

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ»
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МІЖНАРОДНИЙ НАУКОВО-ТЕХНІЧНИЙ ЖУРНАЛ

№ 1

2021

ЗАСНОВАНО У ЛИПНІ 2001 р.

РЕДАКЦІЙНА КОЛЕГІЯ:

Головний редактор

М.З. ЗГУРОВСЬКИЙ, акад. НАН України

Заступник головного редактора

Н.Д. ПАНКРАТОВА, чл.-кор. НАН України

Члени редколегії:

П.І. АНДОН, акад. НАН України

А.В. АНІСІМОВ, чл.-кор. НАН України

Г.-В. ВЕБЕР, проф., Турція

О.Д. ГВІШАНИ, акад. РАН, Росія

П.О. КАСЬЯНОВ, проф., д.ф.-м.н.,
Україна

В.М. КУНЦЕВИЧ, акад. НАН України

О.А. ПАВЛОВ, проф., д.т.н., Україна

Л. САКАЛАУСКАС, проф., Литва

А.М. САЛЕМ, проф., Єгипет

І.В. СЕРГІЄНКО, акад. НАН України

Х.-М. ТЕОДОРЕСКУ, акад. Румунської
Академії

Е.О. ФАЙНБЕРГ, проф., США

Я.С. ЯЦКІВ, акад. НАН України

У номері:

• **Прогресивні інформаційні технології, високопродуктивні комп'ютерні системи**

• **Проблемно і функціонально орієнтовані комп'ютерні системи та мережі**

• **Математичні методи, моделі, проблеми і технології дослідження складних систем**

• **Методи аналізу та управління системами в умовах ризику і невизначеності**

• **Нові методи в системному аналізі, інформатиці та теорії прийняття рішень**

АДРЕСА РЕДАКЦІЇ:

03056, м. Київ,

просп. Перемоги, 37, корп. 35,

ННК «ІПСА» КПІ ім. Ігоря Сікорського

Тел.: **204-81-44**; факс: **204-81-44**

E-mail: journal.iasa@gmail.com

<http://journal.iasa.kpi.ua>

NATIONAL ACADEMY OF SCIENCES OF UKRAINE
EDUCATIONAL AND SCIENTIFIC COMPLEX
«INSTITUTE FOR APPLIED SYSTEM ANALYSIS»
OF THE NATIONAL TECHNICAL UNIVERSITY OF UKRAINE
«IGOR SIKORSKY KYIV POLYTECHNIC INSTITUTE»

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

INTERNATIONAL SCIENTIFIC AND TECHNICAL JOURNAL

№ 1

2021

IT IS FOUNDED IN JULY 2001

EDITORIAL BOARD:

The editor – in – chief

M.Z. ZGUROVSKY, Academician of
NASU

Deputy editor – in – chief

N.D. PANKRATOVA, Correspondent
member of NASU

Associate editors:

F.I. ANDON, Academician of
NASU

A.V. ANISIMOV, Correspondent
member of NASU

E.A. FEINBERG, Prof., USA

A.D. GVISHIANI, Academician of RAS,
Prof., Ukraine

P.O. KASYANOV, Prof., Ukraine

V.M. KUNTHEVICH, Academician of NASU

A.A. PAVLOV, Prof., Ukraine

L. SAKALAUŠKAS, Prof., Lithuania

A.M. SALEM, Prof., Egypt

I.V. SERGIENKO, Academician of NASU

H.-N. TEODORESCU, Academician of
Romanian Academy

G.-W. WEBER, Prof., Turkey

Ya.S. YATSKIV, Academician of NASU

In the issue:

• **Progressive information technologies, high-efficiency computer systems**

• **Problem- and function-oriented computer systems and networks**

• **Mathematical methods, models, problems and technologies for complex systems research**

• **Methods of system analysis and control in conditions of risk and uncertainty**

• **New methods in system analysis, computer science and theory of decision making**

THE EDITION ADDRESS:

03056, Kyiv,
av. Peremogy, 37, building 35,
Institute for Applied System Analysis
at the Igor Sikorsky Kyiv Polytechnic Institute
Phone: **204-81-44**; Fax: **204-81-44**
E-mail: journal.iasa@gmail.com
<http://journal.iasa.kpi.ua>

Шановні читачі!

Навчально-науковий комплекс «Інститут прикладного системного аналізу» Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» видає міжнародний науково-технічний журнал

«СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ».

Журнал публікує праці теоретичного та прикладного характеру в широкому спектрі проблем, що стосуються системних досліджень та інформаційних технологій.

Провідні тематичні розділи журналу:

Теоретичні та прикладні проблеми і методи системного аналізу; теоретичні та прикладні проблеми інформатики; автоматизовані системи управління; прогресивні інформаційні технології, високопродуктивні комп'ютерні системи; проблеми прийняття рішень і управління в економічних, технічних, екологічних і соціальних системах; теоретичні та прикладні проблеми інтелектуальних систем підтримання прийняття рішень; проблемно і функціонально орієнтовані комп'ютерні системи та мережі; методи оптимізації, оптимальне управління і теорія ігор; математичні методи, моделі, проблеми і технології дослідження складних систем; методи аналізу та управління системами в умовах ризику і невизначеності; евристичні методи та алгоритми в системному аналізі та управлінні; нові методи в системному аналізі, інформатиці та теорії прийняття рішень; науково-методичні проблеми в освіті.

Головний редактор журналу — ректор Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», академік НАН України Михайло Захарович Згуровський.

Журнал «Системні дослідження та інформаційні технології» включено до переліку фахових видань ВАК України.

Журнал «Системні дослідження та інформаційні технології» входить до таких наукометричних баз даних: Scopus, EBSCO, Google Scholar, DOAJ, Index Copernicus, реферативна база даних «Україніка наукова», український реферативний журнал «Джерело», наукова періодика України.

Статті публікуються українською, російською та англійською мовами.

Журнал можна передплатити. **Наш індекс 23918.** Якщо ви не встигли передплатити журнал, його можна придбати безпосередньо в редакції за адресою: 03056, м. Київ, просп. Перемоги, 37, корп. 35.

Завідувачка редакції **С.М. Шевченко**

Редакторка **Р.М. Шульженко**

Молодша редакторка **Л.О. Тарин**

Комп'ютерна верстка, дизайн **А.А. Патіюхи**

Свідоцтво про реєстрацію КВ № 23234–13074 ПР від 22.03.2018 р.

Підписано до друку 26.03.2021. Формат 70x108 1/16. Папір офс. Гарнітура Times.

Спосіб друку – цифровий. Ум. друк. арк. 14,411. Обл.-вид. арк. 28,56. Наклад 100 пр. Зам. № 11/04

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Свідоцтво про державну реєстрацію: ДК № 5354 від 25.05.2017 р.

просп. Перемоги, 37, м. Київ, 03056.

ФОП Пилипенко Н.М., вул. Мічуріна, б. 2/7, м. Київ, 01014.

Виписка з Єдиного державного реєстру № 2 070 000 0000 0214697 від 17.05.2019 р.,

тел.: (044) 361 78 68.

Dear Readers!

Educational and Scientific Complex «Institute for Applied System Analysis» of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» is published of the international scientific and technical journal

**«SYSTEM RESEARCH AND
INFORMATION TECHNOLOGIES».**

The Journal is printing works of a theoretical and applied character on a wide spectrum of problems, connected with system researches and information technologies.

The main thematic sections of the Journal are the following:

Theoretical and applied problems and methods of system analysis; theoretical and applied problems of computer science; automated control systems; progressive information technologies, high-efficiency computer systems; decision making and control in economic, technical, ecological and social systems; theoretical and applied problems of intellectual systems for decision making support; problem- and function-oriented computer systems and networks; methods of optimization, optimum control and theory of games; mathematical methods, models, problems and technologies for complex systems research; methods of system analysis and control in conditions of risk and uncertainty; heuristic methods and algorithms in system analysis and control; new methods in system analysis, computer science and theory of decision making; scientific and methodical problems in education.

The editor-in-chief of the Journal is rector of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», academician of the NASU Michael Zaharovich Zgurovsky.

The articles to be published in the Journal in Ukrainian, Russian and English languages are accepted. Information printed in the Journal is included in the Catalogue of periodicals of Ukraine.

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

1 • 2021

ЗМІСТ

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ВИСОКОПРОДУКТИВНІ КОМП'ЮТЕРНІ СИСТЕМИ

<i>Matsuki Y., Bidyuk P.I.</i> Simulating angular momentum of gravitational field of a rotating black hole and spin momentum of gravitational waves	7
<i>Редько І.В., Яганов П.О., Зилевич М.О.</i> Редукційне концептування оракульних схем	21

ПРОБЛЕМНО І ФУНКЦІОНАЛЬНО ОРІЄНТОВАНІ КОМП'ЮТЕРНІ СИСТЕМИ ТА МЕРЕЖІ

<i>Dodonov A., Mukhin V., Zavgorodnii V., Kornaga Ya., Zavgorodnya A.</i> Method of searching for information objects in unified information space	34
<i>Globa L.S., Gvozdetska N.A., Novogrudska R.L.</i> Ontological model for data processing organization in information and communication networks	47
<i>Maslianko P.P., Sielskyi Y.P.</i> Data science — definition and structural representation	61

МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ, ПРОБЛЕМИ І ТЕХНОЛОГІЇ ДОСЛІДЖЕННЯ СКЛАДНИХ СИСТЕМ

<i>Белас О.М., Белас А.О.</i> Загальна методика прогнозування нелінійних нестационарних процесів на основі математичних моделей з використанням статистичних даних	79
<i>Концеба С.М., Ліщук Р.І., Скуртол С.Д., Родащук Г.Ю., Васильченко І.П.</i> Прогнозування обсягів виробництва сільськогосподарської продукції за допомогою методів інтелектуального аналізу даних	87
<i>Naderan M.</i> Review methods for breast cancer detection using artificial intelligence and deep learning methods	98

МЕТОДИ АНАЛІЗУ ТА УПРАВЛІННЯ СИСТЕМАМИ В УМОВАХ РИЗИКУ І НЕВИЗНАЧЕНОСТІ

<i>Горелова Г.В., Саак А.А.</i> Сценарное когнитивное моделирование тенденций развития сложной системы «молодежь, рынок труда, качество жизни»	103
<i>Volkova V.N., Leonova A.E., Loginova A.V.</i> Models for development of the information-control complex of the enterprise	121

НОВІ МЕТОДИ В СИСТЕМНОМУ АНАЛІЗІ, ІНФОРМАТИЦІ ТА ТЕОРІЇ ПРИЙНЯТТЯ РІШЕНЬ

<i>Бондаренко В.Г.</i> О некоторых статистиках фрактального броуновского движения	131
<i>Zrazhevsky G., Zrazhevskaya V.</i> Quintile regression based approach for dynamical VaR and CVaR forecasting using metalog distribution	139
Відомості про авторів	151

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

1 • 2021

CONTENT

PROGRESSIVE INFORMATION TECHNOLOGIES, HIGH-EFFICIENCY COMPUTER SYSTEMS	
<i>Matsuki Y., Bidyuk P.I.</i> Simulating angular momentum of gravitational field of a rotating black hole and spin momentum of gravitational waves	7
<i>Redko I.V., Yahanov P.O., Zylevich M.O.</i> Reduction conceptualization of oracle schemes	21
PROBLEM- AND FUNCTION-ORIENTED COMPUTER SYSTEMS AND NETWORKS	
<i>Dodonov A., Mukhin V., Zavgorodnii V., Kornaga Ya., Zavgorodnya A.</i> Method of searching for information objects in unified information space	34
<i>Globo L.S., Gvozdetska N.A., Novogradska R.L.</i> Ontological model for data processing organization in information and communication networks	47
<i>Maslianko P.P., Sielskyi Y.P.</i> Data science — definition and structural representation	61
MATHEMATICAL METHODS, MODELS, PROBLEMS AND TECHNOLOGIES FOR COMPLEX SYSTEMS RESEARCH	
<i>Belas O.M., Belas A.O.</i> General methods of forecasting nonlinear nonstationary processes based on mathematical models using statistical data	79
<i>Kontseba S.M., Lishchuk R.I., Skurtol S.D., Rodashchuk H.Yu., Vasylchenko I.P.</i> Forecasting of agricultural production volumes using methods of data mining	87
<i>Naderan M.</i> Review methods for breast cancer detection using artificial intelligence and deep learning methods	98
METHODS OF SYSTEM ANALYSIS AND CONTROL IN CONDITONS OF RISK AND UNCERTAINTY	
<i>Gorelova G.V., Saak A.A.</i> Scenario cognitive modeling of development trends of the complexity system “youth, labor market, quality of life”	103
<i>Volkova V.N., Leonova A.E., Loginova A.V.</i> Models for development of the information-control complex of the enterprise	121
NEW METHODS IN SYSTEM ANALYSIS, COMPUTER SCIENCE AND THEORY OF DECISION MAKING	
<i>Bondarenko V.G.</i> On some statistics of fractional brownian motion	131
<i>Zrazhevsky G., Zrazhevskaya V.</i> Quintile regression based approach for dynamical VaR and CVaR forecasting using metalog distribution	139
Information about the authors	151

SIMULATING ANGULAR MOMENTUM OF GRAVITATIONAL FIELD OF A ROTATING BLACK HOLE AND SPIN MOMENTUM OF GRAVITATIONAL WAVES

Y. MATSUKI, P.I. BIDYUK

Abstract. In this research, we simulated the angular momentum of gravitational field of a rotating black hole and the spin momentum of gravitational waves emitted from the black hole. At first, we calculated energy densities of the rotating gravitational field and spinning gravitational waves as the vectors, which were projected on the spherical curved surface of the gravitational field and of the gravitational waves. Then we calculated the angular momentum and the spin momentum as the vectors perpendicular to the curved surface. The earlier research by Paul Dirac, published in 1964, did not select the curved surface to calculate the motion of quantum particles; but, instead, he chose the flat surface to develop the theory of quantum mechanics. However, we pursued the simulation of the gravitational waves in spherical polar coordinates that form the spherical curved surface of the gravitational waves. As a result, we found that a set of anti-symmetric vectors described the vectors that were perpendicular to the spherical curved surface, and with these vectors we simulated the angular momentum of the rotating black hole's gravitational field and the spin momentum of gravitational waves. The obtained results describe the characteristics of the rotation of a black hole and of spinning gravitational waves.

Keywords: gravitational waves, angular momentum, curvature tensor, stress-energy tensor, black hole.introduction

INTRODUCTION

Research question

In our previous two researches [1, 2, 3], we reported that the antigravity and anti-gravitational waves appear when a black hole rotates, but we also reported that further research is needed to identify the vectors, which are perpendicular to the rotating axis of the gravitational field and the gravitational waves. Then, in this new research, in order to further investigate this problem, we assumed as if the “sin φ component” to the rotational axis of φ represents the perpendicular direction of the curved surface described by spherical polar coordinates.

Theory of movement of the curved surface in the time-space coordinates

Dirac [4] explained two types of coordinate systems that describe time and space: one is in the flat space-like surface (Fig. 1), and another is the curved space-like surface (Fig. 2). In each figure three-dimensional surfaces (S_1, S_2, S_3, S'_1 in Fig. 1

and S in Fig. 2) are placed in four-dimensional time-space, where x_0 is for time and x_1, x_2, x_3 are for the space. (Einstein's special theory of relativity is explained in Fig.1, while the general theory of relativity [5] is explained in Fig. 2) Dirac [4] described that Fig. 2 represents a three-dimensional curved surface in a four-dimensional space-time, which has the property of being everywhere space-like, and the normal (perpendicular) vector to the surface must be in the light-cone in the Fig. 2. Dirac predicted that this perpendicular movement to the curved space must be physically meaningful.

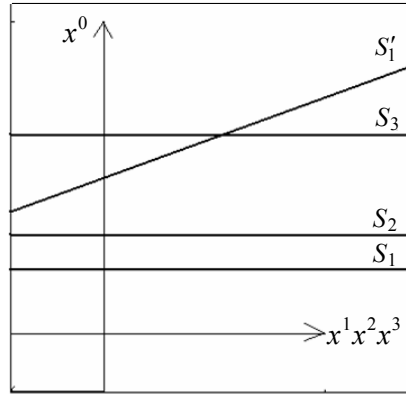


Fig. 1. Flat space-like surface (adapted from Reference [4])

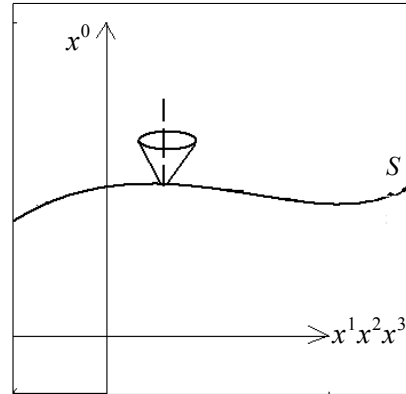


Fig. 2. Curved space like surface

In our research, we simulated the gravitational field and gravitational waves in the curved space-like surface using the spherical polar coordinate system so that we could still use the orthogonal transformation for modelling the rotation of a black hole. In our previous research [6], we simulated the energy density of the gravitational field and of the gravitational waves, also with the spin angular momentum of the gravitational waves on the flat surface; but, not the movements of the vectors perpendicular to the curved surface. In this article we report the result of our next research about the simulation of perpendicular component of the movement of the curvature tensors.

Curvature tensors

In this research, we used the same curvature tensors that we derived for our previous researches [2, 3], but we reorganized the components in the following formula for the gravitational field:

$$R_{\mu\nu} = \begin{bmatrix} R_{11} & 0 & 0 \\ 0 & R_{22} & 0 \\ 0 & 0 & R_{33} \end{bmatrix},$$

where
$$R_{11} = \frac{20}{3(\rho - \tau)^2} + \frac{11\mu}{18m(\rho - \tau)^{4/3}};$$

$$R_{22} = \frac{28}{9\mu^2(\rho - \tau)^{10/3}} + \frac{140m}{9\mu^2(\rho - \tau)^4} + \frac{4}{\sin^2 \theta} + \cot^2 \theta,$$

and
$$R_{33} = \frac{-28}{9\mu^2(\rho - \tau)^{10/3} \sin^2 \theta} + \frac{140m}{9\mu^3(\rho - \tau)^4 \sin^2 \theta} + \frac{4}{\sin^2 \theta} + \frac{11 \cot^2 \theta}{\sin^2 \theta},$$

then we formulated the body vector of the black hole: $R = [R_{11} \ R_{22} \ R_{33}]^T$.

