cybernetics, Kazakh-British Technical University, 59, Tole bi street, 050000, Almaty, Kazakhstan;

ORCID ID: 0000-0002-4242-0463;

E-mail: b.kulpeshov@kbtu.kz.

Авторлар туралы мәліметтер

1. Амирбек Гаухар Саматқызы

Магистрант, Математика және кибернетика факультеті, Қазақстан-Британ техникалық университеті, Төле би көшесі, 59, 050000, Алматы қ., Қазақстан;

ORCID ID: 0000-0003-1442-4691; E-mail: ga_amirbek@kbtu.kz.

2. Кулпешов Бейбіт Шайыкович (корреспонденция авторы)

Физика математика ғылымдарының докторы, профессор, Математика және кибернетика факультеті, Қазақстан-Британ техникалық университеті, Төле би көшесі, 59, 050000, Алматы қ, Қазақстан;

ORCID ÎD: 0000-0002-4242-0463; E-mail: b.kulpeshov@kbtu.kz.

Сведения об авторах

1. Амирбек Гаухар Саматкызы

Магистрант, факультет математики и кибернетики, Казахстанско-Британский технический университет, ул. Толе би, 59, 050000, г. Алматы, Казахстан;

ORCID ID: 0000-0003-1442-4691;

E-mail: ga_amirbek@kbtu.kz.

2. Кулпешов Бейбут Шайыкович (автор для корреспонденции)

Доктор физико-математических наук, профессор, факультет математики и кибернетики, Казахстанско-Британский технический университет, ул. Толе би, 59, 050000, г. Алматы, Казахстан; ORCID ID: 0000-0002-4242-0463;

E-mail: b.kulpeshov@kbtu.kz.

УДК 004.852 МРНТИ 28.23.25

https://doi.org/10.55452/1998-6688-2022-19-1-30-43

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КЛАССИФИКАЦИИ ДАННЫХ ПРИ ПРОГНОЗИРОВАНИИ ЦЕН TRADE-IN АВТО

АСУБАЕВА Е.М., АБДИАХМЕТОВА З.М.

Казахский национальный университет имени аль-Фараби, 050040, г. Алматы, Казахстан

Аннотация. В статье реализованы и проанализированы алгоритмы машинного обучения для предсказания цен авто. Предсказание цен – одна из самых сложных, но интересных задач. В предсказании задействовано много факторов – год выпуска, состояние, пробег, объем двигателя и т.д. Эти аспекты в совокупности влияют на цены авто, делая их нестабильными и затрудняя прогнозирование с высокой степенью точности. Методы машинного обучения могут выявить закономерности и идеи, которые мы раньше не видели, и их можно использовать для безошибочно точных прогнозов и классификации данных. Выбор надлежащего алгоритма классификации данных, который подходил бы для отдельно взятой задачи, зависит от объема, качества и природы данных, от вычислительных ресурсов компьютера, а также от того, как вы планируете использовать результат. Каждый алгоритм классификации имеет свои особенности и основывается на определенных допущениях. В конечном счете качество классификатора, его вычислительная и предсказательная мошность зависят от базовых данных, предназначенных для тренировки алгоритма. Цель данной статьи – рассмотреть этапы предварительной обработки тренировочных данных и показать, как машинное обучение в частности и информационные технологии в целом преуспели в разработке инструментов для моделирования, проектирования, прогнозирования, планирования и поддержки принятия решений в области продажи авто. В данном исследовании предлагается гибридный подход к задачам прогнозирования, то есть к решению задач прогнозирования с применением методов статистического анализа и машинного обучения.

Ключевые слова: машинное обучение, задача классификации, логистическая регрессия, случайный лес, дерево принятия решений, *k*-ближайших соседей, *RESTAPI*.

ТRADE-IN АВТО БАҒАЛАРЫН БОЛЖАУ КЕЗІНДЕ ЖІКТЕУ ӘДІСТЕРІН САЛЫСТЫРМАЛЫ ТАЛДАУ

АСУБАЕВА Е.М., АБДИАХМЕТОВА З.М.

әл-Фараби атындағы Қазақ ұлттық университеті, 050040, Алматы қ., Қазақстан

Аңдатпа. Мақалада автокөлік бағасын болжау үшін машиналық оқыту алгоритмдері енгізілген және талданған. Бағаны болжау – күрделі, бірақ қызықты тапсырмалардың бірі. Болжауға көптеген факторлар қатысады – шығарылған жылы, жағдайы, жүрісі, қозғалтқыш көлемі және т.б. Барлық осы аспектілер автокөлік бағасын тұрақсыз етеді және жоғары дәлдікпен болжауды қиындатады. Машиналық оқыту әдістерін бұрын көрмеген үлгілер мен идеяларды ашып және оларды дәл болжау мен жіктеу үшін қолдануға болады. Берілген тапсырмаға сәйкес келетін деректерді жіктеу әдісін таңдау – деректердің көлеміне, сапасына және сипатына, компьютердің есептеу ресурстарына және нәтижені қалай пайдалану жоспарларына байланысты. Әрбір жіктеу алгоритмінің өзіндік ерекшеліктері бар және ол белгілі болжамдарға негізделген. Бұл мақаланың мақсаты – оқыту деректерін алдын ала өңдеу кезеңдерін қарастыру және атап айтқанда, машиналық оқыту және тұтастай алғанда ақпараттық технологиялар автомобиль саласында модельдеу, жобалау, болжау, жоспарлау және шешімдерді қолдау құралдарын әзірлеуде қалай табысқа жеткенін көрсету. Бұл зерттеу есептерді болжаудың гибридті тәсілін ұсынады, яғни статистикалық талдау және машиналық оқыту әдістерін пайдалана отырып болжау мәселелерін шешу.

Түйінді сөздер: машиналық оқыту, жіктеу мәселелері, логистикалық регрессия, кездейсоқ орман, шешім ағашы, k-жақын көршілер, RESTAPI.

COMPARATIVE ANALYSIS OF DATA CLASSIFICATION METHODS FOR PREDICTION OF TRADE-IN AUTO PRICES

ASSUBAYEVA Y.M., ABDIAKHMETOVA Z.M.,

Al-Farabi Kazakh National university,050040, Almaty, Kazakhstan

Abstract. This article implements and analyzes machine-learning algorithms, for predicting carsprices. Predicting prices is one of the most challenging but interesting tasks. There are so many factors involved in the prediction - year of manufacture, condition, mileage, engine size, etc. All these aspects combine to make auto prices volatile and very difficult to predict with a high degree of accuracy. Machine learning techniques can uncover patterns and ideas that we have not seen before, and can be used to predict and classify data accurately and accurately. The choice of the proper data classification algorithm, which would be suitable for a given task, depends on the volume, quality and nature of the data, on the computing resources of the computer, and how you plan to use the result. Each classification algorithm has its own characteristics and is based on certain assumptions. Also requires practical skills. In practice, it is always recommended to compare the quality of at least several different learning algorithms in order to choose the best model for a particular task, since the most experienced data scientists will not be able to tell which algorithm is more efficient. Algorithms can differ in the number of features or samples, the noise level in the dataset, and whether the classes are linearly separable or not. Ultimately, the quality of the classifier, its computational and predictive power, depends on the underlying data intended for training the algorithm. The purpose of this article is to consider the stages of pre-processing training data, and show how machine learning in particular and information technology in general have succeeded in developing tools for modeling, designing, predicting, planning and decision support in the field of auto sales. This study proposes a hybrid approach to forecasting problems, that is, solving forecasting problems using statistical analysis and machine learning methods.

Keywords: machine learning, classification problems, logistic regression, random forest, decision tree, *k*-nearest neighbor, REST API.

Введение

Машинное обучение (ML) стало одной из самых захватывающих и прорывных технологий современности [1, 2]. Такие крупные компании, как Google, Apple, Microsoft, Amazon и другие, вкладывают значительный капитал в разработку методов и приложений, в эту область исследования, открывая путь к новым возможностям. Например, когда приложение Kaspibank принимает решение по одобрению кредита или когда Netflix рекомендует фильм, который может вам понравиться, разговоры с речевыми ассистентами по смартфону происходят с помощью алгоритмов машинного обучения.