And for the simulation of the gravitational waves, we used the following formula, as three diagonal components of a 3×3 matrix:

$$\text{For } \zeta = \eta = 1: \frac{-16}{9(\rho - \tau)^2} - \frac{2}{3(\rho - \tau)^2} - \frac{2\mu}{81m(\rho - \tau)^{10/3}} + \frac{\mu}{9m(\rho - \tau)^{7/3}} -$$

$$- \frac{\mu}{9m(\rho - \tau)^{4/3}} - \frac{2}{(\rho - \tau)^2} + \frac{224m}{9\mu^3(\rho - \tau)^4} + \frac{224m}{9\mu^3(\rho - \tau)^4 \sin^2 \theta}.$$

$$\text{For } \zeta = \eta = 2: \frac{24 \cos \theta}{\sin^4 \theta} + \frac{8}{\sin^2 \theta}, \text{ and for } \zeta = \eta = 3: \frac{64}{9(\rho - \tau)^2} - 32 \cot^2 \theta$$

and similarly we formulated the wave vector: $g = [g_1 \ g_2 \ g_3]^T$.

The curvature tensors of gravitational waves, which penetrate the boundary of a black hole [2], are:

$$g^{\zeta\eta\rho\sigma, \zeta\eta} + g_{,\sigma}^{\zeta\eta} (g_{\rho\zeta, \eta} - (1/2)g_{\zeta\eta, \rho}) + g_{,\rho}^{\zeta\eta} (g_{\sigma\zeta, \eta} - (1/2)g_{\zeta\eta, \sigma}) +$$

$$+ (1/2)g^{\rho\zeta} g_{\beta\beta} g_{\rho\zeta, \beta} g_{\eta\zeta, \beta\sigma} + (1/2)g^{\rho\zeta} g_{\beta\beta} g_{\rho\zeta, \beta\sigma} g_{\eta\zeta, \beta} + (1/2)g_{,\sigma}^{\rho\zeta} g_{\beta\beta} g_{\rho\zeta, \beta} g_{\eta\zeta, \beta} +$$

$$+ (1/2)g^{\rho\zeta} g_{\beta\beta, \sigma} g_{\rho\zeta, \beta} g_{\eta\zeta, \beta} + (1/2)g^{\sigma\zeta} g_{\beta\beta} g_{\sigma\zeta, \beta} g_{\zeta\eta, \beta\rho} + (1/2)g^{\sigma\zeta} g_{\beta\beta} g_{\sigma\zeta, \beta\rho} g_{\zeta\eta, \beta} +$$

$$+ (1/2)g_{,\rho}^{\sigma\zeta} g_{\beta\beta} g_{\sigma\zeta, \beta} g_{\eta\zeta, \beta} + (1/2)g^{\sigma\zeta} g_{\beta\beta, \rho} g_{\sigma\zeta, \beta} g_{\eta\zeta, \beta}.$$

Distortion of time and space in strong gravity

We used the same assumption of our previous research [3] for simulating the distortion of time and space, as shown in Fig. 3 and Fig. 4. In these figures, τ is a relative time in the coordinate system, which expands and shrinks depending on the distance r , where $\tau = t + f(r)$; and ρ is the relative distance, which expands and shrinks depending on the time t , where $\rho = t + g(r)$; and $f(r)$ and $g(r)$ are functions of r . For the simulation we assumed Case-1: $f(r) = \log r$; and $g(r) = e^r$ (non-linear); and Case-2: $f(r) = 1/r$, and $g(r) = r$ (linear).

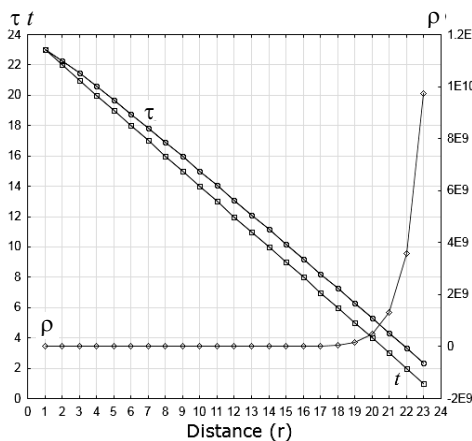


Fig. 3. Time and distance from the center of the gravitational field, Case-1 (non-linear distortion): $f(r) = \log r$ and $g(r) = e^r$

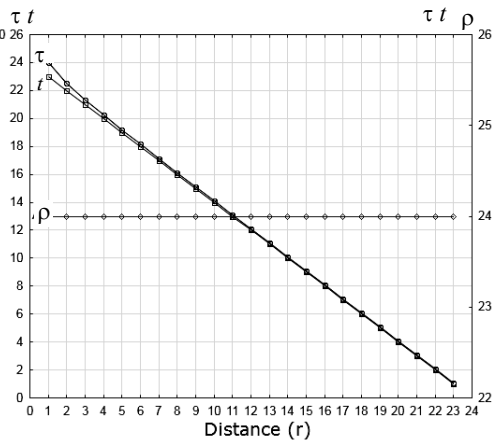


Fig. 4. Time and distance from the center of the gravitational field, Case-2 (linear distortion): $f(r) = (1/4)$ and $g(r) = r$

ALGORITHM

We used the same algorithm that we used for our previous research [3] to simulate the relative strengths (intensities) of the curvature tensors, which are reflected by the stress-energy tensor that is placed at the end of the distance r in Fig. 3 and Fig. 4.

Einstein’s field equation [5] that rules the motion of particles in the gravitational field is as follows: $(R^{\mu\nu} - (1/2)g^{\mu\nu}R)_{,\nu} = 0$. Then $R_{\mu\nu} - (1/2)g_{\mu\nu}R = kT$, where T is the stress-energy tensor and k is a constant [7]. By calculating, c and $V(c)$, as shown below, we estimated the relative strength of each component of $R_{\mu\nu}$ to the stress-energy tensor in the system of spherical polar coordinates:

$$H = kT - R_{\mu\nu} = kT - (c_1X_1 + c_2X_2 + c_3X_3),$$

and

$$H^2 = \{kT - (c_1X_1 + c_2X_2 + c_3X_3)\}^2,$$

where $c_1, c_2,$ and c_3 are the coefficients, which make a column vector c . And $X = [X_1 \ X_2 \ X_3]$, then $H = kT - Xc$. Then we set the constraint $X'H = 0$, then $X'(kT - Xc) = 0$, where X' is transpose matrix of X .

Then $X'Xc = X'kT$, $c = (X'X)^{-1}X'kT$ and $\Sigma = V(c) = \hat{\sigma}^2(X'X)^{-1}$, where $V(c) = \sigma^2$ is the variance of the c and $\hat{\sigma}^2 = e'e/(n-l)$, where $e = MkT$, $M = I - X(X'X)^{-1}X'$, n is the number of rows of each column of X (in this simulation $n = 23$), l is the number of columns of X , I is a 23×23 unit matrix that holds 1 on all diagonal elements and 0 for the other elements, $(X'X)^{-1}$ is the inverse matrix of $X'X$, and e' is the transpose vector of e .

Rotation of the black hole (an object), which contains strong gravity that distorts time and space

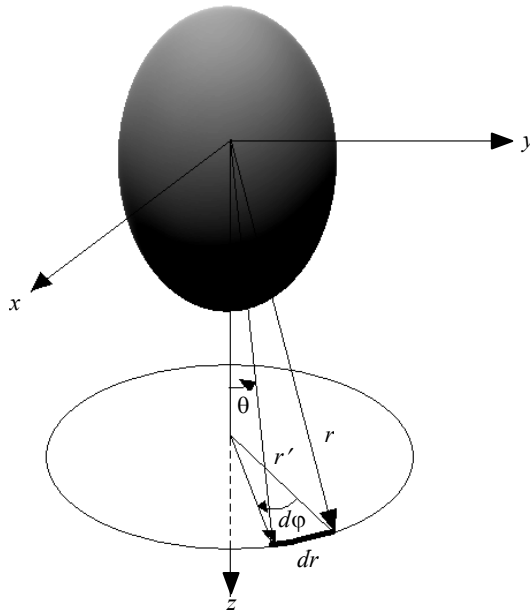


Fig. 5. Rotation of an object

When an object rotates as shown in Fig. 5, its coordinate system will be transformed by the transformation matrix D of the Euler’s angles [7]. For the rotation around one axis of φ the tensors of the object’s coordinate system will be multiplied by the matrix

$$D = \begin{bmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

And then the curvature tensor $R_{\mu\nu}$ will be transformed to:

$$DR_{\mu\nu} = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} R_{11} & 0 & 0 \\ 0 & R_{22} & 0 \\ 0 & 0 & R_{33} \end{bmatrix} = \begin{bmatrix} \cos\varphi \cdot R_{11} & \sin\varphi \cdot R_{22} & 0 \\ -\sin\varphi \cdot R_{11} & \cos\varphi \cdot R_{22} & 0 \\ 0 & 0 & R_{33} \end{bmatrix}.$$

Here the components $\sin\varphi \cdot R_{22}$ and $-\sin\varphi \cdot R_{11}$ are anti-symmetrical, which are perpendicular to the rotational axis $z = x_3$ for φ of Fig. 5.

Given the above transformed curvature tensor after the rotation, at first we calculated the relative strength of the principal moment of the rotation by the diagonal components of $DR_{\mu\nu}$, which are

$$\begin{bmatrix} dR_{11} & 0 & 0 \\ 0 & dR_{22} & 0 \\ 0 & 0 & R_{33} \end{bmatrix} = \begin{bmatrix} \cos\varphi \cdot R_{11} & 0 & 0 \\ 0 & \cos\varphi \cdot R_{22} & 0 \\ 0 & 0 & R_{33} \end{bmatrix}$$

to formulate $H = kT - (c_1 dR_{11} + c_2 dR_{22} + c_3 R_{33})$, then the algorithm follows as explained above.

And then we also calculated the anti-symmetrical components of $DR_{\mu\nu}$, which are

$$\begin{bmatrix} 0 & \sin\varphi \cdot R_{22} & 0 \\ -\sin\varphi \cdot R_{11} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & R_{22}d\Omega_3 & 0 \\ -R_{11}d\Omega_3 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

to calculate $\begin{bmatrix} dR_{11} \\ dR_{22} \\ dR_{33} \end{bmatrix} = \begin{bmatrix} R_{22} \cdot d\Omega_3 \\ -R_{11} \cdot d\Omega_3 \\ 0 \end{bmatrix}$ to formulate $H = kT - (c_1 \cdot R_{22} \cdot d\Omega_3 -$

$-c_2 \cdot R_{11} \cdot d\Omega_3)$, then the same algorithm follows as explained above.

Here, $\begin{bmatrix} 0 & d\Omega_3 & 0 \\ -d\Omega_3 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \varepsilon$ is an infinitesimal rotation operator; while in general;

$\varepsilon = \begin{bmatrix} 0 & d\Omega_3 & -d\Omega_2 \\ -d\Omega_3 & 0 & d\Omega_1 \\ d\Omega_2 & -d\Omega_1 & 0 \end{bmatrix}$, according to Reference [7]; but in this our simula-

tion $d\Omega_1 = d\Omega_2 = 0$, and $d\Omega_3 = \sin\varphi$. It calculates rotated vector as the cross-product of $R_{\mu\nu}$ and $d\Omega$,

$$\begin{bmatrix} dR_{11} \\ dR_{22} \\ dR_{33} \end{bmatrix} = R_{\nu\nu} d\Omega = \begin{bmatrix} R_{11} \\ R_{22} \\ R_{33} \end{bmatrix} \begin{bmatrix} d\Omega_1 \\ d\Omega_2 \\ d\Omega_3 \end{bmatrix} = \begin{bmatrix} R_{22}d\Omega_3 - R_{33}d\Omega_2 \\ R_{33}d\Omega_1 - R_{11}d\Omega_3 \\ R_{11}d\Omega_2 - R_{22}d\Omega_1 \end{bmatrix}.$$

For the simulation of gravitational waves, simply R_{11} , R_{22} and R_{33} , are replaced by the diagonal components of the gravitational waves, and henceforward, the same algorithm follows.

SIMULATION

Input data

Time t is set as shown in Fig. 3 for Case-1 and in Fig. 4 for Case-2, with which its slope to the distance r from the center of the gravitational field is a constant. For simulating the spatial expansion of the gravitational field we assumed, as if θ becomes larger in far distance. On the other hand, for simulating the flow of gravitational waves we assumed, that θ becomes smaller in far distance, as shown in Fig. 6. For simulating the rotation of the object we set two cases assuming φ_1 (the rotation1) and φ_2 (the rotation 2) also as shown in Fig. 6. With these settings $\sin \theta$, $\cos \theta$, $\cot \theta$ and $\cos \varphi$ behave like it is shown in Fig. 7.

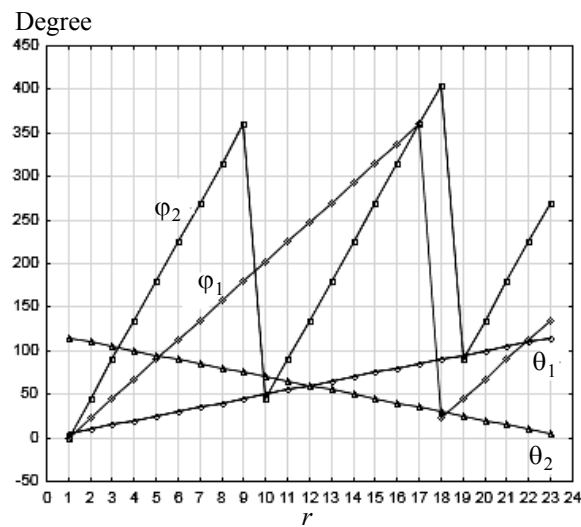


Fig. 6. Angles θ and φ for simulating gravitational field and gravitational waves

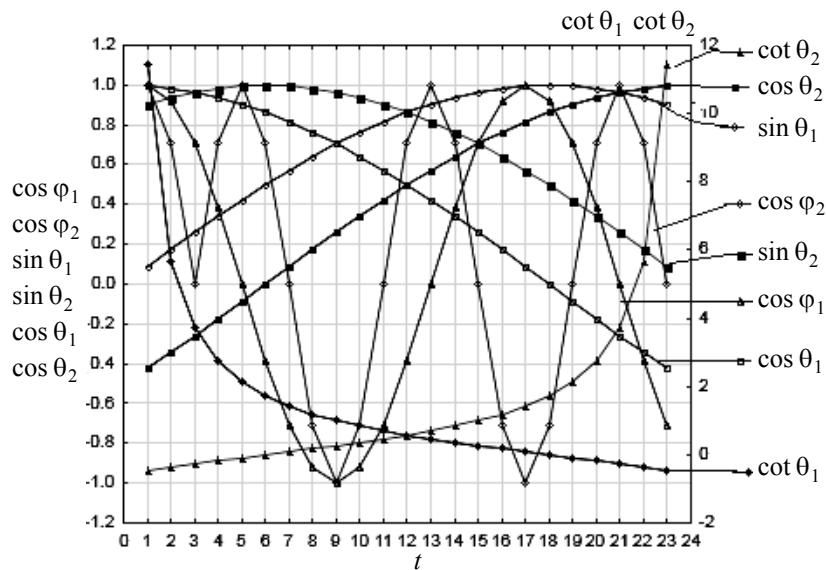


Fig. 7. $\sin \theta$, $\cos \theta$, $\cot \theta$, and $\cos \varphi$ of the simulated gravitational field

In addition, for this simulation we set the stress-energy tensor $kT=1$; because, the purpose of this simulation is to measure the order of magnitude of the relative strength of each component of $R_{\mu\nu}$ and the gravitational waves to the stress-energy tensor.

RESULT

Gravitational field

Fig. 8 (Table 1) shows the relative strengths of the gravitational field of the black hole, which are the gravitational field energy, projected on the spherical curved surface, and the angular momentum on the perpendicular vector to the surface, in Case-1 (non-linear distortion of the time and space) and Case-2 (linear distortion of time and space). As the rotation becomes more frequent from the rotation 1 to the rotation 2, the angular momentum (the perpendicular vector) changes from positive to negative. It means that the direction of the angular momentum changes, depending on the frequency of the rotation of the black hole. On the other hand, on the curved surface the gravitational field energy is negative (gravity) before the rotation in Case-1, but it changes to positive (antigravity) in the rotation 1, and then to negative (gravity) again in the rotation 2. It means that the antigravity appears, depending on the frequency of the rotation of the black hole. In Case-2, the gravitational field energy is positive (but smaller than in Case-1, and closer to zero) when the black hole doesn't rotate; while the gravitational field energy (negative) becomes larger when the black hole rotates faster. The angular momentum of Case-2 changes as it changes in Case-1.

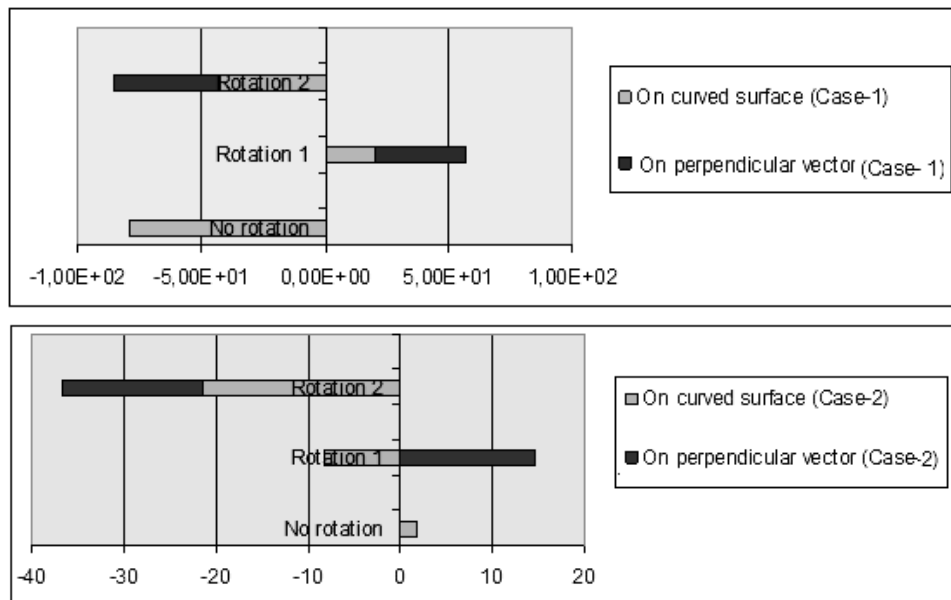


Fig. 8. Gravitational fields on curved surface and on perpendicular direction from the surface

Fig. 9 (Table 2) and Fig. 10 (Table 3) show the strengths of gravitational field energy, projected on the spherical curved surface in Case-1 and in Case-2, in

3 directions (projected on the spherical curved surface in the coordinates of r , θ , and φ , which are the generalized coordinates of $x_1 = x$, $x_2 = y$, and $x_3 = z$ of Cartesian coordinate system). In Case-1, only the r direction appears on the surface in all cases (no rotation, the rotation 1 and the rotation 2); while in Case-2, in addition to the direction of r , the component of θ appears as the black hole rotates.

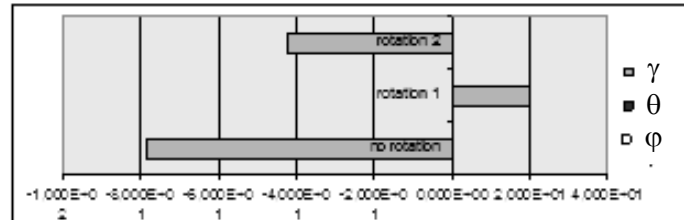


Fig. 9. Gravitational field energy in 3 directions on the spherical curved surface, Case-1

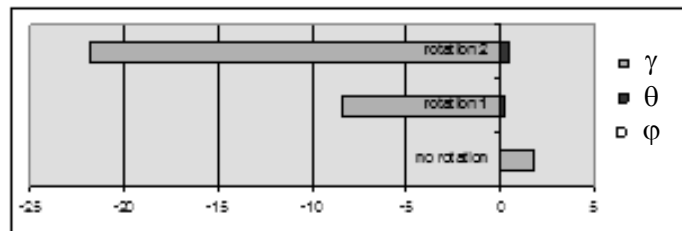


Fig. 10. Gravitational field energy in 3 directions on the spherical curved surface, Case-2

Fig. 11 (Table 4) and Fig. 12 (Table 5) show the strengths of the rotation's angular momentum in two directions (r and θ , which are perpendicular to the rotating axis, φ ($x_3 = z$ of Fig. 5)). Similar to the gravitational field energy, only the vector's component of r appears in Case-1; while the vector's component of θ also appears in Case-2. In both Case-1 and Case-2, as the frequency of the rotation increases from the rotation 1 to the rotation 2, the direction of the momentum changes from plus to minus. It suggests that the rotation of a black hole reverses its direction of the momentum when the frequency of the rotation changes.

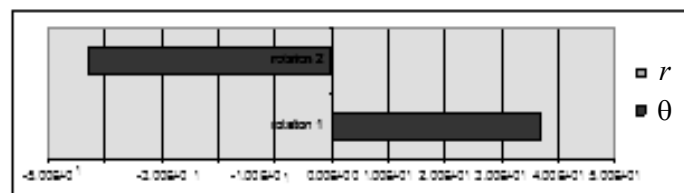


Fig. 11. Rotation momentum of the gravitational field in two directions of r and θ , Case-1

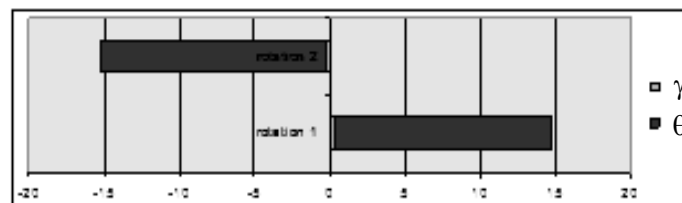


Fig. 12. Rotation momentum of the gravitational field in two directions of r and θ , Case-2

Gravitational waves

Fig 13 (Table 6) shows the strength of the gravitational waves emitted from the black hole, which are the energy (projected on the spherical curved surface of the wave) and the spin momentum of the rotation (projected on the perpendicular direction to the surface), in Case-1 and Case-2. On spherical curved surface the energy of gravitational waves are not affected by the rotation; while on the perpendicular direction the rotational momentum (spin) appears, and it changes its direction from positive to negative when the frequency of the rotation changes from the rotation 1 to the rotation 2. It suggests, that the gravitationa waves make spin as the waves move on the direction of r as the emitter (the black hole) rotates, and it changes its spinning direction when the frequency of the rotation changes from the rotation 1 to the rotation 2.

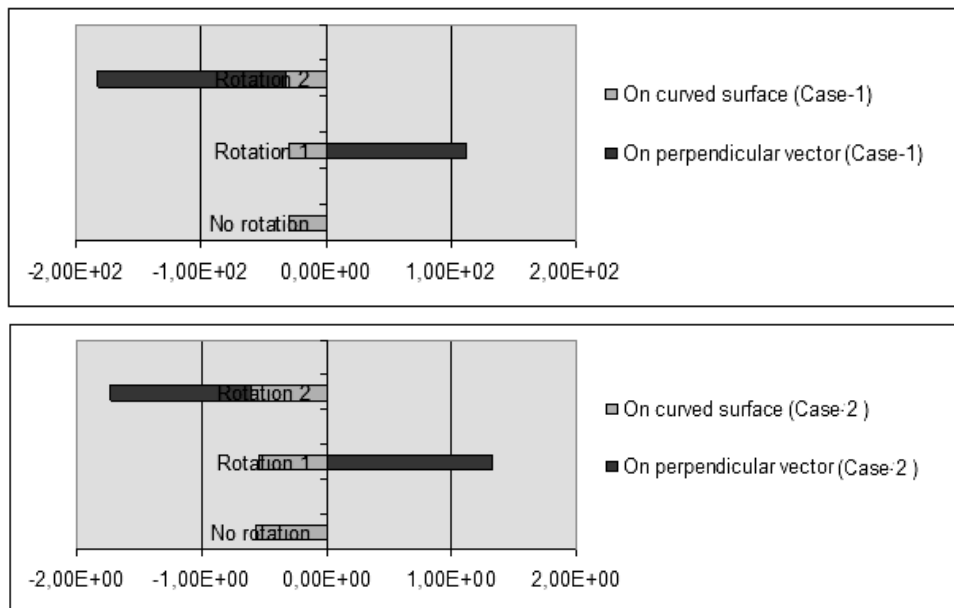


Fig. 13. Gravitational waves on curved surface and on perpendicular direction from the surface

Fig. 14 (Table 7) and Fig. 15 (Table 8) show the energy density of gravitational waves, projected on the spherical curved surface in three directions in Case-1 and Case-2. These figures suggest that the waves have negative energy density on the direction of r in Case-1; while the negative energy density appears also on the direction of rotation φ in Case-2. These figures suggest the appearance of anti-gravitational waves on the spherical curved surface. (The anti-gravitational waves must have negative sign [1], while gravitational waves must have positive sign). This finding is different from our previous report [3], in which either the gravitational waves or the anti-gravitational waves didn't appear when the black hole didn't rotate. The difference came from the configuration in the algorithm to formulate $H = kT - (c_1X_1 + c_2X_2 + c_3X_3)$. In this new research we reorganized the components of the curvature tensor into three vectors, X_1 , X_2 and X_3 ; while in our previous report [3] we calculated the relative strength of every component of the curvature tensor by, $H = kT - (c_1X_1 + c_2X_2 + \dots + c_nX_n)$.

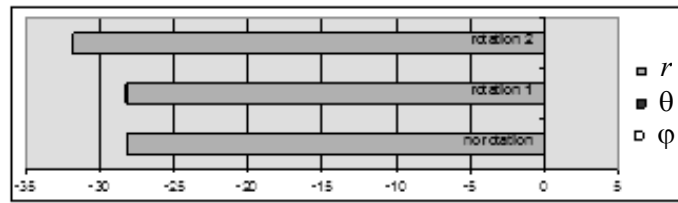


Fig. 14. Gravitational waves energy density on the curved surface, Case-1

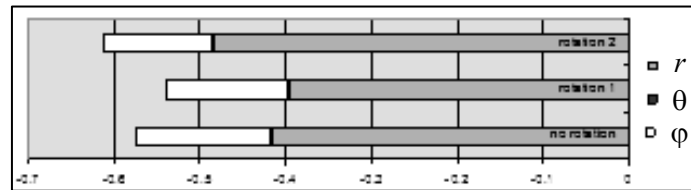


Fig. 15. Gravitational waves energy density on the curved surface, Case-2

Fig. 16 (Table 9) and Fig. 17 (Table 10) show the spin momentum of gravitational waves, projected in the directions of r and θ in Case-1 and Case-2. In

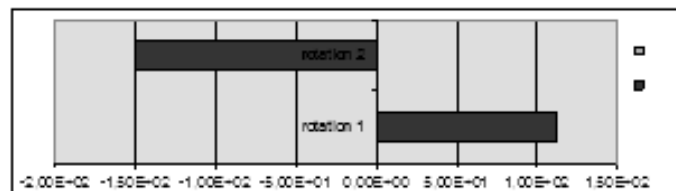


Fig. 16. Spin momentum of gravitational waves in two directions of r and θ , Case-1

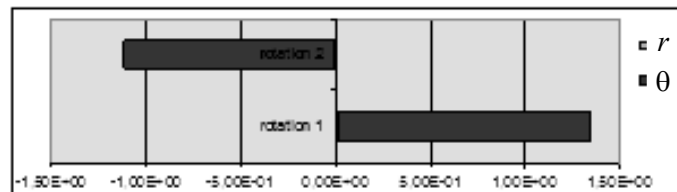


Fig. 17. Spin momentum of gravitational waves in two directions of r and θ , Case-2

both cases the vector component of θ appears in positive direction in the rotation 1, and in the negative direction in the rotation 2. These figures suggest that the gravitational waves changes its direction of spin as the rotation's speed of the black hole changes. its dsrection of spin as the rotation's speed of the black hole changes

PHYSICAL MEANING OF THE RESULTS

The components of $\cos\varphi$ in this analysis, which are the projections of the gravitational field and the gravitational waves on the spherical curved surface, are comparable to the vector component on the curved surface shown in Fig. 2 (at the beginning of this article). And the movements of these vectors are the movement of the curved surface itself. On the other hand, the vector components of $\sin\varphi$ are perpendicular to the curved surface, which is shown in the light-cone of Fig. 2, and the movement perpendicular to the curved surface must have real physical meaning. However, the earlier research by Paul Dirac [4] reported that it was problematic to quantize the movement of a quantum particle (gravitation is one of

them) such as calculating its momentum in the direction of the vectors, perpendicular to the curved surface. Henceforward, the theory of quantum mechanics was not developed on the curved surface (Fig. 2), but on the flat surface (Fig. 1).

In this research we simulated the gravitational waves on the surface of spherical polar coordinates as a surrogate of the general curved surface. And we used the cross product of anti-symmetrical vectors for simulating the components of $\sin \varphi$ as the mathematical model of the momentum in the light-cone of the general curved surface (Fig. 2).

CONCLUSIONS AND RECOMMENDATIONS

In this research we investigated the angular momentum of the gravitational field and the spin momentum of the gravitational waves by simulating the perpendicular component that is the “ $\sin \varphi$ component” to the rotational axis of φ . We used the system of the spherical polar coordinates so that we could simulate the rotation with the orthogonal transformation of Euler’s angles.

The result of the simulation shows that the rotating black hole can produce the antigravity and anti-gravitational waves; and the gravitational waves changes their spinning direction as the frequency of the black hole’s rotation changes. These findings are consistent with our previous researches: [1] in which we reported that the negative flow of gravitational waves must have the clockwise spin, while the positive flow has the counter-clockwise spin; and [3] in which we reported that the antigravity and anti-gravitational waves appear when the black hole rotates.

In this research we used the system of spherical polar coordinates as the surrogate of the general curved surface; however, in near future, the developed computer technologies must increase a possibility of using general curved surface of Einstein’s gravitational field equation also for solving the equation of motion of quantum particles.

Table 1. Strengths of gravitational field

Case	Case-1		Case-2	
	On curved surface	On perpendicular vector	On curved surface	On perpendicular vector
No rotation	-78,55	—	1,770	—
Rotation 1	20,00	36,90	-8,178	14,77
Rotation 2	-41,96	-43,00	-21,31	-15,29

Table 2. Strength of gravitational field on principal axis z , Case-1

Diagonal Components of $R_{\mu\nu}$	C and $\sqrt{V(c)}$ of $R_{\mu\nu}$ before the rotation	Diagonal Components of rotated $R_{\mu\nu}$	C and $\sqrt{V(c)}$ (Rotation 1)	C and $\sqrt{V(c)}$ (Rotation 2)
R_{11}	-78,68 (26,49)	$\cos \varphi \cdot R_{11}$	20,01 (58,22)	-42,05 (52,30)
R_{22}	0,1307 ($3,369 \cdot 10^{-2}$)	$\cos \varphi \cdot R_{22}$	$-5,516 \cdot 10^{-3}$ ($8,030 \cdot 10^{-2}$)	$8,903 \cdot 10^{-2}$ ($6,829 \cdot 10^{-2}$)
R_{33}	$-6,803 \cdot 10^{-5}$ ($2,557 \cdot 10^{-4}$)	R_{33}	$-3,290 \cdot 10^{-4}$ ($3,869 \cdot 10^{-4}$)	$-2,990 \cdot 10^{-4}$ ($5,403 \cdot 10^{-4}$)s
The values in the blackets are $\sqrt{V(c)}$				

Table 3. Strength of gravitational field on principal axis z , Case-2

Diagonal Components of $R_{\mu\nu}$	C and $\sqrt{V(c)}$ of $R_{\mu\nu}$ before the rotation	Diagonal Components of rotated $R_{\mu\nu}$	C and $\sqrt{V(c)}$ (Rotation 1)	C and $\sqrt{V(c)}$ (Rotation 2)
R_{11}	1,767 (7,364)	$\cos \varphi \cdot R_{11}$	-8,368 (11,85)	-21,79108 (16,36)
R_{22}	$2,862 \cdot 10^{-3}$ (0,1469)	$\cos \varphi \cdot R_{22}$	0,1924 (0,2427)	0,48490 (0,3369)
R_{33}	$1,110 \cdot 10^{-5}$ ($2,224 \cdot 10^{-3}$)	R_{33}	$-2,854 \cdot 10^{-3}$ ($3,673 \cdot 10^{-3}$)	$-7,281 \cdot 10^{-3}$ ($5,099 \cdot 10^{-3}$)

Table 4. Strength of the perpendicular vector to the principal axis z of gravitational field, Case-1

Rotation	C and $\sqrt{V(c)}$ (Rotation 1)	C and $\sqrt{V(c)}$ (Rotation 2)
$dx_1 = R_{22} \cdot d\Omega_3 =$ $= \sin \varphi \cdot R_{22}$	$9,077 \cdot 10^{-2}$ ($5,072 \cdot 10^{-2}$)	$-4,816 \cdot 10^{-2}$ ($5,931 \cdot 10^{-2}$)
$dx_2 = -R_{11} \cdot d\Omega_3 =$ $= -\sin \varphi \cdot R_{11}$	36,83 (46,33)	-42,94 (45,44)

Table 5. Strength of the perpendicular vector of the rotating gravitational field, Case-2

Rotation	C and $\sqrt{V(c)}$ (Rotation 1)	C and $\sqrt{V(c)}$ (Rotation 2)
$dx_1 = R_{22} \cdot d\Omega_3 =$ $= \sin \varphi \cdot R_{22}$	0,28213 (0,2621)	-0,22852 (0,2597)
$dx_2 = -R_{11} \cdot d\Omega_3 =$ $= -\sin \varphi \cdot R_{11}$	14,48 (16,65)	-15,06 (16,69)

Table 6. Strengths of gravitational waves

Case	Case-1		Case-2	
	On curved surface	On perpendicular vector	On curved surface	On perpendicular vector
No rotation	-28,13	—	-0,5738	—
Rotation 1	-28,21	$1,125 \cdot 10^2$	-0,5396	1,341
Rotation 2	-31,84	$-1,505 \cdot 10^2$	-0,6123	-1,113

Table 7. Strength of gravitational waves, Case-1

Components of gravitational tensor	C and $\sqrt{V(c)}$ of $R_{\mu\nu}$ before the rotation	C and $\sqrt{V(c)}$ (After Rotation 1)	C and $\sqrt{V(c)}$ (After Rotation 2)
$x = x_1$ component	-28,06 (14,20)	-28,15 (15,52)	-31,84 (19,70)
$y = x_2$ component	$-4,594 \cdot 10^{-4}$ ($2,398 \cdot 10^{-4}$)	$5,535 \cdot 10^{-4}$ ($2,987 \cdot 10^{-4}$)	$1,167 \cdot 10^{-3}$ ($1,181 \cdot 10^{-3}$)
$z = x_3$ component	$-6,855 \cdot 10^{-2}$ ($3,083 \cdot 10^{-2}$)	$-5,955 \cdot 10^{-2}$ ($2,711 \cdot 10^{-2}$)	$-8,950 \cdot 10^{-3}$ ($7,179 \cdot 10^{-3}$)

Table 8. Strength of gravitational waves. Case-2

Components of gravitational tensor	C and $\sqrt{V(c)}$ of $R_{\mu\nu}$ before the rotation	C and $\sqrt{V(c)}$ (After Rotation 1)	C and $\sqrt{V(c)}$ (After Rotation 2)
$x = x_1$ component	-0,4164 (0,2513)	-0,3967 (0,2734)	-0,4850 (0,3506)
$y = x_2$ component	$-1,116 \cdot 10^{-3}$ ($4,614 \cdot 10^{-4}$)	$-1,009 \cdot 10^{-3}$ ($4,484 \cdot 10^{-4}$)	$-8,918 \cdot 10^{-4}$ ($4,123 \cdot 10^{-4}$)
$z = x_3$ component	-0,1562 ($6,097 \cdot 10^{-2}$)	-0,1419 ($5,927 \cdot 10^{-2}$)	-0,1264 ($5,444 \cdot 10^{-2}$)

Table 9. Strength of the perpendicular vector to the principal axis z of gravitational waves. Case-1

Rotation	C and $\sqrt{V(c)}$ (Rotation 1)	C and $\sqrt{V(c)}$ (Rotation 2)
$dx_1 = R_{22} \cdot d\Omega_3 =$ $= \sin \varphi \cdot R_{22}$	$-9,438 \cdot 10^{-4}$ ($5,800 \cdot 10^{-3}$)	$6,081 \cdot 10^{-3}$ ($3,813 \cdot 10^{-3}$)
$dx_2 = -R_{11} \cdot d\Omega_3 =$ $= -\sin \varphi \cdot R_{11}$	$1,125 \cdot 10^2$ ($1,426 \cdot 10^2$)	$-1,505 \cdot 10^2$ ($1,027 \cdot 10^2$)

Table 10. Strength of the perpendicular vector of the rotating gravitational waves. Case-2

Rotation	C and $\sqrt{V(c)}$ (Rotation 1)	C and $\sqrt{V(c)}$ (Rotation 2)
$dx_1 = R_{22} \cdot d\Omega_3 =$ $= \sin \varphi \cdot R_{22}$	$1,025 \cdot 10^{-2}$ ($5,865 \cdot 10^{-3}$)	$-5,437 \cdot 10^{-3}$ ($4,966 \cdot 10^{-3}$)
$dx_2 = -R_{11} \cdot d\Omega_3 =$ $= -\sin \varphi \cdot R_{11}$	1,33031 (1,053)	-1,10749 (0,8583)

REFERENCES

1. Y. Matsuki and P.I. Bidyuk, "Analysis of negative flow of gravitational waves", *System Research & Information Technology*, no. 4, pp. 7–18, 2019.
2. Y. Matsuki and P.I. Bidyuk, "Numerical simulation of gravitational waves from a black hole, using curvature tensors", *System Research & Information Technology*, no. 1, pp. 54–67, 2020.
3. Y. Matsuki and P.I. Bidyuk, "Simulating the rotation of a black hole and antigravity", *System Research & Information Technology*, no. 3, pp. 124–137, 2020.
4. P.A.M. Dirac, *Lectures on quantum mechanics*; originally published by the Belfer Graduate School of Science. Yeshiva University, New York, 1964, 90 p.
5. P.A.M. Dirac, *General Theory of Relativity*. Florida University, A Wiley-Interscience Publication, John Wiley & Sons, New York, 1975, 70 p.
6. Y. Matsuki and P.I. Bidyuk, "Calculating energy density and spin momentum density of moon's gravitational waves in rectilinear coordinates", *System Research & Information Technology*, no. 3, pp. 7–17, 2019.
7. H. Goldstein, C.P. Poole, and J.L. Safko, *Classical Mechanics*. 3rd edition; published by Pearson Education Inc., 2002, 690 p.