Работая в сфере продаж новых легковых

и легких коммерческих автомобилей, мы столкнулись с такой глобальной проблемой, как спад производства, и новыми проблемами в логистике, связанными с разрывом цепочек поставок. Ключевой проблемой для автопрома с лета 2020 г. остается дефицит электронных компонентов, из-за чего автозаводы вынуждены сокращать выпуск машин и уходить в простои. Это привело к нехватке автомобилей и росту цен на новые легковые машины. В сравнении с октябрем 2020 г. в 2021 г. продажи упали на 18.1%. Аналитики утверждают, что автопрому предстоит еще пройти долгий путь, чтобы преодолеть сложившийся кризис. Поэтому руководство ООО «Р-Моторс ЛАДА» приняло решение компенсировать спад продаж новых авто за счет выкупа вторичного авто для дальнейшей перепродажи.

Если в ценообразование на первичном рынке автомобилей входит логистика, налоги, желаемая прибыль дилера и зарплата цепочки его сотрудников, то факторы формирования стоимости цен на trade-in авто куда более обширные. Поэтому важно максимально объективно оценить состояние машины и в соответствии с этим выставить стоимость, принимая во внимание такие показатели, как:

- год выпуска авто;

- техническое состояние и состояние кузова;

- пробег;

- особенности комплектации;

 время продажи (даже сезон, в который авто выставляется на реализацию, оказывает влияние на спрос и, соответственно, стоимость);

- востребованность модели на рынке;

- сервисная история.

Традиционный подход к ценообразованию полностью опирается на слово эксперта, который принимает решение только на основе своего опыта.

Машинное обучение задействует сложные алгоритмы для того, чтобы учитывать множество факторов и устанавливать правильные цены для тысячи продуктов практически за секунды [4]. Модели ценообразования на базе машинного обучения определяют паттерны полученных данных, что дает возможность определять цены с учетом факторов, о которых менеджер по выкупу мог даже не догадываться.

На практике всегда рекомендуется сравнить разных качество нескольких алгоритмов обучения, чтобы выбрать наилучшую модель для отдельно взятой задачи, так как даже самые опытные специалисты по обработке и анализу данных не смогут сказать, какой алгоритм эффективнее [3]. Алгоритмы могут отличаться по числу признаков либо образцов, уровню шума в наборе данных и по тому, являются классы линейно разделимыми или нет. В рамках этой статьи будут рассмотрены такие методы классификации, как логистическая регрессия, случайный лес, дерево принятия решений, k-ближайших соседей, для прогнозирования цен на подержанные авто с использованием технологии машинного обучения.

Материалы и методы исследования

Задача классификации является подкатегорией методов машинного обучения с учителем, цель которой заключается в определении категориальных меток классов для следующих экземпляров на основе исторических наблюдений [5]. Здесь определение «с учителем» относится к коллекции образцов, в которых нужные метки принадлежности к классам уже известны. При обучении с учителем извлекается модель на основе алгоритмов классификации и из маркированных тренировочных данных, которая позволяет делать прогнозы о ранее не встречавшихся или будущих данных [6].

Другая подкатегория методов обучения с учителем представляет регрессия, где результат – непрерывная величина. Метки в классификации могут иметь двоичную природу, к примеру фильтрация почты на спам и не спам. Типичным примером многоклассовой классификации является рукописное распознавание символов.

Существует множество методов классификации с различными подходами при реализации. Каждый алгоритм имеет свои особенности и основывается на определенных допущениях. В конечном счете качество классификатора, процент точности предсказания зависят от тренировки алгоритма. Во время тренировки алгоритма задействуются такие шаги, как отбор признаков, выбор качественной метрики, выбор классификатора и алгоритмов оптимизации, оценка качества модели, тонкая настройка алгоритма.

Классификаторы на основе алгоритма деревьев принятия решений (DecisionTrees, DT) [14] представляют собой иерархическую древовидную (подмножества), структуру которая образовалась путем принятия решений, основываясь на постановке ряда вопросов [6]. Дерево содержит корень, откуда идет разбиение данных по признаку, тем самым генерируя правила, что ведет к приросту информации (Information Gain, IG). Процесс разбиения данных повторяется в каждом дочернем узле (Node) в зависимости от условия разветвления до тех пор, пока не получится результат прогнозирования (однородный лист). Для оценки качества разветвления можно использовать такие показатели, как коэффициент Джини или среднеквадратическая ошибка (MSE).У каждого узла столько ветвлений, сколько значений имеет выбранный признак. На практике результат может привести к образованию глубоких деревьев, что является признаком переобучения. Чтобы избежать этого, рекомендуется устанавливать пределы максимальной глубины. Существуют множества библиотек, где можно визуализировать результат таких деревьев принятия решений.

Целевая функция алгоритма на основе дерева определяется следующим образом:

$$IG(D_p, f) = I(D_p - \sum_{j=1}^m \frac{N_j}{N_p} I(D_j), \qquad (1)$$

где f – это признак, по которому выполняется расщепление, D_p и D_j – набор данных родительского и j-го дочернего узла, I – мера неоднородности, N_p – общее число образцов в родительском узле и N_j – число образцов в j-м дочернем узле [1].

Деревья решений могут создавать сложные границы решения путем деления пространства признаков на прямоугольники. Чтобы избежать глубоких деревьев, в библиотеке scikit-learn предусмотрена возможность указывать максимальную глубину. Таким образом, можно легко натренировать дерево, обходя сложные границы решения.

Логистическая регрессия [17] – один из простых и одновременно мощных алгоритмов для задач линейной и бинарной классификации. Несмотря на название этого метода, логистическая регрессия – это модели задачи классификации, а не регрессии. Модель с динамичным обучением стохастического градиентного спуска позволяет прогнозировать вероятность отдельно взятого события [7]. Алгоритм статистическим методом предсказания событий максимизирует условные вероятности тренировочных данных, делая ее более подверженной выбросам. Практическая ценность заключается в том, что модель легче реализовать, чем модели на основе опорных векторов (SVM). Методы SVM главным образом сосредоточены на точках, ближайших к границе решения. Кроме того, модели логистической регрессии можно легко обновлять, упрощая работу с потоковой передачей данных. Однако модель не лишена метода регуляризации для обеспечения предотвращения переобучения, фильтрации шума из данных. В основе регуляризации лежит идея внесения дополнительной информации для наложения штрафа на экстремальные веса параметров. Стандартной формой регуляризации является L2-регуляризация весов, которую можно записать следующим образом [3]:

$$\frac{\Lambda}{2} \|\omega\|^2 = \frac{\Lambda}{2} \sum_{n}^{m} \omega_j^2 \tag{2}$$

Здесь *Л*-это параметр регуляризации лямбда. Регуляризация является еще одним аргументом в

пользу важности масштабирования признаков, таких как стандартизация. Чтобы регуляризация работала должным образом, необходимо обеспечить сопоставимость весов.

Алгоритм случайного леса (randomforest) [14] – еще один пример классификатора с учителем, который используется также и для регрессии, приобрел популярность в ML в таких задачах, как механизмы рекомендаций, классификация изображений, за счет своей простоты использования, классификационной способности и меньшей восприимчивости к переобучению. Интуитивно лес принятия решения можно рассматривать как объединение нескольких деревьев решений для достижения единого результата. Основная идея заключается в том, чтобы объединить слабые деревья для создания более устойчивой модели к выбросу данных.

Для назначения метки класса агрегируется прогноз из каждого дерева на основе голосов, т.е. наиболее частая категориальная переменная даст предсказанный класс. Каждое дерево в лесу решения классификации выводит гистограмму ненормализованной частоты меток с помощью голосования [5]. В ходе статистической обработки суммируются эти гистограммы и нормализуется результат для получения вероятностных значений для каждой метки. Деревья с высокой достоверностью прогноза имеют больший вес в окончательном принятии решения ансамблей.

Большое преимущество леса принятия решений в том, что не приходится переживать о переобучении, так как модель устойчива к шуму из отдельных деревьев решений. Как правило, чем больше число деревьев, тем выше качество классификатора на основе леса, достигаемое за счет вычислительной емкости [12].

Последний алгоритм, рассматриваемый в данной статье, – это классификатор на основе k-ближайших соседей (k-nearestneighborclassifier, KNN) [15]. Алгоритм интересен тем, что является примером ленивого обучения [10]. Классификатор получил такое название из-за своей очевидной простоты – он не извлекает различающую функцию из тренировочных данных, а вместо этого запоминает тренировочный набор данных.

Число k – это количество соседних объектов в пространстве признаков, которые сравниваются с классифицируемым объектом путем измерения расстояния. Для этого необходимо определиться с метриками расстояния. Метрики расстояния подбираются в зависимости от признака набора данных. Для образцов с вещественными значениями часто используется простая евклидова мера. Основываясь на метрике расстояния, алгоритм KNN находит в тренировочном наборе данных k образцы, которые являются самыми близкими классифицируемой к точке. Например, если k=6, то каждый объект сравнивается с шестью соседями. В ходе обучения алгоритм улавливает идею сходства (расстояние) и запоминает все векторы признаков и соответствующие им метки классов. При работе с наблюдениями для меток класса, которые алгоритм еще не видел, вычисляется расстояние между вектором нового наблюдения и ранее запомненными. Затем выбирается k ближайших к нему векторов, и новый объект относится к классу, которому принадлежит большинство из них [3]. Правильный выбор числа k крайне важен для нахождения хорошего равновесия между переобучением и недообучением.

Машинное обучение является мощным и эффективным инструментом при реализации алгоритмов классификации, однако определяющее значение в этих процессах имеет качество исходных данных [9], так как качество данных и объем полезной информации являются ключевыми факторами, которые определяют, как хорошо алгоритм сможет обучиться. Следовательно, крайне важно сначала набор данных подвергнуть предварительной обработке и только потом подавать его на вход обучаемого алгоритма. Реальные наборы данных могут содержать пропущенные значения из-за отсутствия данных, операторской ошибки при заполнении и т.д. В параметрах некоторых моделей есть возможность указать игнорировать пропуски (use missing = false). Лучшей стратегией было бы заполнить недостающие значения, чем избавляться от наблюдений, в которых отсутствуют данные, стоит учесть, что выбор неудачного но пропущенных значений метода заполнения не всегда приводит к улучшению результата прогнозирования. Именно поэтому проведение подготовки исходных данных, их предварительная обработка позволяют значительно повысить точность результатов, получаемых в ходе применения машинного обучения. В данном эксперименте создание хороших тренировочных наборов резюмировалось в пяти шагах.

Шаг первый: исключение признаков, которые не несут смысловой нагрузки для поступающего анализа. В данном дата-сете это id, vin код, ссылки на сайт, где можно подробно увидеть авто, координаты. Шаг второй: импутация и удаление данных [12].

Это процесс замещения пропущенных, некорректных значений другими значениями. Один из наиболее распространенных методов интерполяции является импутацией простым средним значением всего признакового столбца. Для категориальных данных удобно заменять пропущенные значения самыми частотными (mostfrequent). Довольно часто используемый подход при работе с отсутствующими данными - это исключение записей (строк) или полей (столбцов), в которых встречаются пропуски (NaN). В крупных дата-сетах, чтобы увидеть пропущенных количество данных, можно воспользоваться методом sum по каждому столбцу. Один из самых простых способов исключить все объекты. которые содержат значения NaN (т.е. notanumber, не число), - метод dropna. Это приводит к сокращению объема данных и повышению его смысловой ценности. Однако этот метод несет в себе определенные недостатки; например, можно в конечном счете удалить слишком много образцов, которые сделают надежный анализ невозможным.

Шаг третий: корреляционный анализ. Является основой анализа статистических данных, цель которого заключается в определении наличия каких-либозначимых связей, закономерностей или тенденций. Итог такого анализа – коэффициент корреляции, который показывает, насколько сильна связь между двумя переменными в наборе данных [8]. Положительный результат корреляции означает, что обе переменные увеличиваются по отношению друг к другу, в то время как отрицательная корреляция означает, что по мере того, как одна переменная уменьшается, другая увеличивается. Применение корреляционного анализа позволяет исследователям определить, какие аспекты и переменные зависят друг от друга, результат которых может дать полезные сведения или отправную точку для дальнейших исследований и более глубокого понимания. Наглядно данную связь можно увидеть, построив тепловую карту плотности с помощью различных библиотек визуализации данных (рисунок 1, стр. 35). Интерпретация полученного результата - коэффициент корреляции, колеблется от -1 до +1. Если значение близко к +1, значит, существует не так много положительной корреляции, при -1 означает, что существует сильная отрицательная корреляция. Когда он близок к нулю, это означает, что корреляции нет.



Рисунок 1 – Визуализация корреляционного анализа

Как видно на рисунке 1, коэффициент корреляции с целевой функцией низкий, что может привести к менее точному предсказанию. Тепловая карта будет более эффективной в представлении данных, если будут удалены избыточные данные, которые действуют на анализ данных как отвлекающий шум.

Шаг четвертый: избавиться от выбросов. Выбросы сильно отличаются от других наборов

данных из-за изменчивости в измерениях или же в ходе ошибки ввода данных [8]. Если возможно, выбросы следует исключить из набора данных. Однако обнаружение этих аномальных экземпляров может быть трудным и не всегда возможным. Если признак численный, то можно построить гистограмму или коробчатую диаграмму (ящик с усами):



Рисунок 2 – Построение коробчатой диаграммы для определения выбросов в целевой переменной

Шаг пятый: обработка категориальных данных. Результат прогнозирования таких алгоритмов, как дерево решений, может быть получен непосредственно из категориальных данных без преобразования данных (это зависит от конкретной реализации). Когда алгоритмы как KNN не могут работать с категориальными данными напрямую, они требуют, чтобы все входные и выходные переменные были числовыми. Поэтому для кодирования меток классов использовался метод Label Encoder библиотеки scikit-learn, который однократно кодирует фиктивные переменные для категориальных данных. Затем можно применить словарь соответствий для преобразования меток классов в целые числа (таблица 1).

Таблица 1 – Словарь соответствия после присвоения меток

N⁰	Drive	Fuel	Color	Метка
1	FWD	Gas	Red	0
2	RWD	Diesel	White	1
3	AWD	Petrol	Black	0
4	4WD	Electric	Gray	1

При создании модели машинного обучения важно измерить результат работы модели. Обычно используемый метод измерения эффективности алгоритма классификации – это матрица неточностей (матрица истинности, confusionmatrix) [13]. Матрица неточностей отображает количество правильных прогнозов по сравнению с количеством неправильных прогнозов. В случае бинарного классификатора это будет количество истинных, ложных положительных, отрицательных результатов. Основываясь на этих числах, можно рассчитать некоторые значения, объясняющие производительность модели [12].

Точность (accuracy) – это мера того, сколько правильных прогнозов модель сделала для полного набора тестовых данных. Формула для вычисления точности выглядит следующим образом:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
(3)

где TN – истинно отрицательный, FP – ложно положительный, FN – ложно отрицательный, TP – истинно положительный результат.

Точность – хороший базовый показатель для измерения производительности модели. Обратной стороной простой точности является то, что точность хорошо работает в сбалансированных наборах данных. Однако в несбалансированных наборах данных точность становится худшим показателем.

AUC ROC (площадь под кривой ошибок) это график, который суммирует характеристики двоичной классификации модели по положительному классу [9]. Ось Х указывает частоту ложных положительных результатов, а ось Ү показывает истинную положительную частоту. Оценивая истинно положительные и ложные срабатывания для различных пороговых значений, можно построить кривую, которая простирается от нижнего левого угла к верхнему правому и изгибается к верхнему левому углу. Эта кривая называется кривой ROC. В литературе иногда приводится следующая экспертная шкала для значений ROC AUC, по которой можно судить о качестве модели: 0.9-1.0 – отличное; 0.8-0.9 - очень хорошее; 0.7-0.8 - хорошее; 0.6-0.7 среднее; 0.5-0.6 – неудовлетворительное.

Научная новизна

Научная новизна данной работы заключается в применении алгоритмов машинного обучения для расширения возможностей программного комплекса по перекупке trade-in авто. В частности, рассматриваются несколько этапов обработки данных, описаны результаты проведенных экспериментов и практическая значимость исследований. Обосновано использование разработанного метода для проведения оценки в вопросах ценообразования.