Received 02.05.2020

INFORMATION ON THE ARTICLE

Petro I. Bidyuk, ORCID: 0000-0002-7421-3565, Educational and Scientific Complex “Institute for Applied System Analysis” of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: pbidyuke_00ukr.net

Yoshio Matsuki, ORCID: 0000-0002-5917-8263, National University of Kyiv-Mohyla Academy, Ukraine, e-mail: matsuki@wdc.org.ua

МОДЕЛЮВАННЯ КУТОВОГО МОМЕНТУ ГРАВІТАЦІЙНОГО ПОЛЯ ОБЕРТОВОЇ ЧОРНОЇ ДІРИ І СПІН-МОМЕНТУ ГРАВІТАЦІЙНИХ ХВИЛЬ / Й. Мацукі, П.І. Бідюк

Анотація. Змодельовано момент імпульсу гравітаційного поля обертової чорної діри і спіну-моменту гравітаційних хвиль, що випромінюються з чорної діри. Спочатку обчислено питому енергію обертового гравітаційного поля і спіну-моменту гравітаційних хвиль як векторів, що проєктуються на сферичну криволінійну поверхню гравітаційного поля та гравітаційних хвиль. Обчислено момент імпульсу та спіну-момент як вектори, перпендикулярні до криволінійної поверхні. У своєму дослідженні, опублікованому в 1964 р., Поль Дірак обрав не криволінійну поверхню для обчислення руху квантових частинок, а плоску поверхню для розроблення теорії квантової механіки. У цій роботі зроблено спробу змодельовати гравітаційні хвилі у сферичних полярних координатах, які утворюють сферичну криволінійну поверхню гравітаційних хвиль. З’ясовано, що множина антисиметричних векторів описує вектори, перпендикулярні до сферичної криволінійної поверхні; з такими векторами змодельовано момент імпульсу гравітаційного поля обертової чорної діри і спіну-моменту гравітаційних хвиль. Отримані результати описують характеристики обертання чорної діри та обертання гравітаційних хвиль.

Ключові слова: гравітаційні хвилі, кутовий момент, тензор кривизни, тензор енергії напруження, чорна діра.

МОДЕЛИРОВАНИЕ УГЛОВОГО МОМЕНТА ГРАВИТАЦИОННОГО ПОЛЯ ВРАЩАЮЩЕЙСЯ ЧЕРНОЙ ДЫРЫ И СПИН-МОМЕНТА ГРАВИТАЦИОННЫХ ВОЛН / Й. Мацуки, П.И. Бидюк

Аннотация. Смоделирован момент импульса гравитационного поля вращающейся черной дыры и вращающего момента гравитационных волн, излучающихся из черной дыры. Сначала вычислена удельная энергия вращающегося гравитационного поля и вращающихся гравитационных волн как векторов, проецирующихся на сферическую криволинейную поверхность гравитационного поля и гравитационных волн. Вычислены момент импульса и вращающий момент как векторы, перпендикулярные к криволинейной поверхности. В своем исследовании, опубликованном в 1964 г., Поль Дирак выбрал не криволинейную поверхность для вычисления движения квантовых частиц, а плоскую поверхность для разработки теории квантовой механики. В этой работе предпринята попытка смоделировать гравитационные волны в сферических полярных координатах, которые образуют сферическую криволинейную поверхность гравитационных волн. В результате выяснено, что множество антисимметричных векторов описывает векторы, перпендикулярные к сферической криволинейной поверхности; с такими векторами смоделированы момент импульса гравитационного поля вращающейся черной дыры и спин-момент гравитационных волн. Полученные результаты описывают характеристики вращения черной дыры и вращения гравитационных волн.

Ключевые слова: гравитационные волны, угловой момент, тензор кривизны, тензор энергии напряжения, черная дыра.

РЕДУКЦІЙНЕ КОНЦЕПТУВАННЯ ОРАКУЛЬНИХ СХЕМ

І.В. РЕДЬКО, П.О. ЯГАНОВ, М.О. ЗИЛЕВІЧ

Анотація. Роботу спрямовано на розвиток інтерсуб'єктивної парадигми та активної ролі суб'єкта у редукційному концептуванні. Для цього загальну оракульну схему концептування конкретизовано взаємодоповненням композиції і декомпозиції як експлікацій синтезу та аналізу сутностей. Прагматико-обумовлене збагачення цього взаємодоповнення здійснено із залученням композиційного програмування та іменних моделей даних, функцій і композицій. Розглянуто оракульну схему редукції, сенс якої полягає в тому, що вона, спираючись на наявні композиції, природним чином імплементує парадигму «поділяй і володарюй» у розумінні активної ролі суб'єкта в концептуванні, підтримуючи реальне взаємодоповнення декомпозиційного і композитного методів концептування. Наведено репрезентативні приклади редукційного концептування, які обґрунтовують технологію вирішення програмістських завдань.

Ключові слова: програмування, інтерсуб'єктивна парадигма, оракульна схема, концептування, редукція.

ВСТУП

Зародження, розвиток, становлення, стагнація та занепад будь-яких технічних систем і пов'язаних з ними технологічних процесів відбуваються за об'єктивними законами діалектики, проходячи ряд характерних етапів від бурхливого зростання до системної кризи, яка спонукає шукати вихід із ситуації, що склалася. Часто результативним вирішенням є зміна парадигми та створення і розвиток нової концепції, завдяки якій система у своєму зростанні виходить на новий рівень. Програмування як вид інформатико-технологічної діяльності не є винятком. Сьогодні стрімкий розвиток програмістської творчості, попри беззаперечні успіхи, нагромадив чимало проблем, ігнорування яких або відсилення їх на людський фактор веде до «бігу по колу» і відтворення помилок, що дедалі більше поглиблює кризу.

Розглянемо зміст програмування, яке зазвичай сприймають як діяльність з отримання результату — готової програми, тобто як процес реалізації плану (програми) досягнення цього результату. У такому разі варто зазначити, що програмування як діяльність провадиться в певних умовах, які формуються у невідривному зв'язку з програмою. Можна стверджувати про наявність обумовленості між програмуванням і програмою. Цей причинно-наслідковий зв'язок є суб'єктно-об'єктивним, у якому суб'єкт — програміст, будучи вагомим зовнішнім чинником, реалізовує свій задум, спираючись на власні уявлення про способи, методи, евристики, знахідки і запозичення досягнень результату в межах об'єкта — процесу створення програми. Об'єктивізм процесу виражається у можливості схематизації програмування генетичними структурами з властивими їм структурними сутностями. Генезис визначає структуру програми, її внутрішню логічну цілісність, якій підпорядковується результат програмування і нею обумовлюється.

У такому причинно-наслідковому зв'язку місце і роль суб'єктно-об'єктної взаємодії для сучасного розуміння програмування є ключовими і мають розглядатись не поза програмістською діяльністю, як це зазвичай відбувається, а безпосередньо як самостійний об'єкт дослідження. З одного боку, суб'єкт програмування не обмежений ніякими умовами у процесі програмування, крім головної, яку висловлює замовник, — програма повинна забезпечувати очікуваний результат. При цьому програміст «поділяє і володарює», обираючи на власний розсуд методи і засоби досягнення результату, використовуючи власний арсенал прийомів, підтверджуючи своїми діями поняття «мистецтво програмування». Його активна роль полягає не у залученні до схематизації об'єктивного причинно-наслідкового зв'язку між результатом і процесом його отримання у вигляді конкретних схем, а у формуванні власної авторської «функції множини дій», за допомогою якої він вирішує поставлене завдання без належної конкретизації проміжних результатів. Про технологію програмування у цьому випадку говорити важко, оскільки досягнення кожного суб'єкта програмування реально невіддільні від нього, бо не підтримують взаємодоповнення фактографії і фактології інформатико-технологічної діяльності. Загальновідомим у програмістському середовищі є твердження про те, що невдалу програму простіше створити заново, ніж виправити.

Активна роль суб'єкта повинна виразитись у конкретизації діяльності обумовлення, проваджуваної за допомогою плану програмування, актуалізацією значущих для процесу якісних (інтенціональних) суб'єктивних обумовлень. Таким чином, традиційна індивідуально-суб'єктивна парадигма, у межах якої програма розглядається як результат програмування, має змінитись інтерсуб'єктивною парадигмою, згідно з якою програмування — це діяльність суб'єкта зі створення програми (плану) цієї діяльності.

Суб'єкт програмування здійснює обумовлення в активній і пасивній формах. Активна форма — це безпосереднє накладання умови носієм таких умов (суб'єктом обумовлення) на програмування залежно від того, яким арсеналом засобів він оперує. Програміст впливає на результат і на спосіб досягнення цього результату — програмування як процесу. Активна форма виявляється безпосередньо в прагматичі використання для досягнення мети активних видів обумовлень, які створені не тільки в межах вирішення конкретного завдання, але і залучені, зокрема, з інших методів, що розвиваються на основі традиційного математичного апарату для нотації результату, а також денотативних прийомів.

Пасивна форма зумовлена наслідком активізації діяльності суб'єкта і опосередковано впливає на наслідок планування програми. Цей вплив виражається у залученні до процесу вирішення завдань, які вже відбулись, — або в межах цього процесу, або поза ним. Ці вирішення матеріалізовані у результатах діяльності суб'єкта і його власного розуміння такої діяльності. Зазвичай із цим пов'язують досвід, здобутий у процесі провадження активної діяльності та активного обумовлення цієї діяльності. В інтерсуб'єктивній парадигмі активно-пасивне обумовлення збагачує поняття програми як сутності, оскільки визначає роль суб'єкта не поза межами процесу програмування, а безпосередньо у ньому, знаходячи своє відображення, зокрема, у формі семантичного терму, фіксуючи ключові моменти розуміння діяльності, що спонукала до появи результату.

Мета роботи — подальше розвинення інтерсуб'єктивної парадигми та об'єктивізації активної ролі суб'єкта у редукційному концептуванні оракульних схем.

Для досягнення мети роботи загальну оракульну схему концептування конкретизовано взаємодоповненням композиції і декомпозиції як аналогів синтезу та аналізу сутностей, а прагматико-обумовлене збагачення цього взаємодоповнення розглянуто на репрезентативних прикладах редукційного концептування із залученням композиційного програмування та іменних моделей даних, функцій і операцій.

ОРАКУЛЬНА СХЕМАТИЗАЦІЯ КОНЦЕПТУВАННЯ

У працях [1–3] обґрунтовано засадниче для інтерсуб’єктивної парадигми програмування положення, що поза цілісним розумінням програмування немає його продуктивного розуміння. Концептуванню відведено роль носія можливих збагачень цього змістовного положення. Неможливість об’єктивного зведення один до одного цих видів розумінь зумовила відкритозамкненість та суб’єктно-об’єктність схеми концептування [3].

Згадані у цих працях схеми є концептами для різних композитних і композиційних концептувань. Останні отримуються за рахунок актуалізації оракулів [4, 5], що входять до схем. Із самої побудови схем випливає, що вони продукують функціональні сутесутнісні залежності вигляду

$Ent \xrightarrow{Comt} CtMon$ або $Ent \xrightarrow{Comp} CMon$, де $CtMon$ — збагачення оракула Mon за рахунок актуалізації концепту композитом; $Comt$ — оракул «композит»; $Comp$ — оракул «композиція»; $CMon$ — збагачення оракула Mon за рахунок актуалізації концепту композицією [4].

Отже, і композитні, і композиційні монади можуть бути експліковані як композиції сутностей: $CtMon \Rightarrow Comt(Ent_{i_1}, \dots, Ent_{i_k})$ і $CMon \Rightarrow Comp(Ent_{j_1}, \dots, Ent_{j_p})$, де $Ent_{i_1}, \dots, Ent_{i_k}, Ent_{j_1}, \dots, Ent_{j_p}$ — деякі сутності, а \Rightarrow означає «експлікативно зводиться». Наявні у схемі концептування оракули є «реперними точками» реалізації активної ролі суб’єкта в концептуванні. Виходячи з того, що ключовою ланкою реалізації активності суб’єкта є композит як концепт активності суб’єкта і його зв’язок з композицією, можна констатувати, що активність суб’єкта є концептуванням, концептом якого є композит, — базова для суб’єкта генна структура.

Розглядаючи ці експлікативні зведення в контексті концептування і проєктуючи їх на відповідні функціональні залежності, отримуємо такі оракульні схеми: $Ent \xrightarrow{Comt} Comt(Ent_{i_1}, \dots, Ent_{i_k})$ і $Ent \xrightarrow{Comp} Comp(Ent_{j_1}, \dots, Ent_{j_p})$.

Особливість кожної з них полягає у тому, що перша використовує базову композицію, але при цьому на сутності $Ent_{i_1}, \dots, Ent_{i_k}$ ніяких додаткових обмежень не накладається. У другій же, навпаки, композиція не обов’язково базова, швидше вона похідна від покроковості застосування композитів, зате сутності $Ent_{j_1}, \dots, Ent_{j_p}$ є елементарними, тобто достатньо деталізованими для суб’єкта. Це створює можливість передбачення наслідку концептування як композиції елементарних, для суб’єкта, сутностей.

У наведених схемах явно відображений факт взаємодоповнення двох методів дослідження концептування — синтезу та аналізу, композиції і декомпозиції. Синтез поданий в них оракулами $Comt$ і $Comp$, а аналіз (декомпозиція) — $Ent_{i_1}, \dots, Ent_{i_k}, Ent_{j_1}, \dots, Ent_{j_p}$.

Звернемо увагу на те, що перша схема оперує базовими для суб'єкта композиціями (комполітатами). Це дозволяє здійснити ще один крок до продуктивного збагачення розуміння програмування. Будемо говорити про те, що кортеж $\langle Ent_{i_1}, \dots, Ent_{i_k} \rangle \in Comt$ -редукцією сутності Ent , якщо існує функціональна залежність $Ent \xrightarrow{Comt} Comt(Ent_{i_1}, \dots, Ent_{i_k})$. З наведеного розгляду безпосередньо випливає твердження про оракульну схематизацію редукції.

Твердження. Кортеж $\langle Ent_{i_1}, \dots, Ent_{i_k} \rangle \in Comt$ -редукцією сутності Ent , якщо справедлива $Ent = Comt(E_{i_1}, \dots, E_{i_k})$.

Змістовний сенс наведеної оракульної схеми полягає у тому, що вона природним чином імплементує парадигму «поділай і володарюй» в розумінні активної ролі суб'єкта в концептуванні, підтримуючи реальне взаємодоповнення декомпозиційного і композитного методів концептування [3]. Таким чином, композит $Comt$ як актуально задана базова композиція є концептом Ent , тобто Ent є монадою. Цю особливість буде використано у наведених нижче прикладах редукційного концептування.

РЕПРЕЗЕНТАТИВНІ ПРИКЛАДИ РЕДУКЦІЙНОГО КОНЦЕПТУВАННЯ

Природа композитів і композицій в цілому релятивна, що істотно залежить як від суб'єкта, так і від розглядуваних сутностей. У свою чергу, розуміння сутності залежить від композитів, що залучаються у розгляд. Тому реальне проникнення в інтерсуб'єктивну природу композитів і композицій можливе тільки у взаємодії з прагматико-обумовленим збагаченням цього взаємодоповнення. У цьому напрямі отримано велику кількість результатів різної глибини і значущості [6–11]. Детальний розгляд такої фактографії виходить за межі цієї роботи. Тому доцільно поповнити її якісно іншими простими і, разом з тим, репрезентативними прикладами редукційного концептування.

Розв'язання будь-якої задачі, як відомо, є інтеграцією розв'язань її підзадач. Якщо задача проста, то інтеграція тривіальна і, як правило, явно не виділяється. У разі ж коли задача складна, інтеграційний аспект її розв'язання домінує, адже власне саме ним і визначається складність.

Відомо, що рівень складності розв'язання реальних задач визначається, головним чином, складністю їх інтеграційних складових. Зважаючи на це, подальші побудови проведемо «від простого до складного», розглянувши на прикладах нескладних задач числового аналізу їх розв'язання в середовищах мікро- та макроінтеграції. Як платформу розгляду використаємо композиційне програмування та іменну модель даних, функцій і операцій, а як композити — операції мультиплікування \circ , розгалуження IF , циклування WD і найпростіші похідні від них композиції, що уточнюють найбільш вживані та прості способи генезису одних програм з інших [6, 7]. Надалі під даними, функціями та операціями, якщо не зазначено інше, розумітимемо іменні дані, іменні функції та іменні операції відповідно.

Проведене змістовне розгортання і роз'яснення концептування достатньо обгрунтовує точку зору на нього як на покроковість редукувань, що зводиться до пошуку підхожої $Comt$ -редукції для індукованого відповідною функціональною залежністю рівняння $Ent = Comt(E_{i_1}, \dots, E_{i_k})$. Під розв'язком тут розуміється кортеж $\langle Ent_{i_1}, \dots, Ent_{i_k} \rangle$ такий, що справедливо

вою є тотожність $Ent \equiv Comt(Ent_{i_1}, \dots, Ent_{i_k})$. Запис $Comt(Ent_{i_1}, \dots, Ent_{i_k})$ означає застосування k -арного композита $Comt$ до кортежу $\langle Ent_{i_1}, \dots, Ent_{i_k} \rangle$, тобто $Comt(Ent_{i_1}, \dots, Ent_{i_k}) \stackrel{\text{def}}{=} Ap(Comt, \langle Ent_{i_1}, \dots, Ent_{i_k} \rangle)$, де Ap розуміється традиційно як аплікація [9], а $\stackrel{\text{def}}{=}$ як рівність за визначенням.