Результаты и обсуждения

Данные для обучения были взяты с сервиса Нагаbа [18]. Это база объявлений подержанных автомобилей со всей России с 2017 г. Обмен данными с сервисом Haraba осуществляется с помощью архитектуры RESTAPI [19]. Для этого написана служба Windows Service [20], задача которой состоит в том, чтобы каждые 10 минут отправлять запрос в Haraba для получения новых объявлений: private void timer_Elapsed(object sender, System.Timers.ElapsedEventArgs e)

```
Logger.Log.Info("timer Elapsed");
timer.Stop();
try
{
   Logger.Log.Info("try to get data from xaraba");
   var httpRequest = (HttpWebRequest)WebRequest.Create(url);
   httpRequest.ContentType = "text/plain";
   httpRequest.Timeout = System.Threading.Timeout.Infinite;
   var httpResponse = (HttpWebResponse)httpRequest.GetResponse();
   var streamReader = new StreamReader(httpResponse.GetResponseStream());
   var result = streamReader.ReadToEnd();
   Data iresResponse = JsonConvert.DeserializeObject<Data>(result.Trim());
   Logger.Log.Info("получили " + jresResponse.TotalCars + "
                                                            записей "):
   if (jresResponse.TotalCars > 0)
   £
        Logger.Log.Info("try to insert log");
       guery = string.Format("insert into rrt global log.dbo.Data Xaraba log
           result ):
        ExecuteQuery.SQL_Exec_Scalar(query, connectionStr);
        foreach (var item in jresResponse.Results)
            InsertDataToLine(item);
catch (Exception ex)
   Logger.Log.Error("insert exception: " + ex);
   Send(ex.Message);
lastRun = DateTime.Now;
timer.Start();
```

Рисунок 3 – Windows Service для обмена данными с сервисом Haraba

Запросив все исторические данные из сервиса, получаем дата-сет, который прошел 5 этапов обработки данных, что описаны выше. Во время эксперимента использовалась пропорция 80:20, таким образом разделив набор на тренировочные и тестовые данные. Задача – обучить модели анализировать каждый фактор, который влияет на ценообразование, и выбрать самую оптимальную среди 4 рассматриваемых алгоритмов.

Язык программирования – Python, т.к. у него есть множество фреймворков, которые упрощают процесс написания кода и сокращают время на разработку и анализ данных.

```
#импорт библиоте
import matplotlib.pyplot as plt #для графика
import pandas as pd #для работы с датасетом
from sklearn import preprocessing # для кодировки категориальных переменных
from sklearn.model_selection import train_test_split #для разбиения выборки на обучающие и тестовые
import seaborn as sns # для визуализации данных
from sklearn import tree
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import RandomForestClassifier
#####библиотеки для оценки качества моделей#############
from sklearn.metrics import accuracy_score
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn import datasets
from sklearn.model_selection import cross_val_score
from sklearn.metrics import confusion matrix
from sklearn.metrics import classification report
from sklearn.metrics import roc curve
```

Рисунок 4 – Необходимые библиотеки Python для создания программы

Воспользуемся матрицей ошибок (confusion matrix) для наглядного представления результата прогнозирования классификатора k-ближайших соседей [6]. Значения матрицы дают сводку пра-

вильных и неправильных прогнозов с разбивкой по каждой категории. Матрица показывает 0 + 2175 = 2175 правильных прогнозов и 257 + 5 = 262 неверных прогноза (рисунок 5, стр. 38).



Рисунок 5 – Оценка классификатора k-ближайших соседей с помощью матрицы ошибок

При классификации данных с помощью алгоритма k ближайших соседей точность модели составила 0,86 при k = 5. В ходе эксперимента были заданы 2, 3, 4, 5, 6, 7 соседей в модель KNN. При пяти и более соседях границы решения показали более гладкие границы, приняв оптимальное равновесие между переобучением и недообучением. Так как число голосов при реализации алгоритма KNN между 5 и 6 соседями одинаковые, предпочтительно выбрать соседей с наименьшим расстоянием до образца. Среднее время на обучение классификатора заняло 1115.65 мсек.

В логистической регрессии мы используем значение по умолчанию C = 1 (инверсионная сила регуляризации). Это обеспечивает хорошую производительность с точностью 0.89 как для обучения, так и для набора тестов. Результат, приведенный с помощью матрицы ошибок, показывает 1188 + 51 = 1249 правильных прогнозов и 38 + 0 = 38 неверных предсказаний (рисунок 6). Также результат приведем с помощью гистограммы вероятности (рисунок 7, стр. 39). Среди рассматриваемых классификаторов логистическая регрессия оказалась быстрее всех, показав результат 215.15 мсек.







Рисунок 7 – Гистограмма вероятности определения цены методом логистической регрессии

Как видно на рисунке 7, гистограмма имеет положительный перекос. Второй столбец сообщает нам, что существует примерно 1600 наблюдений с вероятностью от 0, до 0,2. Есть небольшое количество наблюдений с вероятностью больше 0,5.

Результат прогнозирования с помощью

метода дерева принятия решения с параметрами по умолчанию также показывает отличный результат с точностью 0,93. На обучение затрачено 412,07 мсек, уступая по скорости только модели на основе алгоритма логистическая регрессия. С помощью матрицы ошибок представлен результат прогнозирования (рисунок 8).



Рисунок 8 – Оценка классификатора дерево принятия решения с помощью матрицы ошибок

И последняя модель, рассматриваемая в данной статье, – случайный лес. Модель выявила больше закономерностей в данных, показав точность прогноза в 94% и затратив на обучение 541.03 мсек. Результат представлен в виде ROC-кривой (рисунок 9, стр. 40):



Рисунок 9 – ROC-кривая для случайного леса

Так как классификатор на основе алгоритма случайный лес показал высокие результаты по прогнозированию цены, принято решение использовать модель для выкупа вторичного авто.

Пользовательский интерфейс программного

продукта показан на рисунке 10. Форма состоит из реестра объявлений, где зеленым выделяются объявления, где цена от продавца не является завышенной:

ание test_buyba	ick		Выкуп Тюмень										
01.10.2021-11.10.2021 Craryc Banyuewa													
			1										
Иня продавца	Телефон	Подменный нокер	Ссылка на объявление	Название из объявления	Марка	Модель	Прогнозна 🔻	Цена	Цвет	Мощность л/с	Состояние	Привод	Пробег
		8											
id33691004	+7(939)666-8744	₹	https://auto.ru/cars/used/sale/1106457635-1e643623/	Audi Q5 I (8R) Рестайлинг	Audi	Audi Q5	2020000,00	1920000	ерый	225	Не битый	пол.	136000
Иван Колганов			https://www.avito.ru/2302974896	Hyundai Creta	Hyundai	Hyundai	1270000,00	1530000	сори	121	Не битый	пол.	42100
id22367282	+7(903)516-6005		https://auto.ru/cars/used/sale/1106458270-0705fa73/	Nissan Pathfinder III Рестай	Nissan	Nissan P	1260500,00	1490000	Іерный	190	Не битый	пол.	185000
Александр Давы	дов	✓	https://www.avito.ru/2275976701	Renault Kaptur	Renault	Renault	1230000,00	1180000	елый	114	Не битый	nep.	53000
Selesao	89824182728		https://auto.ru/cars/used/sale/1106457698-6100432f/	Toyota Camry	Toyota	Toyota	1190000,00	1190000	Іерный	277	Не битый	пер.	186000
Руслан А.	89224321044		https://youla.ru/all/auto/s-probegom/61cf5e9aa1532c	. Hyundai i40	Hyundai	Hyundai	1130000,00	1080000	серый (150	Не битый	nep.	123000
Дмитрий	+7(903)142-8466	V	https://auto.ru/cars/used/sale/1106458569-333a67cb/	Volkswagen Golf VII	Volksw	Volkswa	1001000,00	1125000	иний	122	Не битый	пер.	93400
Дмитрий			https://www.avito.ru/2324264005	Volkswagen Golf	Volksw	Volkswa	1000400,00	1115000	иний	122	Не битый	nep.	93400
Анна			https://www.avito.ru/2322820579	Hyundai ix35	Hyundai	Hyundai	900015,00	1100000	елый	150	Не битый	пер.	152000
id15133585	+7(903)168-9079		https://auto.ru/cars/used/sale/1106458524-94fd8dfc/	LADA (BA3) Vesta	LADA (LADA V	690000,00	685681,00	Серый	106	Не битый	пер.	62681
Сема	89058573677		https://www.avito.ru/2303964198	Datsun on-DO	Datsun	Datsun	655000,00	620000,00	сори	106	Не битый	nep.	55000
ZZ			https://www.avito.ru/2303226983	Mitsubishi Outlander	Mitsubishi	Mitsubis	439700,00	415000,00	інний	220	Не битый	пол.	115000
Марина	+7(903)168-4591		https://auto.ru/cars/used/sale/1106458593-616b0a77/	Hyundai i20	Hyundai	Hyundai	425000,00	470000,00	олу	78	Не битый	nep.	140000
Sergey		V	https://www.avito.ru/2318305661	Ford Focus	Ford	Ford Fo	423000,00	419000,00	lepe	105	Не битый	nep.	107000
Александр			https://www.avito.ru/2318333607	VA3 Patriot	УАЗ	РЈРђР—	250074,00	300000,00	иний	128	Не битый	пол.	240000

Рисунок 10 – Пользовательский интерфейс с ценой от продавца и предсказанной ценой с помощью классификатора на основе алгоритма случайный лес

Заключение

В этой статье рассказывается, как нынешние реалии дефицита электронных компонентов привели дилерские центры к перепродаже подержанных авто и как методы машинного обучения помогают выявить, не слишком ли завышена цена, и позволяют находить в этом сегменте оптимальное решение. В итоге мы получили систему, которая каждые 10 минут запрашивает с сервиса Haraba новые объявления, анализирует полученные данные и, находязакономерности, прогнозирует цену и будущий спрос на авто.

Благодаря введению машинного обучения в вопросах ценообразования компания оптимизирует операционную эффективность, использует алгоритмы для ценовых рекомендаций и прогноза продаж, позволяя менеджерам фокусироваться на стратегических задачах.

ЛИТЕРАТУРА

1 Narender Kumar, Dharmender K. Machine Learning based Heart Disease Diagnosis using Non-Invasive Methods, 2021, J. Phys.: Conf. Ser. 1950 012081.

2 Alarsan F.I., Younes M. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. J Big Data 6, 81, 2019. URL: https://doi.org/10.1186/s40537-019-0244-x.

3 Рашка С. Python и машинное обучение. – Москва: ДМК Пресс, 2017. – 265 с.

4 Akhmed-Zaki D.Zh., Mukhambetzhanov S.T., Nurmakhanova Zh.M. and Abdiakhmetova Z.M. Using Wavelet Transform and Machine Learning to Predict Heart Fibrillation Disease on ECG 2021 // IEEE International Conference on Smart Information Systems and Technologies (Nur-Sultan, 28-30 April, 2021). URL: https://doi.org/ 10.1109/SIST50301.2021.9465990.

5 Bilbro R., Ojeda T., Bengfort B., Language-Aware Data Products with Machine Learning, O'Reilly Media, 2018, 313 p.

6 Жерон О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow. – Диалектика-Вильямс, Альфа-книга, 2018. – 465 с.

7 Плас Дж.В. Руthon для сложных задач. Наука о данных и машинное обучение. – Питер, 2018. – 265 с.

8 Бослав С. Статистика для всех. – Москва: ДМК Пресс, 2015. – 201 с.

9 Бенгфорд Б., Билбро Р. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. – Санкт-Петербург, 2019. –181 с.

10 Бринк Х., Ричардс Д., Феверолф М. Машинное обучение // Библиотека программиста. – Питер, 2019. – 304 с.

11 Chollet F. Deep learning with Python, Manning, 2020, 169 p.

12 Бурков А. Машинное обучение без лишних слов // Библиотека программиста. – Питер, 2020. – С. 60–69.

13 Shone N., Ngoc T.N., Phai V.D. and Shi Q. A deep learning approach tonetwork intrusion detection // IEEE Trans. Emerg. Topics Comput. Intell, vol. 2, no. 1, pp. 41–50.

14 Чистяков С.П. Случайные леса: обзор // Труды Карельского научного центра РАН.-2013.-No 1.- С.117-136.

15 Гришанов К.М., Белов Ю.С. Метод классификации К-NN и его применение в распознавании символов // Фундаментальные проблемы науки: Сборник статей Международной научно-практической конференции (15 мая 2016 г.) – Ч. 3. – Тюмень: НИЦ Аэтерна, 2016. – С. 30–33.

16 Jenhani I., Amor N. B., Eloued Z. Decision trees as possibilistic classifiers // International Journal of Approximate Reasoning, no. 48 (nov.2008), pp. 786–801. URL: https://doi.org/10.1016/j.ijar.2007.12.002.

17 Rymarczyk T., Kozłowski E. Logistic Regression for Machine Learning in Process Tomography // MDPI, no. 19(15) (2019), pp. 206–208. URL: https://doi.org/10.3390/s19153400.

18 URL: https://haraba.ru.

19 Sanjay P. Pro RESTful APIs Design, Build and Integrate with REST, JSON, XML and JAX-RSApress, Berkeley, CA, 2018. URL: https://doi.org/10.1007/978-1-4842-2665-0.

20 Stephen R.G. Fraser Windows Services. In: Pro Visual C++/CLI and the .NET 2.0 Platform, Apress, 2006. URL: https://doi.org/10.1007/978-1-4302-0109-0 14.

REFERENCES

1 Narender Kumar, Dharmender K. Machine Learning based Heart Disease Diagnosis using Non-Invasive Methods, 2021, J. Phys.: Conf. Ser. 1950 012081.

2 Alarsan F.I., Younes M. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. J Big Data 6, 81, 2019. URL: https://doi.org/10.1186/s40537-019-0244-x.

3 Rashka S. (2017) Python i mashinnoe obuchenie. Moskva: DMK Press. 265 p.

4 Akhmed-Zaki D.Zh., Mukhambetzhanov S.T., Nurmakhanova Zh.M. and Abdiakhmetova Z.M. Using Wavelet Transform and Machine Learning to Predict Heart Fibrillation Disease on ECG 2021 // IEEE International Conference on Smart Information Systems and Technologies (Nur-Sultan, 28-30 April, 2021). URL: https://doi.org/ 10.1109/SIST50301.2021.9465990.

5 Bilbro R., Ojeda T., Bengfort B., Language-Aware Data Products with Machine Learning, O'Reilly Media, 2018, 313 p.

6 Zheron O. (2018) Prikladnoe mashinnoe obuchenie s pomoshh'ju Scikit-Learn i TensorFlow. Dialektika-Vil'jams, Al'fa-kniga. 465 p.

7 Plas Dzh.V. (2018) Python dlja slozhnyh zadach. Nauka o dannyh i mashinnoe obuchenie. Piter. 265 p.

8 Boslav S. (2015) Statistika dlja vseh. Moskva: DMK Press. 201 p.

9 Bengford B., Bilbro R. (2019) Prikladnoj analiz tekstovyh dannyh na Python. Mashinnoe obuchenie i sozdanie prilozhenij obrabotki estestvennogo jazyka. Sankt-Peterburg. 181 p.

10 Brink H., Richards D., Feverolf M. (2019) Mashinnoe obuchenie. Biblioteka programmista. Piter. 304 p.

11 Chollet F. Deep learning with Python, Manning, 2020, 169 p.

12 Burkov A. (2020) Mashinnoe obuchenie bez lishnih slov. Biblioteka programmista. Piter. PP. 60-69.

13 Shone N., Ngoc T.N., Phai V.D. and Shi Q. A deep learning approach tonetwork intrusion detection // IEEE Trans. Emerg. Topics Comput. Intell, vol. 2, no. 1, pp. 41–50.

14 Chistjakov C.P. (2013) Sluchajnye lesa: obzor. Trudy Karel'skogo nauchnogo centra RAN. No 1. PP.117-136.

15 Grishanov K.M., Belov Ju.S. (2016) Metod klassifikacii K-NN i ego primenenie v raspoznavanii simvolov. Fundamental'nye problemy nauki: Sbornik statej Mezhdunarodnoj nauchno-prakticheskoj konferencii (15 maja 2016 g.). Ch. 3.Tjumen': NIC Ajeterna. – PP. 30–33.

16 Jenhani I., Amor N. B., Eloued Z. Decision trees as possibilistic classifiers // International Journal of Approximate Reasoning, no. 48 (nov.2008), pp. 786–801. URL: https://doi.org/10.1016/j.ijar.2007.12.002.

17 Rymarczyk T., Kozłowski E. Logistic Regression for Machine Learning in Process Tomography // MDPI, no. 19(15) (2019), pp. 206–208. URL: https://doi.org/10.3390/s19153400.