Розглянемо застосування апарату простої WD -редукції, рухаючись від простого до складного. Така редукція є кортежем $\langle g, p \rangle$, що є розв'язком рівняння $f = WD(g, p)$, де g, p — деякі функція та предикат. Для пошуку розв'язку корисними є наслідки згаданого твердження про схематизацію редукції, які продуктивно збагачують його, при цьому істотно звужуючи коло пошуку редукції.

Наслідок 1. $\langle g, p \rangle$ є WD -редукцією функції f , якщо $f = WD(p, g)$.

Наслідок 2. Щоб кортеж $\langle g, p \rangle$ був WD -редукцією функції f , необхідно, щоб виконувалась рівність $g \circ f = f$, де \circ — операція мультиплікування.

Проілюструємо це на конкретних прикладах. Для цього, передусім, звернемося до найпростішого класу задач числового аналізу, що складається з однієї задачі — обчислення \sqrt{x} із заданою точністю ε , де x і ε — додатні дійсні числа.

Відомо, що послідовність y_0, y_1, y_2, \dots , у якій $y_0 = a, y_{i+1} = \frac{1}{2} \left(y_i + \frac{x}{y_i} \right)_{i=0,1,2,\dots}$, де a — деяке додатне дійсне число, незалежно від a , збігається до \sqrt{x} . Звідси випливає, що процес обчислення \sqrt{x} із заданою точністю може бути зведений до деталізації іменної функції f [7, 11], яка іменній множині $\{(u, x), (v, \varepsilon), (w, a)\}$ ставить у відповідність іменну множину $\{(w, y_n)\}$, де y_n — перший член зазначеної послідовності, для якого виконується умова $|y_n^2 - y_{n-1}^2| < \varepsilon$.

Знайдемо WD -редукцію функції $f_{\sqrt{x}}$. Розглянемо наступні іменну функцію $g : \{(u, x), (v, \varepsilon), (w, a)\} \rightarrow \{(u, x), (v, \varepsilon), (w_{pr}, a), \left(w, \frac{1}{2} \left(a + \frac{x}{a} \right) \right)\}$ та імен-

ний предикат $p : \{(v, \varepsilon), (w, a), (w_{pr}, b)\} \rightarrow \begin{cases} \text{True}, & |a^2 - b^2| \geq \varepsilon, \\ \text{False}, & |a^2 - b^2| < \varepsilon. \end{cases}$ Легко переко-

натися у справедливості рівності $f_{\sqrt{x}} = g \circ f_{\sqrt{x}}$ і, як наслідок, у тому, що $\langle g, p \rangle$ дійсно є WD -редукцією функції. Адже дійсно, якщо

$|a^2 - \left(\frac{1}{2} \left(a + \frac{x}{a} \right) \right)^2| < \varepsilon$, то $f_{\sqrt{x}} \left(\left\{ (u, x), (v, \varepsilon), \left(w, \frac{1}{2} \left(a + \frac{x}{a} \right) \right) \right\} \right) = \{(w, y_n)\}$. От-

же, можна зробити висновок: $f_{\sqrt{x}} \equiv WD(g, p)$. При цьому правильність висновку безпосередньо впливає з побудови. Особливістю цього вирішення є те, що його синтаксичне оформлення зводиться до трансляції відповідного семантичного терму в обрану мову програмування.

```

Procedure G(x:double; var rz, zr:double);
begin
    zr:=rz;
    rz:=0,5*(rz+x/rz);
end;
    
```

Рис. 1. Приклад процедури G

Продемонструємо це на прикладі мови програмування Pascal. Почнемо із функції редукції g та предиката p . Процедура G , очевидно, подає функцію g у синтаксисі мови Pascal (рис. 1).

ж синтаксисі подає предикат p (рис. 2).

Насамкінець подамо шукану функцію $f : \{(u, x), (v, \varepsilon), (w, a)\} \rightarrow \{(w, y_n)\}$

```

Function P(rz,zr: double):boolean;
begin
    if abs(sqrt(rz)-sqrt(zr))>=v then
        P:=True
    else
        P:=False;
end;
    
```

Рис. 2. Приклад процедури-функції P

у Pascal-синтаксисі (з деякими неприциповими скороченнями, зробленими заради компактності викладення матеріалу, рис. 3).

тись в автоматичному режимі за допомогою спеціалізованого програмного забезпечення — дефінітора мови Pascal [3, 7]. Таким чином, необтяженість

Значимо, що синтаксичне подання шуканого розв'язку може породжуватись семантичної специфікації розв'язання задачі не тільки не заважає отриманню його синтаксичного подання, а є суттєвою перевагою такої специфікації перед надто конкретизованими високо рівневими мовами програмування.

```

Program F;
var u,v,w,wpr: double;
Procedure G(x:double; var rz,zr:double);
...
Function P(rz,zr:double):boolean;
begin
... // Блок уведення значень u,v,w
if (u>0) and (v>0) and (w>0) then
begin
G(u, w, wpr);
while P(w, wpr) do G(u, w, wpr);
writeln(w);
end;
end.
    
```

Рис. 3. Приклад Pascal-подання шуканої функції

Даний приклад демонструє задачу, що розв'язується «в один крок». Для розв'язань таких задач не потрібно залучати оракульні схеми концептування, тому і

інтеграційна складова тут є тривіальною. Наступний приклад вже на макрорівні ілюструє інтеграційний аспект концептування і значно виразніше демонструє переваги редукційного концептування. Розглянемо клас спеціальних рівнянь вигляду $x = \varphi(x)$, де функція $\varphi(x)|_{x \in R}$ задовольняє такі дві умови:

- 1) вона визначена і неперервно диференційовна на всій числовій прямій;
- 2) існує таке дійсне число $p < 1$, що для всіх x справедлива нерівність

$$|\varphi'(x)| \leq p.$$

Відомо, що стосовно такого класу рівнянь метод простих ітерацій збігається. Причому розв'язком є границя послідовності $\{x_i\}_{i=0,1,2,\dots}$, де x_0 — будь-яке дійсне число, а $x_{i+1} = \varphi(x_i)$, $i = 0,1,2,\dots$

Із наведених умов випливає, що концептування задачі пошуку розв'язку рівнянь вигляду $x = \varphi(x)$ можна звести до деталізації функції

$f : \{(v, x_0), (u, \varepsilon)\} \rightarrow \{(v, x_n)\}$, де x_n — перший член послідовності наближень, для якого виконується умова $|x_{n-1} - x_n| < \varepsilon$. Очевидно, що кортеж $\langle g, p \rangle$, де $g : \{(v_{pr}, a), (v, b)\} \rightarrow \{(v_{pr}, b), (v, \varphi(b))\}$, а $p : \{(v_{pr}, a), (v, b), (u, \varepsilon)\} \rightarrow$
 $\rightarrow \begin{cases} \text{True}, & |a - b| \geq \varepsilon \\ \text{False}, & |a - b| < \varepsilon \end{cases}$ є розв'язком рівняння $f = WD(E_1, E_2)$. Тобто $f \equiv WD(g, p)$.

Особливість отриманого розв'язку полягає в тому, що він є вже оракульною схемою розв'язання класу задач, інакше кажучи, є схемою монади. Оракулом тут є $\varphi : \{(v, b)\} \rightarrow \{(v, \varphi(b))\}$ — іменна специфікація $\varphi(x)|_{x \in R}$. Схема перетворюється в конкретну монаду після заміни φ у схемі конкретною функцією, що задовольняє наведені вище дві умови. Такими, наприклад, є функції $\varphi : \{(v, b)\} \rightarrow \left\{ \left(v, \frac{\cos(b)}{2} \right) \right\}$, $\varphi : \{(v, b)\} \rightarrow \left\{ \left(v, \frac{\sin(b) + \cos(b)}{3} \right) \right\}$ і т.ін.

У цій схемі оракульну взаємодію подано рудиментарно, оскільки в ній усього один оракул — φ . При цьому неможливість коректного Pascal-подання отриманої схеми через потенційну нескінченність класу описуваних нею окремих розв'язків не заважає використовувати її для породження Pascal-специфікації конкретних монад. Для цього необхідно лише монадизувати оракул φ .

Нехай $\varphi(x) = \sum_{i=1}^{\infty} (-1)^i \frac{x^{2i}}{2(2i)!}$, $|x| < \infty$, що задовольняє наведені вище умови, і

Pascal-платформа реалізації оракула φ містить, крім стандартних математичних функцій мови Pascal, і функцію факторіала FC .

Тоді для будь-якого дійсного a послідовність y_0, y_1, y_2, \dots , де $y_0 = 0$,

$y_{k+1} = \sum_{i=1}^{k+1} (-1)^i \frac{a^{2i}}{2(2i)!} = y_k + (-1)^{k+1} \frac{a^{2(k+1)}}{2(2(k+1))!} \Big|_{k=0,1,2}$, збігається до $\varphi(a)$. Для

обчислення $\varphi(a)$ із заданою точністю до дійсного додатного числа ε корисною є функція q , що іменній множині $\{(v, a), (t, \varepsilon), (w, b)\}$ ставить у відповідність іменну множину $\{(w, y_n)\}$, де y_n — перший член зазначеної

послідовності, для якого виконується умова $\left| \frac{a^{2n}}{2(2n)!} \right| < \varepsilon$. Очевидно, що

$\Phi = q_1 \circ q$, де $q_1 : \{(w, b), (s, m)\} \rightarrow \{(w, 0), (s, 0)\}$, а $\Phi : \{(v, a), (t, \varepsilon)\} \rightarrow$
 $\rightarrow \{(w, y_n)\}$ — іменна функція, яка обчислює наближене значення $\varphi(a)$ із

заздалегідь заданою точністю. Тобто пара $\langle q_1, q \rangle$ є \circ -редукцією функції Φ . Функція q_1 обнулює «комірки» w і s , тобто подається операторами присвоювання $w := 0$; $s := 0$; і тому не потребує подальшої деталізації. Що стосується функції q , то аналогічно до попереднього, пара $\langle r, p \rangle$, де

$r : \{(v, a), (w, b), (s, k)\} \rightarrow \left\{ \left(w, b + (-1)^{k+1} \frac{a^{2(k+1)}}{(2(k+1))!} \right), (s, k+1) \right\} \Big|_{a, b \in R, k \in N}$ та

$$p : \{(t, \varepsilon), (v, a), (s, k)\} \rightarrow \begin{cases} \text{True, } \left| \frac{a^{2n}}{2(2n)!} \right| \geq \varepsilon, \\ \text{False, } \left| \frac{a^{2n}}{2(2n)!} \right| < \varepsilon \end{cases} \in \text{WD-редукцією функції } q. \text{ I, та}$$

ким чином, $\Phi = q_1 \circ \text{WD}(r, p)$.

Відповідне подання функції r здійснюється Pascal-процедурою (рис. 4).

```

Procedure R(var v, w: double; var s: integer);
begin
  w:=w+(((exp((s+1)*ln(-1))*(exp(2*(s+1)*ln(v))))/(2*FC((2*(s+1))))));
  s:=s+1;
end;
    
```

Рис. 4. Приклад процедури R

Предикат p подається процедурою, аналогічною до наведеної у попередньому прикладі процедури-функції P (рис. 5).

```

Function P(v, t: double; s:integer):boolean;
begin
  if s>0 then
  begin
    if abs((((exp((s+1)*ln(-1))*(exp(2*(s+1)*ln(v))))/(2*FC((2*(s+1))))))>=t
    then
      P:=True
    else
      P:=False;
    end;
  end;
end;
    
```

Рис. 5. Приклад процедури-функції P

Тепер нескладно надати Pascal-подання для функції Φ (рис. 6).

```

Function FI(v,t:double):real;
Var
  i:integer;
  w:double;
begin
  w:=0; i:=0;
  while P(abs((((exp((i+1)*ln(-1))*(exp(2*(i+1)*ln(v))))/(2*FC((2*(i+1))))))>t
  do
  begin
    R(v,w,i);
  end;
  F:=w;
end;
    
```

Рис. 6. Приклад процедури-функції FI

Продовжуючи побудови, приходимо до Pascal-програми (рис. 7).

Розглянуті приклади демонструють важливі особливості редуційного концептування оракульних схем. Це, по-перше, можливість переходу від розв'язань окремих задач до розв'язань класів подібних задач з гарантова-

ною коректністю отримуваних редукційним концептуванням розв'язків за їх побудовою. І, по-друге, отримані концептуванням розв'язки можуть бути реалізовані на різних синтаксичних програмних платформах.

```

Program F;
  var u,w,wpr,t:double;
      s:integer;
  Function G(e:double; x:double):double;
  begin
      x:=FI(x,e); G:=x;
  end;
  Function P(x, v:double):boolean;
  begin
      if abs(x- G(x))>=v then P:=True else P:=False;
  end;
  Procedure R(x:double; var rz,zr:double; var s:integer);
  begin
      ...
  end;
  Function FI(v,t:double):real;
  begin
      ...
  end;
  // Блок уведення значень u,t,w
  ...
  begin
      if (t>0) then
          begin
              w:=0;
              while P(w, t) do G(w, t);
              readln(w);
          end;
      end.
  
```

Рис. 7. Приклад програми F

Прагнення до першого було і залишається спонукальним мотивом створення більшості парадигм програмування. У тому чи іншому вигляді важливість цього знайшла своє відображення, зокрема, у тьюрінговських лекціях Джона Бекуса, Едгера Дейкстра, Джона Мак-Карті, Роберта Флойда [12]. Видатні вчені, які зробили вагомий внесок у розбудову засад, у тому числі і комп'ютерної науки та програмування у цілому, такі як Алонзо Чорч, Стівен Кліні, Гаскелл Каррі, Ніклаус Вірт, Алан Кей та їх послідовники у своїй діяльності також керувались і керуються проблемами підвищення продуктивності та забезпечення коректності отримуваних розв'язків [13–26]. Реінжиніринг програмного забезпечення з лавиноподібним зростанням інвестицій у його розроблення став самоочевидним трендом розвитку інформаційних технологій та програмування [27]. Віддаючи належне визначним досягненням, зробленим у цьому напрямі, тим не менше, необхідно звернути увагу на те, що всі вони з об'єктивних причин слабо інтегровані між собою. Адже їх адекватна інтеграція потребувала розгляду ряду питань, для вирішення яких не було достатньої фактографії. Такими, зокрема, є питання об'єктивізації суб'єктивних впливів на розв'язання задач переходу від систем, орієнтованих на замкнуті у конкретиці розв'язки, до систем, орієнтованих на розв'язання класів подібних задач.

Що ж до другого, то спрямованість редукційного концептування на притаманні програмуванню генетичні структури дозволяє враховувати активну роль суб'єкта у ньому, на відміну від більшості традиційних підходів, які орієнтовані на синтаксичну нотацію результатів свідомого чи несвідомого застосування таких генетичних структур. Таким чином, маємо достатньо підстав розглядати редукційне концептування оракульних схем як продуктивну експлікацію програмування.

ВИСНОВКИ

У роботі обґрунтовано ключові поняття інтерсуб'єктивної парадигми програмування із залученням оракульних схем концептування. Розглянуто експлікативні зведення концептування до нових оракульних схем, які є концептами для різних композитних і композиційних концептувань. Інструментом такого зведення є конкретизація оракулів, що входять до цих схем. Продуктивне збагачення розуміння програмування знайшло своє відображення у твердженні про оракульну схематизацію редукції.

Розвинуто і збагачено новим змістом програмування як діяльність, орієнтовану не тільки на нотацію одержуваних розв'язків, але таку, яка здатна забезпечувати продуктивну діяльність суб'єкта для їх отримання із залученням розвинених засобів програмування, адаптованих для реалізації активної ролі суб'єкта. Для цього необхідно вирішувати проблему побудови програмологічних засад продуктивної діяльності і розвитку адаптивних інформатико-технологічних середовищ.

Наведено репрезентативні приклади редукційного концептування, які обґрунтовують оракульну схематизацію як технологію вирішень програмістських завдань.

Подальші дослідження з цієї проблематики будуть спрямовані на розширення змістовних природничо-наукових досліджень, фактографії, що підтверджує їх, і розвиток відповідної фактології для оракульної схематизації концептування як продуктивного засобу специфікації розв'язків класів задач і розвинених на її основі редукційних методів програмування.

ЛІТЕРАТУРА

1. І.В. Редько та П.О. Яганов, "Концептуальна модель технологічного середовища програмування", *Наукові вісті КІП*, № 1, с. 18–26, 2020. doi: 10.20535/kpi-sn.2020.1.197953.
2. I. Redko, P. Yahanov, and M. Zylevich, "Concept-Monadic Model of Technological Environment of Programming", *2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC)*, Kyiv, Ukraine, 2020, pp. 125–130. doi: 10.1109/SAIC51296.2020.9239204.
3. І.В. Редько, Д.І. Редько, та Т.Л. Захарченко, *Концептологічні основи проектування*. Київ: Компринт, 2016.
4. В.Р. Карымов, "Арифметическая и гиперарифметическая вычислимость относительно вычислений с ограничениями", *Математика и механика*, № 5, с. 48–52, 2010.
5. M. Trakhtenbrot, "Mutation Patterns for Temporal Requirements of Reactive Systems", *2017 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, Tokyo, 2017, pp. 116–121. doi: 10.1109/ICSTW.2017.27
6. А.И. Мальцев, *Алгоритмы и рекурсивные функции*. Москва: Наука, 1965.