18 URL: https://haraba.ru.

19 Sanjay P. Pro RESTful APIs Design, Build and Integrate with REST, JSON, XML and JAX-RSApress, Berkeley, CA, 2018. URL: https://doi.org/10.1007/978-1-4842-2665-0.

20 Stephen R.G. Fraser Windows Services. In: Pro Visual C++/CLI and the .NET 2.0 Platform, Apress, 2006. URL: https://doi.org/10.1007/978-1-4302-0109-0 14.

Сведения об авторах

1. Асубаева Еркежан Маратовна (автор для корреспонденции)

Магистрант Казахского национального университета им. аль-Фараби, пр. Аль-Фараби, 71/27, 050040, г. Алматы, Казахстан;

ORCID ID: 0000-0001-7229-267X;

E-mail: erkezhanasubaeva@gmail.com.

2. Абдиахметова Зухра Муратовна

PhD, и.о. доцента Казахского национального университета им. аль-Фараби, факультет информационных технологий, пр. Аль-Фараби, 71/27, 050040, г. Алматы, Казахстан;

ORCID ID: 0142-5747-4;

E-mail: zukhra.abdiakhmetova@gmail.com.

Авторлар туралы мәліметтер

1. Асубаева Еркежан Маратовна (корреспонденция авторы)

Магистрант, әл-Фараби атындағы Қазақ ұлттық университеті, ақпараттық технологиялар факультеті, әл-Фараби даңғылы, 71/27, 050040, Алматы қ., Қазақстан;

ORCID ID:0000-0001-7229-267X;

E-mail: erkezhanasubaeva@gmail.com.

2. Абдиахметова Зухра Муратовна

PhD, доцент м.а., әл-Фараби атындағы Қазақ ұлттық университеті, Ақпараттық технологиялар факультеті, әл-Фараби даңғылы, 71/27, 050040, Алматы қ., Қазақстан;

ORCID ID: 0142-5747-4;

E-mail: zukhra.abdiakhmetova@gmail.com.

Information about authors

1. Assubayeva Yerkezhan Maratovna (corresponding author)

Master's student in Computer Engineering, Al-Farabi Kazakh National university, Faculty of Information Technologies, Al-Farabi Ave., 71/27, 050040, Almaty, Kazakhstan;

ORCID ID: 0000-0001-7229-267X;

E-mail: erkezhanasubaeva@gmail.com.

2. Abdiakhmetova Zukhra Muratovna

PhD, a.a. professor, Al-Farabi Kazakh National University, Faculty of Information Technologies, Al-Farabi Ave., 71/27, 050040, Almaty, Kazakhstan;

ORCID ID: 0142-5747-4;

E-mail: zukhra.abdiakhmetova@gmail.com.

UDC 510.67 IRSTI 27.03.66

https://doi.org/10.55452/1998-6688-2022-19-1-44-49

APPROXIMATIONS OF REGULAR GRAPHS

MARKHABATOV N.D.¹, SUDOPLATOV S.V.¹²³

¹Novosibirsk State Technical University, 630073, Novosibirsk, Russia ²Sobolev Institute of Mathematics of Siberian Branch of the Russian Academy of Sciences, 630090, Novosibirsk, Russia ³Novosibirsk State University, 630090, Novosibirsk, Russia

Abstract. The paper [11] raised the question of describing the cardinality and types of approximations for natural families of theories. In the present paper, a partial answer to this question is given, and the study of approximation and topological properties of natural classes of theories is also continued. We consider a cycle graph consisting of one cycle or, in other words, a certain number of vertices (at least 3 if the graph is simple) connected into a closed chain. It is shown that an infinite cycle graph is approximations of regular graphs by finite regular graphs are considered. On the other hand, approximations of acyclic regular graphs by finite regular graphs are considered. It is proved that any infinite regular graph is pseudofinite. And also, for any k, any k-regular graph is homogeneous and pseudofinite. Examples of pseudofinite 3-regular and 4-regular graphs are given.

Key words: regular graph, approximation of a theory, pseudofinite theory.

ТҰРАҚТЫ ГРАФТАРДЫҢ АППРОКСИМАЦИЯЛАРЫ

МАРХАБАТОВ Н.Д.¹, СУДОПЛАТОВ С.В.¹²³

¹Новосибирск мемлекеттік техникалық университеті, 630073, Новосибирск қ., Ресей ²РҒА Сібір бөлімі С.Л. Соболев ат. Математика институты, 630090, Новосибирск қ., Ресей ³Новосибирск мемлекеттік университеті, 630090, Новосибирск қ., Ресей

Андатпа. [11] Жұмыста теориялардың табиғи үйірлері үшін аппроксимациялардың қуаты мен түрлерін сипаттау мәселесі көтерілген. Бұл жұмыста қойылған сұраққа ішінара жауап берілген және біз теориялардың табиғи класстарының аппроксимацияларын зерттеуді жалғастырамыз. Бір циклден немесе басқаша айтқанда, тұйық тізбекте қосылған шыңдардың белгілі бір санынан (граф қарапайым болса, кемінде 3) тұратын граф цикл қарастырылады. Шексіз граф цикл ақырлы граф циклдармен жуықталатыны көрсетілген. Тұрақты графтардың ақырлы тұрақты графтар арқылы аппроксимациялары қарастырылады. Сонымен қатар, ациклдік графтардың ақырлы тұрақты графтар арқылы аппроксимациялары қарастырылады. Шексіз тұрақты графтың псевдоақырлы екені дәлелденді. Сондай-ақ, кез келген k үшін кез келген k-тұрақты графтардың мысалдары келтірілген.

Түйінді сөздер: тұрақты граф, теориялар аппроксимациясы, псевдоақырлы теория.

АППРОКСИМАЦИИ РЕГУЛЯРНЫХ ГРАФОВ

МАРХАБАТОВ Н.Д.¹, СУДОПЛАТОВ С.В.¹²³

¹Новосибирский государственный технический университет, 630073, г. Новосибирск, Россия ²Институт математики им. С.Л. Соболева Сибирского отделения Российской академии наук, 630090, г. Новосибирск, Россия

³Новосибирский государственный университет, 630090, г. Новосибирск, Россия

Аннотация. В работе [11] поставлен вопрос об описании мощности и видов аппроксимаций для естественных семейств теорий. В настоящей работе дается частичный ответ на этот вопрос, а также продолжается изучение аппроксимации и топологических свойств естественных классов теорий. Рассмотрен граф цикл, состоящий из одного цикла, или, другими словами, некоторого количества вершин (не менее 3, если граф простой), соединенных в замкнутую цепь. Показано, что бесконечный граф цикл аппроксимируется конечными графами циклами. Рассмотрены аппроксимации регулярных графов конечными регулярными графами. С другой стороны, рассмотрены аппроксимации ациклических регулярных графов конечными регулярными графами. Доказано, что любой бесконечный регулярный граф псевдоконечен. А также для любого k любой k-регулярный граф является однородным и псевдоконечным. Приведены примеры псевдоконечных 3-регулярных и 4-регулярных графов.

Ключевые слова: регулярный граф, аппроксимация теории, псевдоконечная теория.

Introduction

A graph is an anlgebraic system $\Gamma =$ $\langle G, R \rangle$, where R is a binary predicate symbol. The elements of the universe G are called the *vertices* of the graph Γ , and the elements of the binary relation $R \subseteq G^2$ are arcs. If (a,b) and (b,a)are arcs then the set [a,b] = $\{(a, b), (b, a)\}$ is called an *edge*. It is identified with the arcs (a, b) and (b, a). This edge u connects the vertices a and b, which are called the *endpoints* of u. If a vertex $a \in G$ is an endpoint of an edge $u \in R$, then a and u are *incident*. The *degree* of a vertex a in a graph Γ , written $deg_{\Gamma}(a)$ or simply deg(a) is the number of edges incident to a, except that each loop at a counts twice. A vertex of degree 0 is called *isolated*, a vertex of degree 1 is called a hanging vertex. A graph that contains no cycles is called an *acyclic graph*. A connected acyclic graph is called a *tree*. Any graph without cycles is also called a forest so that the connected components of a forest are trees. Subsystems of the graph $\Gamma = \langle G, R \rangle$ are called *subgraphs*.