7. И.А. Басараб, Н.С. Никитченко, и В.Н. Редько, *Композиционные базы данных*. Киев: Лыбидь, 1992.
8. Д.Б. Буй, Е.В. Шишацкая, S. Fabunmi, K. Mohammed, “Математичні основи алгоритмів лінеаризації: рефлексивно-транзитивне замикання бінарних відношень”, *Вісник Харківського національного університету імені В.Н. Каразіна*, № 29, с. 19–33, 2016.
9. Х. Барендрегт, *Лямбда-исчисление. Его синтаксис и семантика*. Москва: Мир, 1985.
10. Дж. Бекус, “Алгебра функциональных программ: мышление функционального уровня, линейные уравнения и обобщенные определения”, *Математическая логика в программировании*. Москва: Мир, 1991, с. 8–53.
11. В.Н. Редько, “Основы прогамологии”, *Кибернетика и системный анализ*, № 1, с. 35–57, 2000.
12. Р. Эшенхерст, *Лекции лауреатов премии Тьюринга*. Москва: Мир, 1993.
13. H.D. Curry, R. Hindley, and J.P. Seldin, *Combinatory Logic. Studies in Logic*. Amsterdam: North-Holland Co., 1972.
14. У. Дал, Э. Дейкстра, и К. Хоор, *Структурное программирование*. Москва: Мир, 1975.
15. Э. Дейкстра, *Дисциплина программирования*. Москва: Мир, 1978.
16. J. McCarthy, “Recursive Functions of Symbolic Expressions and Their Computation by Machine”, *Communications of the ACM*, vol. 3, pp.184–195, 1960.
17. Н. Вирт, *Алгоритмы и структуры данных*. Москва: ДМК Пресс, 2010.
18. Р. Пенроуз, *Глава 2: Лямбда-исчисление Черча. Новый ум короля. О компьютерах, мышлении и законах физики*. Великобритания: Издательство Оксфордского университета, 2011.
19. В.В. Соколов, “Застосування функціональної та реляційної моделей в об’єктно-орієнтованому програмуванні”, *Information Technology and Security*, № 1, с. 54–63, 2017.
20. О.В. Олецкий та Є.В. Івохін, “Формалізація процедури формування динамічної рівноваги альтернатив у багатоагентному середовищі у процесах прийняття рішень більшістю голосів”, *Кибернетика та системний аналіз*, № 1, с. 55–66, 2021.
21. Р. Мартин, *Часть II. Парадигмы программирования. Чистая архитектура*. СПб: Питер, 2018.
22. A. Appel, *Software Foundations Volume 3: Verified Functional Algorithms*. 2018. [Online]. Available: <https://softwarefoundations.cis.upenn.edu/vfa-current/index.html>
23. Miltner S. Padhi, T. Millstein, and D. Walker, “Data-driven inference of representation invariants”, *41st ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2020)*. Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–15. doi: <https://doi.org/10.1145/3385412.3385967>
24. V. Premtoon, J. Koppel, and A. Solar-Lezama, “Semantic code search via equational reasoning”, *41st ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2020)*. Association for Computing Machinery, New York, NY, USA, 2020, pp. 1066–1082. doi: <https://doi.org/10.1145/3385412.3386001>
25. R. Ji, J. Liang, Y. Xiong, L. Zhang, and Z. Hu, “Question selection for interactive program synthesis”, *41st ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2020)*. Association for Computing Machinery, New York, NY, USA, 2020, pp. 1143–1158. doi: <https://doi.org/10.1145/3385412.3386025>
26. Z. Guo et al., “Program synthesis by type-guided abstraction refinement”, *41st ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2020)*. Association for Computing Machinery, New York, NY, USA, 2020, pp. 28. doi: <https://doi.org/10.1145/3371080>
27. M. Majthoub, M.H. Qutqut, and Y. Odeh, “Software Re-engineering: An Overview”, *2018 8th International Conference on Computer Science and Information Technology (CSIT)*, Amman, 2018, pp. 266–270. doi: [10.1109/CSIT.2018.8486173](https://doi.org/10.1109/CSIT.2018.8486173).

Надійшла: 16.12.2020

INFORMATION ON THE ARTICLE

Igor V. Redko, ORCID: 0000-0002-3121-1412, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: redkoigor@ukr.net

Petro O. Yahanov, ORCID: 0000-0001-7358-9846, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: piteryahanov@gmail.com

Maksym O. Zylevich, ORCID: 0000-0003-1646-0557, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: zila@meta.ua

РЕДУКЦИОННОЕ КОНЦЕПТИРОВАНИЕ ОРАКУЛЬНЫХ СХЕМ / И.В. Редько, П.А. Яганов, М.О. Зылевич

Аннотация. Работа направлена на развитие интересубъективной парадигмы и активной роли субъекта в редукционном концептировании. Для этого общая оракульная схема концептирования конкретизирована взаимодействием композиции и декомпозиции как экспликаций синтеза и анализа сущностей. Прагматико-обусловленное обогащение этого взаимодействия осуществлено с привлечением композиционного программирования и именных моделей данных, функций и композиций. Рассмотрена оракульная схема редукции, смысл которой заключается в том, что она, опираясь на имеющиеся композиции, естественным образом имплементирует парадигму «разделяй и властвуй» в смысле активной роли субъекта в концептировании, поддерживая реальное взаимодействие декомпозиционного и композиционного методов изучения концептирования. Приведены репрезентативные примеры редукционного концептирования, которые обосновывают технологию решения программистских задач.

Ключевые слова: программирование, интересубъективная парадигма, оракульная схема, концептирование, редукция.

REDUCTION CONCEPTUALIZATION OF ORACLE SCHEMES / I.V. Redko, P.O. Yahanov, M.O. Zylevich

Abstract. This work is aimed at developing an intersubjective paradigm and an active role of the subject in reduction conceptualization. For this purpose, the general oracular scheme of conceptualization is concretized by the complementarity of composition and decomposition as explications of synthesis and analysis of entities. Pragmatically conditioned enrichment of this complementarity is carried out with the involvement of compositional programming and nominal models of data, functions, and compositions. The oracular scheme of reduction is considered, the meaning of which is that it, based on existing compositions, naturally implements the paradigm of “divide and conquer” in understanding the active role of the subject in conceptualization, supporting the real complementarity of decomposition and compositional methods of conceptualization. Representative examples of reduction conceptualization are shown, which substantiate the technology of solving programming problems.

Keywords: programming, intersubjective paradigm, oracle scheme, conceptualization, reduction.

REFERENCES

1. I. Redko and P. Yahanov, “Conceptual model of the technological environment of programming”, *KPI Science News*, vol.1, no.1, pp. 18–26, 2020. doi: 10.20535/kpi-sn.2020.1.197953.
2. I. Redko, P. Yahanov, and M. Zylevich, “Concept-Monadic Model of Technological Environment of Programming”, *2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC)*, Kyiv, Ukraine, pp. 125–130, 2020. doi: 10.1109/SAIC51296.2020.9239204.
3. I.V. Redko, D.I. Redko, and T.L. Zakharchenko, *Conceptual basis of programming*. Kyiv, Ukraine: Komprynt, 2016.
4. V.R. Karymov, “Arithmetic and hyperarithmetic computability with respect to constrained computations”, in *Mathematics and Mechanics*, no. 5, pp. 48–52, 2010.

5. M. Trakhtenbrot, "Mutation Patterns for Temporal Requirements of Reactive Systems", *2017 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, Tokyo, 2017, pp. 116–121. doi: 10.1109/ICSTW.2017.27
6. A.I. Maltsev, *Algorithms and recursive functions*. Moscow, Russia: Nauka, 1965.
7. I. Basarab, N. Nykytchenko, and V. Redko, *Composite databases*. Kyiv, Ukraine: Lybid, 1992.
8. D.B. Buy, E.V. Shishatskaya, S. Fabunmi, and K. Mohammed, "Mathematical foundations of linearization algorithms: reflexive-transitive closure of binary relations", in *Bulletin of V.N. Karazin Kharkiv National University*, no. 29, pp. 19–33, 2016.
9. H.P. Barendregt, *The Lambda Calculus. Its syntax and semantics*. Moscow, Russia: Mir, 1985.
10. J.W. Backus, "Algebra of functional programs: functional level thinking, linear equations and generalized definitions", in *Mathematical logic in programming*. Moscow, Russia: Mir, 1991, pp. 8–53.
11. V.N. Redko, "Foundations of programmology", *Kibernetika i sistemnyj analiz*, vol. 1, pp. 35–57, 2000.
12. R. Eshenkherst, *Turing Prize Winners Lectures*. Moscow, Russia: Mir, 1993.
13. H.D. Curry, R. Hindley, and J.P. Seldin, *Combinatory Logic. Studies in Logic*. Amsterdam: North-Holland Co., 1972.
14. U. Dal, E. Deikstra, and K. Koor, *Structured programming*. Moscow, Russia: Mir, 1975.
15. E. Deikstra, *Discipline of programming*. Moscow, Russia: Mir, 1978.
16. J. McCarthy, "Recursive Functions of Symbolic Expressions and Their Computation by Machine", *Communications of the ACM*, vol. 3, pp. 184–195, 1960.
17. N. Virt, *Algorithms and data structures*. Moscow, Russia: DMK Press, 2010.
18. R. Penrose, *Chapter 2: Church's Lambda Calculus. The new mind of the king. About computers, thinking and the laws of physics*. United Kingdom: Oxford University Press, 2011.
19. V.V. Solokov, "Application of functional and relational models in object-oriented programming", in *Information Technology and Security*, no. 1, pp. 54–63, 2017.
20. O.V. Olets'kyi and E.V. Ivokhin, "Formalization of the procedure for the formation of a dynamic balance of alternatives in a multi-agent environment in decision-making processes by a majority vote", in *Cybernetics and Systems analysis*, no. 1, pp. 55–66, 2021."
21. R. Martin, *Part II. Programming paradigms. Clean architecture*. Saint Petersburg, Russia: Piter, 2018.
22. A. Appel, *Software Foundations Volume 3: Verified Functional Algorithms*. 2018. [Online]. Available: <https://softwarefoundations.cis.upenn.edu/vfa-current/index.html>
23. Miltner S. Padhi, T. Millstein, and D. Walker, "Data-driven inference of representation invariants", *41st ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2020)*. Association for Computing Machinery, New York, NY, USA, 2020, pp. 1–15. doi: <https://doi.org/10.1145/3385412.3385967>
24. V. Premtoon, J. Koppel, and A. Solar-Lezama, "Semantic code search via equational reasoning", *41st ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2020)*. Association for Computing Machinery, New York, NY, USA, 2020, pp. 1066–1082. doi: <https://doi.org/10.1145/3385412.3386001>
25. R. Ji, J. Liang, Y. Xiong, L. Zhang, and Z. Hu, "Question selection for interactive program synthesis", *41st ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2020)*. Association for Computing Machinery, New York, NY, USA, 2020, pp. 1143–1158. doi: <https://doi.org/10.1145/3385412.3386025>
26. Z. Guo et al., "Program synthesis by type-guided abstraction refinement", *41st ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI 2020)*. Association for Computing Machinery, New York, NY, USA, 2020, pp. 28. doi: <https://doi.org/10.1145/3371080>
27. M. Majthoub, M.H. Qutqut, and Y. Odeh, "Software Re-engineering: An Overview", *2018 8th International Conference on Computer Science and Information Technology (CSIT)*, Amman, 2018, pp. 266–270. doi: 10.1109/CSIT.2018.8486173.

METHOD OF SEARCHING FOR INFORMATION OBJECTS IN UNIFIED INFORMATION SPACE

**A. DODONOV, V. MUKHIN, V. ZAVGORODNII, Ya. KORNAGA,
A. ZAVGORODNYA**

Abstract. The article discusses the concept and principles of building unified information space and presents a scheme for its formation. The article considers formation of unified information space using a specialized information computer system, which is actually a hardware and software basis for supporting unified information space. The stages of information object identification in unified information space are considered. The article suggests a method for finding missing features of an incoming object by implementing the information objects interaction with each other within unified information space.

Keywords: unified information space, information object, signs, object identification, search method.

INTRODUCTION

Unified information space is an information model of a complex subject area. It includes information objects, connections between them, environment of the space and processes accompanying creation of unified information space [1–5].

Unified information space is formed as a result of processing information about an object, received from various sources.

Here is the contradiction, which is as follows: in order to obtain information features of objects, the heterogeneous data sources there are used, and these sources are characterized by varying degrees of accuracy and different formats for data presenting. At the same time, for the formation of unified information space, the unification of data obtained from heterogeneous sources is required. The implementation of a converting mechanism for such formats also is required. So, the contradiction arises between the heterogeneous nature of sensors for collecting features of objects and the requirement for a unified data presentation. In this case, the same object, the parameters of which are obtained from different sensors, must be uniquely identified anyway.

In the process of forming unified information space, an information computer system collects information from various data sources presented in different forms and / or formats, while the processing of incoming data can be carried out by heterogeneous computer systems [6–8].

To form unified information space, it is required to implement a unified data entry, store data in uniform formats and exchange information between all infor-

mation objects [5, 9, 10]. An information object is a mathematical description of an initial object by its main parameters. It can be represented as a tuple of parameters of a real object, and all values of the parameters are determined by the characteristics of the real object.

Information about objects in unified information space changes dynamically [11–12].

Creation of unified information space is intended to provide a unified description of information objects for all users, so that all users of unified information space perceive the same information object in the same way. This characteristic is the main feature of unified information space [1, 2].

Thus, unified information space is a complex of tools that support the unity of presentation, processing and interpretation of information about information objects. Creation of unified information space is aimed to provide access to general information without limiting the place and time [13–19].

The information computer system, on the basis of which unified information space is formed, performs the following main functions:

- transformation of information about objects and formation of unified information space;
- providing users with information about objects.

The goal of the research is to consider the concept and principles of building unified information space and present a scheme for its formation using a specialized information computer system; describe the stages of identifying an information object in unified information space; propose a method for finding missing features of an incoming object.

SCHEME OF UNIFIED INFORMATION SPACE FORMATION

Fig. 1 shows a general scheme for unified information space formation. An important task of unified information space is to transform the input information in such a way that each information object in unified information space is presented uniquely.

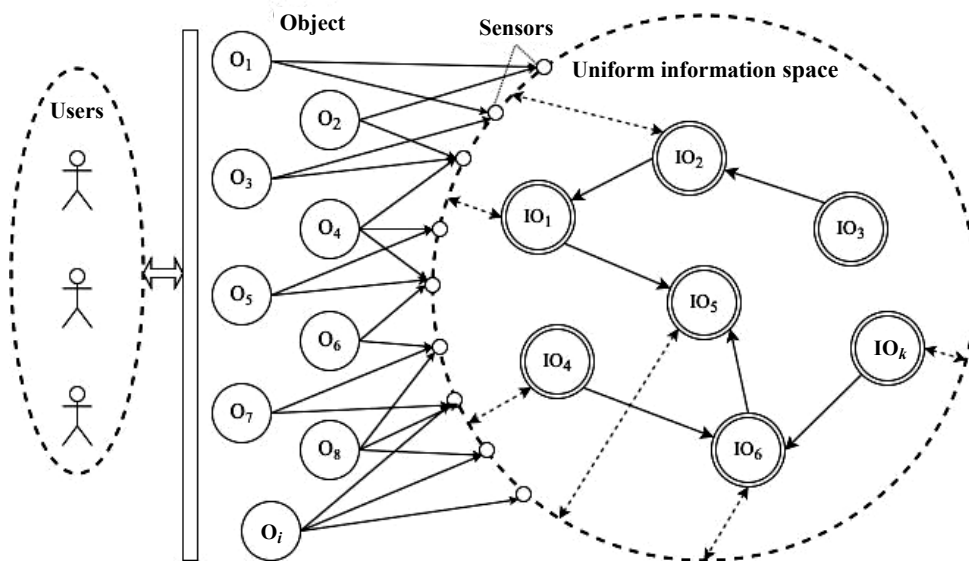


Fig. 1. General scheme of unified information space

In unified information space, the attributes of information objects must be defined in a single format and their number must be the same. Information objects should also be perceived unambiguously by users of unified information space. In fact, the input information in unified information space is heterogeneous in its presentation.

The task of internal mechanisms of an information computer system is to transform heterogeneous information coming in different formats and from different sources into a single set of information object attributes, by which unified information space users can uniquely identify an information object.

STAGES OF INFORMATION OBJECT IDENTIFICATION IN UNIFIED INFORMATION SPACE

Identification of an information object in unified information space makes it possible to unambiguously define it by the corresponding features. To identify objects, an identification method can be used which is based on a step-by-step analysis of the characteristics of an object using requests to the object in order to provide an opportunity to make decisions about its identification.

As it shows Fig.1 the external sources of information for unified information space generally represent objects O_1, O_2, \dots, O_m in different ways, where m is the number of objects operated by unified information space. Information about such objects is received as a set of feature values by reading them using sensors.

Also here is presented the users of unified information space, who's goal is to get some information about the objects there.

There are many information objects IO_1, IO_2, \dots, IO_k , each of which has a set of features P_1, P_2, \dots, P_n , where k is number of information objects, and n is number of an information object features, i.e., $IO_k (P_1, P_2, \dots, P_n)$.

In this case, it is considered that the same maximum number of features is set for each information object. Still the number of features for each specific object may be less than the maximum — some features may be absent (correspond to the NONE value), i. e. this object may simply not have this feature.

In unified information space each IO_k information object is different, that is, there are no two absolutely identical information objects:

$$IO_1 \neq IO_2 \neq IO_3 \neq \dots \neq IO_k.$$

Therefore, unified information space should work as a kind of reference system. For this you need to go through several stages:

1. To form unified information space. It will consist of a set of information objects, each of which is characterized by a set of features. Such information objects will differ from each other, that is, they will be unique. Information objects have connections, so they interact with each other. A connection is understood as the presence of parameters of another information object in the certain object, which are obtained as a result of their interaction with each other. Connections between information objects correspond to the “interaction trajectory”, which is determined by the pre-history of the of objects interactions with each other.

Unified information space is constantly being updated and trained. Formation of unified information space is the procedure of its training, that is, recognition of incoming objects without reference to a specific object.

2. Read the characteristics of the incoming object using sensors, which are a kind of meters. But it happens that sensors may not read some signs of information objects, for example, there will be no access, there will be no information (corresponds to the NULL value).

3. Classification (recognition) of an object. The system receives an object O_m with a set of features, the sensors should read their values, and unified information space should unambiguously answer the question of whether there is an information object in a unified information space with such feature values, or a new information object will be formed.

METHOD OF SEARCHING FOR MISSING FEATURES IN THE INFORMATION OBJECT

Information objects have a local feature memory. When a unique identification of an incoming object occurs, the values of its parameters are added to the memory of corresponding attributes of the information object. Then, from a set of values of each information object attribute, statistical characteristics that describe this attribute are determined — the mathematical expectation M and the variance D , and the more input objects are identified by unified information space, the more accurate they become.

Method of searching for missing features in an information object:

1. Formation of the objects interaction history. Information objects store history of interactions with each other and an information object can answer the question whether it interacted with another information object before, and if so, whether the trajectory of its interaction contains the needed feature. If a given information object does not have an answer to this question, then it refers to other information objects of unified information space.

2. Comparison on the basis of features. An object O_i with a set of parameters (P_1, P_2, \dots, P_n) comes to the input. If the value of each parameter falls within the permissible range of values for the corresponding attribute of a certain information object $(M - D \leq P_i \leq M + D)$, then unified information space uniquely identifies the incoming object, that is, $O_i = IO_k$.

3. Search for missing features. To search for missing features, an information object interacts with every other information object from unified information space. For this, a feature search will be used. This search is based on the trajectory of the objects interaction, and their combination allows gradually narrow the search area for the missing parameters of the objects. And so on, until all the missing features are filled in so that unified information space can uniquely identify the object O_i .

4. Clarification (recognition) of missing features. If there are not enough signs, then it is necessary to turn to other information objects of unified information space and request the necessary missing features based on the trajectory of information objects interaction. For this, it is necessary that each information object retains the trajectory of interaction, i.e. it actually has a kind of global

memory. With such unified information space formation, all trajectories of all information objects interaction are placed in a single data warehouse.

5. The result is displaying the found object, group of objects or establishing the fact that the analyzed object is new.

In brief, the method is the next: the identification of an information object is made according to certain external or internal characteristics of an information object, taking into account the interaction of an information object in a unified information space. To support these actions, each information object is supplied with a set of features that characterize the object to a certain extent. Due to this, the procedure for identifying an information object is reduced to a simple comparison of the characteristics of the information object with the characteristics of the incoming object. If the parameters of an information object in a necessary and sufficient degree coincide with the parameters of the incoming real object, then this object is considered as has been identified.

ORGANIZATION AND CONDUCTING EXPERIMENTAL RESEARCH

We will analyze efficiency of incoming objects search in unified information space.

For the experiments, unified information space of 20,000 information objects was formed.

The percentage of missing parameters in information objects (NONE) was 6%. After that, a single information space is rebuilt by eliminating duplicate information objects. A series of 20 experiments each was carried out, at certain probabilities (25, 20, 15, 10 and 5%, respectively) that the parameter would not be read by the sensors (NULL).

Experiment 1. Let's consider the case when each of 20,000 information objects is described by 7 parameters. Below is a snippet of 10 information objects (Table 1).

Table 1. Fragment of 10 information objects from unified information space

IO	P						
	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇
IO ₁	4 ± 0,5	3 ± 0,9	6 ± 0,7	7 ± 0,1	6 ± 0,9	9 ± 0,4	11 ± 0,6
IO ₂	5 ± 0,1	6 ± 0,6	6 ± 0,9	7 ± 0,4	6 ± 0,6	6 ± 0,1	9 ± 0,2
IO ₃	1 ± 0,3	6 ± 0,9	3 ± 0,2	5 ± 0,2	NONE	10 ± 0,4	8 ± 0,4
IO ₄	1 ± 0,7	6 ± 0,7	7 ± 0,3	6 ± 0,4	6 ± 0,5	NONE	10 ± 0,8
IO ₅	4 ± 0,4	6 ± 0,6	4 ± 0,4	7 ± 0,2	8 ± 0,3	6 ± 0,5	8 ± 0,4
IO ₆	5 ± 0,4	4 ± 0,8	6 ± 0,4	4 ± 0,5	5 ± 0,7	8 ± 0,8	9 ± 0,7
IO ₇	2 ± 0,3	3 ± 0,6	7 ± 0,4	NONE	7 ± 0,6	9 ± 0,7	8 ± 0,6
IO ₈	3 ± 0,4	5 ± 0,5	5 ± 0,4	6 ± 0,2	9 ± 0,6	NONE	7 ± 0,5
IO ₉	2 ± 0,9	2 ± 0,5	3 ± 0,2	4 ± 0,9	8 ± 0,2	6 ± 0,5	11 ± 0,2
IO ₁₀	1 ± 0,2	6 ± 0,7	6 ± 0,4	8 ± 0,7	9 ± 0,3	8 ± 0,3	11 ± 0,5

In this case, the interval length for each parameter was 5 units, for example, for the parameter P₁ [1; 6).

The following variants of the results were obtained:

1. When the sensors read all the values of the parameters of the incoming object and its unique identification has occurred:

New object:

5,8 6,5 6,2 4,8 9,6 9,5 9,7

Search object:

IO 06106 $5 \pm 0,3$ $6 \pm 0,4$ $6 \pm 0,7$ $4 \pm 0,9$ $9 \pm 0,4$ $9 \pm 0,6$ $9 \pm 0,2$

2. When the sensors read all the values of the incoming object parameters and its identification did not occur, there was no information object in unified information space that would describe this incoming object:

New object:

4,9 3,3 4,5 7,9 5,2 6,5 11,7

Object is absent!

In this case, this incoming object becomes a new information object and can be added to unified information space.

3. When the sensors did not read the values of all parameters of the incoming object (NULL), but due to the interaction of information objects with each other in unified information space, a unique identification of the incoming object took place:

New object:

5,2 Null 6,4 6,3 5,5 10,1 9,9

Search object:

IO 02727 $5 \pm 0,1$ $4 \pm 0,4$ $6 \pm 0,5$ $6 \pm 0,8$ $5 \pm 0,6$ $10 \pm 0,8$ $9 \pm 0,2$

IO 13394 $5 \pm 0,6$ $2 \pm 0,2$ $6 \pm 0,8$ $6 \pm 0,4$ $5 \pm 0,2$ $10 \pm 0,1$ $9 \pm 0,9$

IO 14824 $5 \pm 0,1$ $5 \pm 0,3$ $6 \pm 0,9$ $6 \pm 0,9$ $5 \pm 0,4$ $10 \pm 0,2$ $9 \pm 0,4$

ReCreateObject:

5,2 4,1 6,4 6,3 5,5 10,1 9,9

Search object:

IO 02727 $5 \pm 0,1$ $4 \pm 0,4$ $6 \pm 0,5$ $6 \pm 0,8$ $5 \pm 0,6$ $10 \pm 0,8$ $9 \pm 0,2$

4. When the sensors did not read the values of all parameters of the incoming object and, despite the interaction of information objects with each other in unified information space, identification of the incoming object did not occur:

New object:

2,1 4,5 3,5 Null 9,6 9,2 Null

Search object:

IO 08264 $2 \pm 0,3$ $4 \pm 0,2$ $3 \pm 0,7$ None $9 \pm 0,6$ $9 \pm 0,7$ $10 \pm 0,4$

IO 08473 $2 \pm 0,6$ $4 \pm 0,6$ $3 \pm 0,3$ $6 \pm 0,8$ $9 \pm 0,2$ $9 \pm 0,5$ $8 \pm 0,4$

IO 16500 $2 \pm 0,9$ $4 \pm 0,7$ $3 \pm 0,1$ $8 \pm 0,4$ $9 \pm 0,4$ $9 \pm 0,6$ $8 \pm 0,4$

ReCreateObject:

2,1 4,5 3,5 Null 9,6 9,2 8,5

Search object:

IO 08473 $2 \pm 0,6$ $4 \pm 0,6$ $3 \pm 0,3$ $6 \pm 0,8$ $9 \pm 0,2$ $9 \pm 0,5$ $8 \pm 0,4$

IO 16500 $2 \pm 0,9$ $4 \pm 0,7$ $3 \pm 0,1$ $8 \pm 0,4$ $9 \pm 0,4$ $9 \pm 0,6$ $8 \pm 0,4$

ReCreateObject:

2,1 4,5 3,5 5,1 9,6 9,2 8,5

Object absent!

5. When the sensors did not count values of one or several parameters and, after the interaction of information objects with each other in unified information space, it was found that this object does not have this feature (NONE).

New object:

2,7 4,7 4,1 8,3 5,5 Null 7,6

Search object:

IO 12602 $2 \pm 0,5$ $4 \pm 0,4$ $4 \pm 0,3$ $8 \pm 0,7$ $5 \pm 0,8$ $10 \pm 0,9$ $7 \pm 0,8$

ReCreateObject:

2,7 4,7 4,1 8,3 5,5 Null 7,6

ReCreateObject:

2,7 4,7 4,1 8,3 5,5 Null 7,6

ReCreateObject:

2,7 4,7 4,1 8,3 5,5 None 7,6

Object absent!

Generalization of the experiments results made it possible to conclude about search efficiency of incoming objects in unified information space with 7 parameters and an interval length of 5 units, which is presented in Table 2.

Table 2. Search efficiency of incoming objects in unified information space with 7 parameters and an interval length of 5 units

Probability that the parameter will not be read by sensors (NULL),%	Probability of an incoming object identification, %
5	20
10	10
15	25
20	15
25	15

Table 2 shows that search efficiency of incoming objects in unified information space with 7 parameters and an interval length of 5 units is low.

Therefore, it was decided to reduce number of parameters that describe the object, and conduct similar experiments with 4, 5 and 6 parameters with the same initial data. The results of the experiments are presented in Table 3.

Table 3. Search efficiency of incoming objects in unified information space with 4, 5 and 6 parameters and an interval length of 5 units

Probability that the parameter will not be read by sensors (NULL), %	Probability of an incoming object identification depending on the number of parameters, %		
	4	5	6
5	100	100	65
10	100	100	45
15	100	90	85
20	100	100	75
25	100	100	50

Based on the data in Tables 2 and 3, a graph was built for comparing search efficiency of incoming objects in unified information space with 4, 5, 6 and 7 parameters and an interval length of 5 units (Fig. 2).

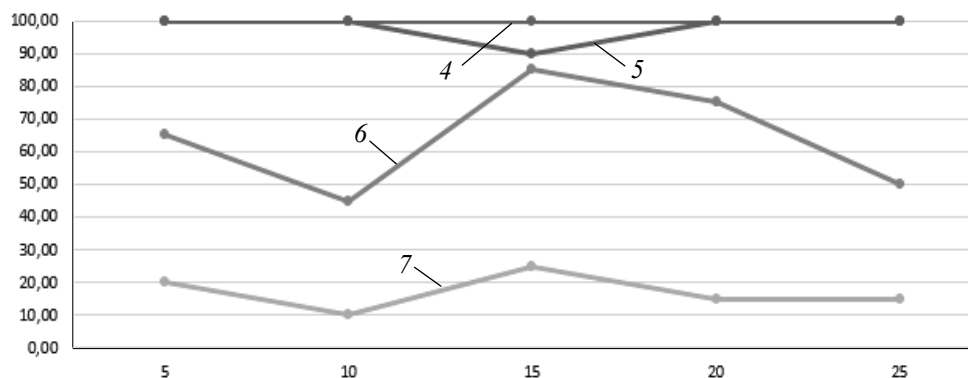


Fig. 2. Comparative graph of search efficiency of incoming objects in unified information space with 4, 5, 6 and 7 parameters and an interval length of 5 units

From Fig. 2, we can conclude that search efficiency of incoming objects in unified information space with an interval length of 5 units was on average: with 4 parameters — 100%, with 5 parameters — 98%, with 6 parameters — 64% and with 7 parameters — 17%.

Accordingly, a decrease in the number of parameters leads to a sharp increase in search efficiency of incoming objects in unified information space.

Experiment 2. When analyzing the conclusions of experiment 1, it was decided to change the lengths of the values intervals of the parameters of the incoming objects in unified information space with a constant value of parameters number.

Let us consider the case when each of 20,000 information objects is described by 7 parameters, but with different interval lengths of parameter values of 3, 4, and 5 units. The results of the experiments are presented in table 4.

Table 4. Search efficiency of incoming objects in unified information space with 7 parameters and an interval length of 3, 4 and 5 units

Probability that the parameter will not be read by sensors (NULL),%	Probability of an incoming object identification depending on the interval length, %		
	3	4	5
5	100	75	20
10	100	40	10
15	95	50	25
20	100	65	15
25	100	50	15

Based on the data in Table 4, a graph was built for comparing search efficiency of incoming objects in unified information space with 7 parameters and an interval length of 3, 4, and 5 units (Fig. 3).

From Fig. 3, we can conclude that search efficiency of incoming objects in unified information space with 7 parameters on average was: with an interval length of 3 units — 99%, 4 units — 56%, 5 units — 17%.

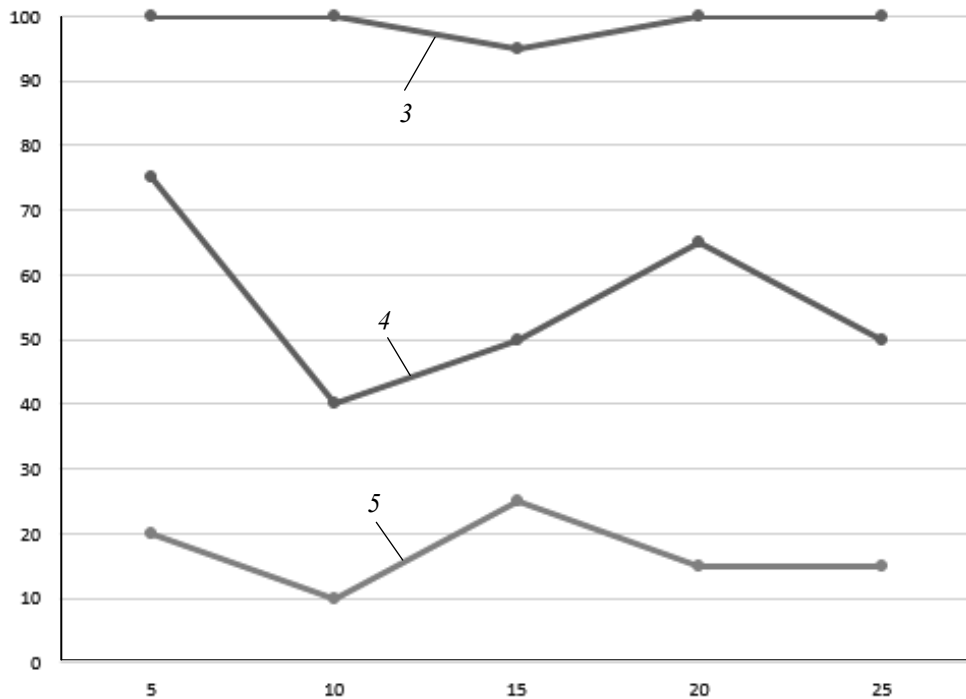


Fig. 3. Comparative graph of search efficiency of incoming objects in unified information space with 7 parameters and an interval length of 3, 4 and 5 units

Accordingly, an increase in the interval length leads to a sharp decrease in search efficiency of incoming objects in unified information space.

Experiment 3. When analyzing the conclusions of experiment 2, it was decided to conduct similar experiments with a small interval length, but a larger number of incoming objects parameters in unified information space.

Let us consider the case when each of 20,000 information objects is described by 8, 9, and 10 parameters with a length of parameter value intervals being 3 units. The results of the experiments are presented in Table 5.

Table 5. Efficiency of searching for incoming objects in unified information space with 8, 9 and 10 parameters with a length of parameter value intervals of 3 units

Probability that the parameter will not be read by sensors (NULL),%	Probability of an incoming object identification depending on the number of parameters, %			
	7	8	9	10
5	100	85	35	25
10	100	80	40	5
15	95	90	35	10
20	100	90	45	25
25	100	95	40	15

Based on the data in Tables 4 and 5, a graph was constructed for comparing searching efficiency of incoming objects in unified information space with 7, 8, 9 and 10 parameters and the length of the parameter values interval of 3 units (Fig. 4).

From Fig. 4, we can conclude that search efficiency of incoming objects in unified information space with length of parameter value interval of 3 units on average was: with 7 parameters — 99%, with 8 parameters — 88%, with 9 parameters — 39%, and with 10 parameters — 16%.

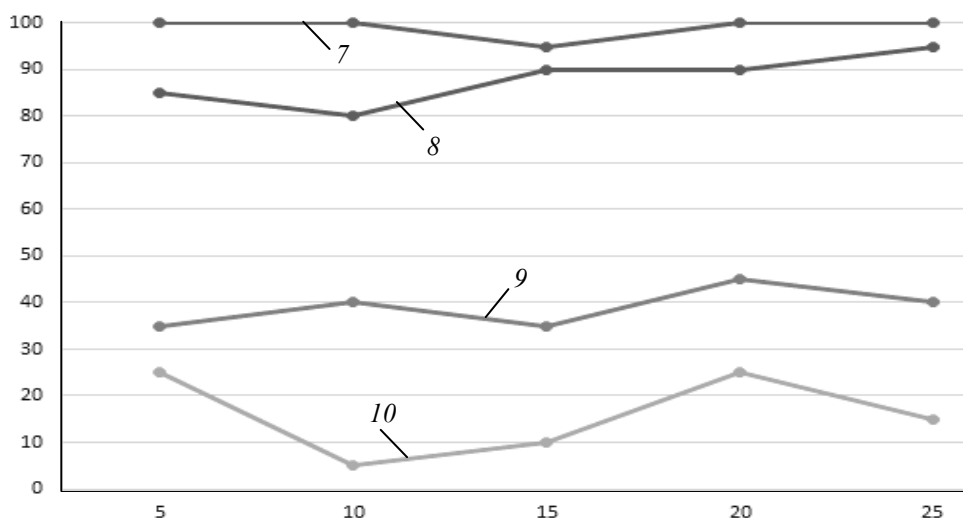


Fig. 4. Comparative graph of search efficiency of incoming objects in unified information space with 7, 8, 9 and 10 parameters and length of the parameter value interval of 3 units

Accordingly, search efficiency of incoming objects in unified information space is the highest with 7 and 8 parameters and with length of the parameter values interval of 3 units.

CONCLUSIONS

The article discusses formation of unified information space using a specialized information computer system, which is actually a hardware and software basis for supporting a single information space.

The stages of information object identification in unified information space are considered. The article proposes the method of searching for missing features of the incoming object by implementing information objects interaction with each other within unified information space.

The experiments described in the article make it possible to evaluate search efficiency of incoming objects in unified information space when the number of incoming parameters and interval of their values change. The experiments have shown that the identification probability depends significantly on the number of parameters of the original object, as well as on the length of the intervals describing values of the object parameters themselves. At the same time, with an increase in number of original object parameters and the interval length of object parameters, search efficiency of incoming objects in unified information space significantly decreases.

Thus, a promising direction of research is the development of specialized methods for identifying objects in unified information space, which will improve object identification efficiency in conditions of an increase in number of the original object parameters and the interval length of object parameters.