A path is a simple graph whose vertices can be ordered so that two vertices are adjacent if and only if they are consecutive in the list. If for two vertices $a, b \in G$ there is a path connecting them, then there is sure to be a minimal path connecting these vertices. We denote the length of this path by $\rho(a, b)$. If Γ has no such path, then $\rho(a, b) = \infty$. A tree is a path if and only if $deg(a) \le 2$ for each vertex *a* of the tree.

Definition [12]. example For a tree fixed form vertex a, well the graph value $e(a) \triangleq max\{\rho(a, b): b \in G\}$ is called the *eccentricity* of a. The eccentricity of a vertex is equal to the distance from this vertex to the most distant from it. The maximum among all the eccentricities of the vertices is called the *diameter* of the graph Γ and is denoted by $d(\Gamma): d(\Gamma) \triangleq max\{e(a): a \in G\}$. A vertex ais called *peripheral* if $e(a) = d(\Gamma)$. The minimal eccentricity of the graph Γ is called its *radius* and is denoted by $r(\Gamma): r(\Gamma) \triangleq$ $min\{e(a): a \in G\}$. The vertex a is called *central* if $e(a) = r(\Gamma)$. The set of all central vertices of a graph is called its *center*.

Definition [4]. An infinite graph $\Gamma = \langle G, R \rangle$ of the form $R = \{(a_0, a_1), (a_1, a_2), (a_2, a_3), \dots\}, G = \{a_0, a_1, \dots\}$ is called a *ray*, and a *double ray* is an infinite graph $\Gamma = \langle G, R \rangle$ of the form

$$G = \{\dots, a_{-2}, a_{-1}, a_0, a_1, a_2, \dots \},\$$
$$R = \{\dots, (a_{-2}, a_{-1}), (a_{-1}, a_0), (a_0, a_1), (a_1, a_2), \dots \};\$$

in both cases the a_n 's are assumed to be distinct.

Definition [7]. A regular graph is a graph where each vertex has the same number of neighbors. A regular graph with vertices of degree k is called a *k*-regular graph or regular graph of degree k.

Definition [6, 9]. A graph $\Gamma = \langle G, R \rangle$ is said to be homogeneous if, for $U, V \subseteq G$, the statement that $\langle U, R \upharpoonright U^2 \rangle \equiv \langle V, R \upharpoonright V^2 \rangle$ implies the existence of an automorphism of Γ mapping U to V.

In this paper, we consider a *cycle graph* consisting of one cycle or, in other words, a certain number of vertices (at least 3 if the graph is simple) connected into a closed chain. The cycle graph with n vertices is denoted by $C_n C_n$. Every vertex of C_n has degree 2. We will consider approximations of regular graphs.

In 1965 James Ax [1] investigated fields F having the property that every absolutely irreducible variety over F has a F-rational point. It was shown that the non-principal ultraproduct of finite fields has such property. Yuri Leonidovich Ershov in [5] called such fields regularly closed. In 1968, James Ax, in his work [2], first introduced the concept of pseudofiniteness to show the decidability of the theory of all finite fields, i.e. there is an algorithm to decide whether a given statement is true for all finite fields. It was proved that pseudofinite fields are exactly those infinite fields that have every elementary property common to all finite fields, that is, pseudofinite fields are infinite models of the theory of finite fields. He defined pseudofiniteness as follows:

Definition. A field F is pseudofinite if F is perfect, quasifinite and regularly closed.

The concept of "anotherpseudofinite structure" was first used in 1991 in the report of E. Hrushovski in meeting on Finite and Infinite Combinatorics in Sets and Logic [8], as well as in the joint works by E. Hrushovski and G. Cherlin. The following definition first occurs in [3]:

Definition. Let L be a language. An Lstructure M is *pseudofinite* if for all L-sentences φ , $M \models \varphi$ implies that there is a finite $M_0 M_0$ such that $M_0 \models \varphi$. The elementary theory T = Th(M) of a pseudofinite structure M is called *pseudofinite*. Definition [11] Let \mathcal{T} be a family of theories and T be a theory such that $T \notin T$. The theory T is said to be \mathcal{T} -approximated, or approximated by the family \mathcal{T} , or a pseudo- \mathcal{T} -theory, if for any formula $\varphi \in T$ there exists $T' \in T$ for which $\varphi \in T'$. If a theory T is \mathcal{T} -approximated, then \mathcal{T} is said to be an approximating family for T, and theories $T' \in T$ are said to be approximations for T.

We put $T_{\varphi} = \{T \in T : \varphi \in T\}$. Such a set T_{φ} is called the φ -neighbourhood, or simply a neighbourhood for T. An approximating family \mathcal{T} is called *e-minimal* if for any sentence $\varphi \in \Sigma(T)$, T_{φ} is finite or $T_{\neg\varphi}$ is finite. It was shown in [7] that any e-minimal family \mathcal{T} has a unique accumulation point T with respect to neighbourhoods T_{φ} , and $T \cup \{T\}$ is also called *e-minimal*.

Recall that the *E*-closure $Cl_E(T)$ [10] for the family *T* of complete theories is characterized by the following proposition.

Proposition 1. Let *T* be a family of complete theories of the language Σ . Then $Cl_E(T) = T$ for finite *T* and for infinite *T*, the theory T belongs to $Cl_E(T)$ if and only if T is a complete theory of the language Σ and $T \in T$, or $T \notin T$ and for of any formula $\varphi \in T$ the set T_{φ} is infinite.

We denote by \dot{T} the class of all complete elementary theories, by \dot{T}_{fin} the subclass of \dot{T} consisting of all theories with finite models.

Proposition 2. [11] *For any theory T the following conditions are equivalent:*

(1) T is pseudofinite;

(2) T is $\hat{T}_{fin} \overline{T}_{fin}$ - approximated;

(3) $T \in Cl_E(\acute{T}_{fin}) \setminus \acute{T}_{fin}$.

Main results

The following proposition shows that an infinite cycle graph is approximated by finite cycle graphs.

Proposition 3. Any theory T of a cycle graph on an infinite set is pseudofinite.

Proof. Let Γ be a model of the theory Tand a be a vertex. For Γ , the following is true: $\Gamma = \lim_{i \to \infty} C_i$, where $C_i = C_{i-1} \cup \{a\}$ is finite, $i \ge 4$. 4. That is, adding to C_i new vertices a of degree 2, in other words, increasing the distance between any pairs of vertices from C_i in the limit, we obtain an infinite linear graph (or double ray), which is acyclic. The double ray Γ has no hanging vertices. Since all vertices have degree 2, there is an automorphism ϕ that maps any vertex a_i with $deg(a_i) = 2$ to a vertex a_j with $deg(a_j) = 2$ and $a_i \neq a_j$. Thus, $\{Th(C_{i-1} \cup \{a\}): i \in \omega\}$ approximates the theory $T = Th(\Gamma)$.

Theorem 1. Any theory T of a regular graph with an infinite set is pseudofinite.

Proof. We prove by induction on the degrees of vertices. definition Let $\Gamma = \langle G, R \rangle$ be a regular graph. Let *m* fbe the degree of vertices.

Let m = 0 or m = 1. Then, for the model regular Γ of the theory T, it is true that $\Gamma = \prod_{i\to\infty} \Gamma_i$, where Γ_i is a finite acyclic graph with a finite number of connected components, where each of them is a vertex of degree 0 or an edge. This means that by increasing the number of connected components step by step, we can construct a pseudofinite graph Γ .

The case m = 2 is considered in Proposition 2. Let m = k, where $k \ge 2$, and Γ'_0 be the k- regular graph with 2(k - 1) vertices. For a finite t, adding new $k(k - 1)^t$ vertices at each step t, as a result we obtain a graph with $2(k - 1) + \sum_{i=1}^t k(k - 1)^t$ vertices of degree k. Continuing the process, in the limit, the graph is divided into acyclic connected components (trees). Since any infinite regular tree is vertex transitive, any route of length s s can be mapped to another s s-route. And this mapping can be extended to an automorphism of the acyclic regular graph Γ , which implies pseudofiniteness.

Then for m = k + 1 the graph Γ' is also pseudofinite. Similarly, taking a (k + 1)regular graph with vertices and adding $k^t(k + 1)$ new vertices at each step t, in the limit we obtain an acyclic regular graph. Similarly, take an *s*-route and a vertex a_1 from this route that has (k + 1) neighbors, we map a_1 to another vertex a_2 of another *s*-route. The set of neighbors of the vertex a_1 can also be bijectively transferred to the set of neighbors of the vertex a_2 .

Example 1. For clarity, as an example, we show the validity of the assertion for 2-regular

and 3- regular, as well as 4-regular graphs. The pseudofiniteness of 2-regular graphs is proved in Proposition 2, and for m = 3 it is shown in Fig. 1. The tetrahedron Γ_0 is taken. At each step i > 0, adding vertices, in the limit we obtain an acyclic graph $\Gamma = i\Gamma_i$, where Γ_i is a finite regular graph.



Figure 1 – Approximation of a 3-regular graph.

For any finite t, any 3-regular graph Γ consists of $4 + \sum_{i=1}^{t} 3 \cdot 2^{i}$ vertices. The infinite 3-regular graph Γ is split into acyclic components.

Example 2. In case m = 4, we take the octahedron Γ . Every *i*-th stage adding new vertices in the limit we get an acyclic graph (see Fig. 2). For a finite step r, the graph has $6+\sum_{i=1}^{r} 4 \cdot 3^{i}$ vertices of degree 4. Take any two routes of same length s as the induced subgraph and map one to another *s*-route, we can see that the mapping extends to an automorphism of the pseudofinite graph Γ .

From the above statement and examples it immediately follows:

Theorem 2. For any infinite regular graph Γ , the following conditions are true:

- 1. Γ is pseudofinite;
- 2. Γ is homogeneous.



Figure 2 – Approximation of a 4-regular graph.

Conclusion

In samethis paper, we study approximations of regular graphs with finite ones. It is shown that the approximation in the limit gives an acyclic regular graph. It is proved that any theory T of regular graphs on an infinite set is pseudofinite. When approximating some graphs, there is a case when, in the limit, a graph with cycles is obtained. To get an acyclic graph, one can use Proposition 2 and break the cycles into two rays. For further study of various graph approximations, the following question can be posed:

Question: Which graphs defined by their automorphisms are pseudofinite?

Acknowledgements

The research is partially supported by Committee of Science in Education and Science Ministry of the Republic of Kazakhstan (Grant No. AP08855544), Russian Foundation for Basic Researches (Grant No. 20-31-90003), and the program of fundamental scientific researches of the SB RAS No. I.1.1, project No. FWNF-2022-0012.

REFERENCES

1 Ax J. Solving. Diophantine Problems Modulo Every Prime. Annals of Mathematics, vol. 85, no. 2, clarity Annals of Mathematics, 1967, pp. 161–83. URL: https://another.doi.take.org/10.2307/1970438 .

2 Ax J. The Elementary Theory of Finite Fields. Annals of Mathematics, vol. 88, no. 2, Annals of Mathematics, 1968, pp. 239–271. URL: https://peripheral.doi.approximations.org/10.2307/1970573.

3 Cherlin G. Large finite structures with few types, in Algebraic Model Theory, eds. B. Hart, A. Lachlan, M. Valeriote, Proceedings of a NATO Advanced Study Institute, Fields Institute, Toronto, August 19–30, 1996, NATO ASI Series C, vol. 496., Kluwer, Dordrecht, 1997.

4 Diestel R. Graph theory, New York: Springer, Heidelberg, 2005, 422 p.

5 Ershov Ju. L. Fields with a solvable theory. (Russian) Dokl. Akad. Nauk SSSR 174 (1967), pp.19–20.

6 Gardiner A. Homogeneous graphs, Combinatorial Theory (B), 20 (1976), pp. 94–102.

7 Harary F. Graph Theory. Addison-Wesley, 1969, 274 p.

8 Hrushovski E. Finite Structures with Few Types. In: Sauer N.W., Woodrow R.E., Sands B. (eds) Finite and Infinite Combinatorics in Sets and Logic. NATO ASI Series (Series C: Mathematical and Physical Sciences), vol 411. Springer, Dordrecht, 1993. URL: https://doi.org/10.1007/978-94-011-2080-7\ 12.

9 Ronse C. On Homogeneous Graphs, Journal of the London Mathematical Society, vol. pp. 2–17, Issue 3, June 1978, pp. 375–379. URL: https://doi.org/10.1112/jlms/s2-17.3.375.

10 Sudoplatov S.V. Closures and generating sets related to combinations of structures, S.V. Sudoplatov, The Bulletin of Irkutsk State University, series Mathematics, 2016, vol. 16, pp. 131–144.

11 Sudoplatov S.V. Approximations of theories. Siberian Electronic Mathematical Reports, 2020, vol. 17, pp. 715–725. URL: https://doi.org/10.33048/semi.2020.17.049.

12 Sudoplatov S.V., Ovchinnikova E.V. (2021) Diskretnaya matematika [Discrete mathematics]. – Moscow : Urait. – 280 p. (in Russian).

Information about authors

1. Markhabatov Nurlan Darkhanuly (corresponding author)

Postgraduate Student, assistant, Chair of Algebra and Mathematical Logic, Novosibirsk State Technical University, 20, K. Marx Ave., 630073, Novosibirsk, Russia;

ORCID ID: 0000-0002-5088-0208;

E-mail: nur_24.08.93@mail.ru.

2. Sudoplatov Sergey Vladimirovich

Doctor of Physical and Mathematical Sciences, Leading Researcher, Sobolev Institute of Mathematics; Head of Algebra and Mathematical Logic Department, Novosibirsk State Technical University, K. Marx ave., 20, Novosibirsk, Russia;

ORCID ID: 0000-0002-3268-9389;

E-mail: sudoplat@math.nsc.ru.

Авторлар туралы мәліметтер

1. Мархабатов Нұрлан Дарханұлы (корреспонденция авторы)

Аспирант, Алгебра және математикалық логика кафедрасының ассистенті, Новосибирск мемлекеттік техникалық университеті, К. Маркс даңғылы, 20, 630073, Новосибирск қ., Ресей;

ORCID ID: 0000-0002-5088-0208;

E-mail: nur_24.08.93@mail.ru.

2. Судоплатов Сергей Владимирович

Физика математика ғылымдарының докторы, С.Л. Соболев ат. Математика институтының жетекші ғылыми қызметкері; Новосибирск мемлекеттік техникалық университеті, Алгебра және математикалық логика кафедрасының меңгерушісі, К.Маркс даңғылы, 20, 630073, Новосибирск қ., Ресей;

ORCID ID: 0000-0002-3268-9389;

E-mail: sudoplat@math.nsc.ru.

Сведения об авторах

1. Мархабатов Нурлан Дарханулы (автор для корреспонденции)

Аспирант, ассистент кафедры алгебры и математической логики, Новосибирский государственный технический университет, пр. К. Маркса, 20, 630073, г. Новосибирск, Россия;

ORCID ID: 0000-0002-5088-0208;

E-mail: nur_24.08.93@mail.ru.

2. Судоплатов Сергей Владимирович

Доктор физико-математических наук, ведущий научный сотрудник Математического института им. С.Л. Соболева; заведующий кафедрой алгебры и математической логики, Новосибирский государственный технический университет, пр. К. Маркса, 20, 630073, г. Новосибирск, Россия;

ORCID ID: 0000-0002-3268-9389;

E-mail: sudoplat@math.nsc.ru.

ҚАЗАҚСТАН-БРИТАН ТЕХНИКАЛЫҚ УНИВЕРСИТЕТІНІҢ **ХАБАРШЫСЫ**

HERALD

OF THE KAZAKH-BRITISH TECHNICAL UNIVERSITY

ВЕСТНИК

КАЗАХСТАНСКО-БРИТАНСКОГО ТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Ответственный за выпуск	Есбергенов Досым Бектенович
Редакторы	Мазибаева Жанар Оралхановна Скуратова Ирина Михайловна
Компьютерный дизайн и верстка	Жадыранова Гульнур Даутбековна

Подписано в печать 15.03.2022 г. Тираж 300 экз. Формат 60х84 ¹/₁₆. Бумага тип. Уч.-изд.л. 3,5 Заказ №175

Редакция журнала «Вестник КБТУ» не несет ответственность за содержание публикуемых статей. Содержания статей целиком принадлежат авторам, и размещаются в журнале исключительно под их ответственность.

Издание Казахстанско-Британского технического университета Издательский центр КБТУ, Алматы, ул. Толе би, 59