REFERENCES

1. V. Mukhin et al., "Method of Restoring Parameters of Information Objects in a Unified Information Space Based on Computer Networks", *International Journal of Computer Network and Information Security*, vol. 12, no. 2, pp.11–21, 2020. doi: 10.5815/ijcnis.2020.02.02
2. Т.А. Ожерельева, "Об отношении понятий информационное пространство, информационное поле, информационная среда и семантическое окружение", *Международный журнал прикладных и фундаментальных исследований*, № 10, с. 21–24, 2014.
3. V.Ya. Tsvetkov, "Information field", *Life Science Journal*, 11(5), pp. 551–554, 2014.
4. В.Г. Бондур, "Информационные поля в космических исследованиях", *Образовательные ресурсы и технологии*, №2 (10), с. 107–113, 2015.
5. V. Mukhin, A. Volokyta, Y. Heriatovych, and P. Rehida, "Method for Efficiency Increasing of Distributed Classification of the Images based on the Proactive Parallel Computing Approach", *Advances in Electrical and Computer Engineering*, no. 18(2), pp. 117–122, May 2018. doi: 10.4316/AECE.2018.02015
6. K. Smelyakov, M. Shuplyiuk, V. Martovytskyi, D. Tovchyrechko, and O. Ponomarenko, "Efficiency of Image Convolution", *2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL), 6-8 Sept. 2019, Sozopol, Bulgaria*, pp. 578–583.
7. K. Smelyakov, O. Ponomarenko, A. Chupryna, D. Tovchyrechko, and I. Ruban, "Local Feature Detectors Performance Analysis on Digital Image", *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), 8–11 Oct. 2019, Kyiv, Ukraine*, pp. 644–648.
8. K. Smelyakov, A. Chupryna, M. Hvozdiev, D. Sandrkin, and V. Martovytskyi, "Comparative efficiency analysis of gradational correction models of highly lighted image", *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), 8–11 Oct. 2019, Kyiv, Ukraine*, pp. 703–708.
9. Z. Hu, V. Mukhin, Ya. Kornaga, A. Volokyta, and O. Herasymenko, "The scheduler for distributed computer system based on the network-centric approach to resources control", *Proc. of the 2017 IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 21–23 Sept. 2017, Bucharest, Romania*, vol. 1, pp. 518–523. doi: 10.1109/IDAACS.2017.8095135
10. А.Н. Тихонов, А.Д. Иванников, И.В. Соловьёв, В.Я. Цветков, и С.А. Кудж, *Концепция сетцентрического управления сложной организационно-технической системой*. Москва: МАКС Пресс, 2010, 135 с.
11. V.Ya. Tsvetkov, "Information interaction", *European Researcher*, vol. 62, no. 11-1, pp. 2573–2577, 2013.
12. V.Ya. Tsvetkov, "Evaluations of Information Assymetry", *Modern Applied Science*, vol. 9, no. 6, pp. 225–261, 2015. doi:10.5539/mas.v9n6p255
13. Z. Hu, V. Mukhin, Ya. Kornaga, O. Herasymenko, and Y. Mostoviy, "The Analytical Model for Distributed Computer System Parameters Control Based on Multi-factoring Estimations", *Journal of Network and Systems Management*, no. 27 (2), pp. 351–365, 2019.
14. V. Mukhin, Y. Kornaga, V. Zavgorodnii, A. Zavgorodnya, O. Herasymenko, and O. Mukhin, "Social Risk Assessment Mechanism Based on the Neural Networks", *International Conference on Advanced Trends in Information Theory (ATIT-2019), 18–20 Dec. 2019*, pp. 179–182.
15. H. Zhenbing, V. Mukhin, Ya. Kornaga, and O. Herasymenko, "Resource Management in a Distributed Computer System with Allowance for the Level of Trust to Computational Components", *Cybernetics and Systems Analysis*, no. 53 (2), pp. 312–322, 2017.
16. V. Mukhin, Ya. Kornaga, Y. Mostovyi, and Y. Bazaka, "A Model For Events Monitoring Heterogeneous Distributed Databases Based on Vector-matrix Operations",

The Far East Journal of Electronics and Communications, vol. 16, issue 3, pp. 645–656, 2016.

17. В.Я. Цветков, “Паралингвистические информационные единицы в образовании”, *Перспективы науки и образования*, № 4, с. 30–38, 2013.
18. А.Н. Тихонов, А.Д. Иванников, И.В. Соловьёв, и В.Я. Цветков, *Основы управления сложной организационно-технической системой. Информационный аспект*. Москва: МаксПресс, 2010, 228 с.
19. Ю.И. Синещук [и др.], “Основные угрозы и направления обеспечения безопасности единого информационного пространства”, *Вестн. С.-Петерб. ун-та МВД*, № 2, с. 150–154, 2013.

Received 21.01.2021

INFORMATION ON THE ARTICLE

Alexander G. Dodonov, Institute of Problems of Information Registration of the National Academy of Sciences of Ukraine, Ukraine, e-mail: dodonovua@gmail.com

Vadym E. Mukhin, ORCID: 0000-0002-1206-9131, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: v.mukhin@kpi.ua

Valerii V. Zavgorodnii, State University of Infrastructure and Technology, Ukraine, e-mail: zavgorodniivalerii@gmail.com

Yaroslav I. Kornaga, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: slovyan_k@ukr.net

Anna A. Zavgorodnya, State University of Infrastructure and Technology, Ukraine, e-mail: annzavgorodnya@gmail.com

МЕТОД ПОИСКА ИНФОРМАЦИОННЫХ ОБЪЕКТОВ В ЕДИНОМ ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ / А.Г. Додонов, В.Е. Мухин, В.В. Завгородний, Я.И. Корнага, А.А. Завгородняя

Аннотация. Рассмотрены понятие и принципы построения единого информационного пространства и представлена схема его формирования. Рассмотрено формирование единого информационного пространства с использованием специализированной информационной компьютерной системы, которая фактически является аппаратно-программным базисом для поддержки единого информационного пространства, а также этапы идентификации информационного объекта в едином информационном пространстве. Предложен метод поиска недостающих признаков входящего объекта путем реализации взаимодействия информационных объектов между собой внутри единого информационного пространства.

Ключевые слова: единое информационное пространство, информационный объект, признаки, идентификация объекта, метод поиска.

МЕТОД ПОШУКУ ІНФОРМАЦІЙНИХ ОБ’ЄКТІВ В ЄДИНОМУ ІНФОРМАЦІЙНОМУ ПРОСТОРІ / О.Г. Додонов, В.Є. Мухін, В.В. Завгородній, Я.І. Корнага, Г.А. Завгородня

Анотація. Розглянуто поняття і принципи побудови єдиного інформаційного простору і подано схему його формування. Розглянуто формування єдиного інформаційного простору з використанням спеціалізованої інформаційної комп’ютерної системи, яка фактично є аппаратно-програмним базисом для підтримання єдиного інформаційного простору, а також етапи ідентифікації інформаційного об’єкта в єдиному інформаційному просторі. Запропоновано метод пошуку відсутніх ознак вхідного об’єкта шляхом реалізації взаємодії інформаційних об’єктів між собою всередині єдиного інформаційного простору.

Ключові слова: єдиний інформаційний простір, інформаційний об’єкт, ознаки, ідентифікація об’єкта, метод пошуку.

REFERENCES

1. V. Mukhin, "Method of Restoring Parameters of Information Objects in a Unified Information Space Based on Computer Networks", *International Journal of Computer Network and Information Security*, vol.12, no.2, pp.11–21, 2020. doi: 10.5815/ijenis.2020.02.02
2. T.A. Ozherelieva, "On the relationship between the concepts of information space, information field, information environment and semantic environment", *International Journal of Applied and Fundamental Research*, no. 10, pp. 21–24, 2014.
3. V.Ya. Tsvetkov, "Information field", *Life Science Journal*, 11(5), pp. 551–554, 2014.
4. V.G. Bondur, "Information fields in space research", *Educational resources and technologies*, no. 2 (10), pp. 107–113, 2015.
5. V. Mukhin, A. Volokyta, Y. Heriatovych, and P. Rehida, "Method for Efficiency Increasing of Distributed Classification of the Images based on the Proactive Parallel Computing Approach", *Advances in Electrical and Computer Engineering*, no. 18(2), pp. 117–122, May 2018. doi: 10.4316/AECE.2018.02015
6. K. Smelyakov, M. Shupyliuk, V. Martovytskyi, D. Tovchyrechko, and O. Ponomarenko, "Efficiency of Image Convolution", *2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL)*, 6–8 Sept. 2019, Sozopol, Bulgaria, pp. 578–583.
7. K. Smelyakov, O. Ponomarenko, A. Chupryna, D. Tovchyrechko, and I. Ruban, "Local Feature Detectors Performance Analysis on Digital Image", *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)*, 8–11 Oct. 2019, Kyiv, Ukraine, pp. 644–648.
8. K. Smelyakov, A. Chupryna, M. Hvozdiev, D. Sandrkin, and V. Martovytskyi, "Comparative efficiency analysis of gradational correction models of highly lighted image", *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)*, 8–11 Oct. 2019, Kyiv, Ukraine, pp. 703–708.
9. Z. Hu, V. Mukhin, Ya. Kornaga, A. Volokyta, and O. Herasymenko, "The scheduler for distributed computer system based on the network-centric approach to resources control", *Proc. of the 2017 IEEE 9th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 21–23 Sept. 2017, Bucharest, Romania, vol. 1, pp. 518–523. doi: 10.1109/IDAACS.2017.8095135
10. A.N. Tikhonov, A.D. Ivannikov, I.V. Soloviev, V.Ya. Tsvetkov, and S.A. Kudzh, *The concept of network-centric control of a complex organizational and technical system*. Moscow: MAKS Press, 2010, 135 p.
11. V.Ya. Tsvetkov, "Information interaction", *European Researcher*, vol. (62), no. 11-1, pp. 2573–2577, 2013.
12. V.Ya. Tsvetkov, "Evaluations of Information Asymmetry", *Modern Applied Science*, vol.9, no. 6, pp. 225–261, 2015. doi:10.5539/mas.v9n6p255
13. Z. Hu, V. Mukhin, Ya. Kornaga, O. Herasymenko, and Y. Mostoviy, "The Analytical Model for Distributed Computer System Parameters Control Based on Multi-factoring Estimations", *Journal of Network and Systems Management*, no. 27 (2), pp. 351–365, 2019.
14. V. Mukhin, Y. Kornaga, V. Zavgorodnii, A. Zavgorodnya, O. Herasymenko, and O. Mukhin, "Social Risk Assessment Mechanism Based on the Neural Networks", *International Conference on Advanced Trends in Information Theory (ATIT-2019)*, 18–20 Dec. 2019, pp. 179–182.
15. H. Zhenbing, V. Mukhin, Ya. Kornaga, and O. Herasymenko, "Resource Management in a Distributed Computer System with Allowance for the Level of Trust to Computational Components", *Cybernetics and Systems Analysis*, no. 53 (2), pp. 312–322, 2017.
16. V. Mukhin, Ya. Kornaga, Y. Mostovyi, and Y. Bazaka, "A Model For Events Monitoring Heterogeneous Distributed Databases Based on Vector-matrix Operations", *The Far East Journal of Electronics and Communications*, vol. 16, issue 3, pp. 645–656, 2016.
17. V.Ya. Tsvetkov, "Paralinguistic information units in education", *Perspectives of Science and Education*, no. 4, pp. 30–38, 2013.
18. A.N. Tikhonov, A.D. Ivannikov, I.V. Soloviev, and V.Ya. Tsvetkov, *Fundamentals of control of a complex organizational and technical system. Informational aspect*. Moscow: MaxPress, 2010, 228 p.
19. Yu.I. Sineschuk [and others], "The main threats and directions for the security ensuring in a single information space", *Bulletin of St. Petersburg. University of the MIA*, no 2, pp. 150–154, 2013.

ONTOLOGICAL MODEL FOR DATA PROCESSING ORGANIZATION IN INFORMATION AND COMMUNICATION NETWORKS

L.S. GLOBA, N.A. GVOZDETSKA, R.L. NOVOGRUDSKA

Abstract. The functioning of modern information and communication networks is impossible without data processing. With the emergence of new network services, the amount of information that needs to be processed increases, while the requirements to the data processing quality become more and more stringent. Therefore, the problem of designing and maintaining a scalable data processing system with a flexible quality of service management is becoming more and more important for a network operator. Such data processing systems have a complex internal structure with many interrelated parameters, which makes them difficult to analyze, manage, and expand. This study proposes to use an ontological model to store, represent, and manipulate the information in the operator's data processing system. The ontological model allows to structure and systematize the data of an information processing system, and transparently reflects the relationships between the parameters of the system to simplify its analysis and scaling. The proposed ontology of a data processing system consists of three related subsystems. The paper describes the proposed ontological model and additionally analyzes the sources of information that needs to be processed in the information and communication network.

Keywords: information and communication network, data processing system, ontology, model, network operator, analysis, scaling, class, relations.

INTRODUCTION

Until the last decade, the term “communication network” primarily meant a set of technical means of communications and facilities designed for routing, switching, transmission, and/or reception of signals between the terminal equipment [1]. However, in modern communication networks, the operation of technical means of communications is not possible without software. Moreover, a big part of hardware functionality is being replaced by software due to the convergence of information and communications technologies. Communication network becomes a cyber-physical system, where the physical and software network components are deeply intertwined. Therefore, the term “information and communications network” is used instead today to emphasize the fact that the set of information and communications systems act as a whole within the information processing. Having an extremely complex internal structure, such a cyber-physical network can not be easily controlled by the man without using the approaches to intellectual data analysis.

In information and communications network data processing plays an essential role. To ensure efficient data processing, telecom operators must manage their data processing system wisely taking into account multiple criteria. Therefore, the relevant problem for the telecom operator is to represent his data processing system in a way to simplify its analysis, maintenance and operation, and to add sup-

port for the system's scalability. To reach these objectives, the system should be represented in a formalized way, so that the parameters of the system are explicitly exposed, the internal dependencies within the structure are easy to analyze and the new component may be easily added to support the system's scalability.

There exist multiple approaches to complex systems' formalization. In particular, the system may be represented with the help of the relational model [2], object-oriented model [2, 3], network-based model [4], ontological model [5, 6]. The relational model focuses on organizing data in two-dimensional tables. The advantage of the relational data model lies in the simplicity and convenience of physical implementation on a computer. The main disadvantages of the relational model include the lack of standard means of identifying individual records and the complexity of describing hierarchical and network relationships. The object-oriented data representation model operates with concepts such as class and object. Classes define a data structure and represent a set of attributes (text string, integer, image, etc.). Instances of a class (objects) have a certain structure and can contain other objects, forming an arbitrary hierarchical structure. As a rule, systems based on an object-oriented data model are functional, flexible, but at the same time, more complex. Being hierarchical, these systems have a limitation: the use case, when a child class has more than one parent class is not supported here. To tackle this problem, the network-based model exists. In this model, the classes and relationships are represented in a form of a graph, which makes this model much more flexible. Ontology is an attempt at a comprehensive and detailed formalization of a certain area of knowledge using a conceptual scheme. By the ontological model, data can be represented as a set of different types of information objects and links between them. The main advantage of the ontological model is its comprehensiveness. In contrast to the object-oriented and network-based model, not only the classes and objects are in focus here, but also the complex semantic relations between them. It helps to describe the system more exhaustively and explicitly. Because of these advantages, we have chosen an ontological model to help the operator to simplify the management and analysis of his data processing system.

In this paper, the usage areas of data processing within the activity of the information and communication network operator are analyzed. In order to simplify the data processing organization, an ontology of the operator's data processing system is proposed. It formalizes the data processing system and exposes its parameters, allowing the operator to analyze and manage the data processing infrastructure with less effort. The paper is structured as follows: in Section 2 we provide an overview of the related work. Section 3 is dedicated to the analysis of the main data sources in the operator's network to reveal the importance of data processing in modern information and communication networks. Section 4 describes the proposed ontological model which aims at simplifying the analysis, management, and scaling of the data processing system. Section 5 provides an example of simplified data processing system analysis. The results of the paper are summarized in the Conclusion.

RELATED WORKS

There exist a number of approaches to formalizing a communication network with the help of an ontological model. The authors of [7] proposed a general ontological model that describes a semantic of the domains relevant for the Next Genera-

tion Networks (NGN). Their main idea was to introduce a central point ontology (core ontology) that defines the main concepts of the mobile domain. Except for the core ontology, the authors generally described a number of sub-ontologies that need to be further refined and extended to further domains. Among them is a communication resources sub-ontology. In this paper we are focused on the data processing system as a communication resource of the network and thus we aim at refining this part of the ontological model proposed in [7]. However, from our point of view the authors did not pay enough attention to the quality of service assessment in the communication network, although this aspect itself contains a large number of concepts and relations that should be considered and formalized.

In [8] an approach to ontology modelling for telecommunications service domain was proposed. Within their telecommunications service domain ontology (TSDO), the authors distinguish the Terminal Capability Ontology, Network Ontology, Service Role Ontology, Service Category Ontology, Charging Ontology and Service Quality Ontology. From our point of view, the main advantage of the proposed concept is taking into account the quality of service parameters as a separate full-fledged ontology. Meanwhile, the authors paid attention to formalizing the communication resource domain as well. In our work we would like to elaborate on this concept and describe the network resources in a relation with its quality parameters in particular. Since the service quality is influenced not only by the network resources but also by the various workloads which have to be processed in the network, in contrast to [8] we would like to pay separate attention to the semantic interoperability of all three domains: network resources, service quality, and workload.

According to the 5G whitepaper [9], in new networks along with traditional Quality of Service parameters such as the packet error rate, transmission latency, and data rate, the new parameters such as network energy efficiency are becoming more and more important. Energy efficiency of data processing in general is a highly relevant topic. According to [10] the amount of power consumed by the data processing facilities around the world comprises near 2% of all electrical power produced worldwide. In order to deal with this problem a number of hardware and software energy efficient approaches to data processing were proposed. Among them are the virtual machines consolidation [11], energy efficient scheduling [12, 13], resource scaling [14]. These approaches have already become standard for the distributed data processing facilities, however, they did not get enough attention in the communication network domain. Of course, energy efficient approaches are used there, but from our point of view, they must be included in the general formalized data processing architecture. In this paper, we try to approach this problem with the proposed ontological model of the data processing system in communications.

DATA PROCESSING RESOURCES IN INFORMATION AND COMMUNICATION NETWORKS

According to the authors of [15], information and communication network resources are divided into information, data processing and storage resources, software. Information resources are information and knowledge transmitted through the information and communication network. Data processing and storage resources are the performance of processors and the amount of memory of com-

puters running on the network, as well as the time during which they are used. Software resources include network operating systems, server software, workstation software, application software, traffic analyzers, network controls, and more. Communication resources are resources that are involved in the transportation and redistribution of information flows in a network. It means that the data processing and storage resources have already become an essential part of modern communication networks.

Alongside with the access and core network, data centers (representing the data processing and storage resources) are becoming the key components of the infrastructure of the communication network operator [16]. Let us briefly overview these key components of the network. Access network is connected to the end (terminal) nodes — equipment installed by users of the network. For example, in the case of building an operator network to provide Internet access services, the end nodes may be subscribers' computers or subscriber routers. The main purpose of the access network is to concentrate the information flows coming through numerous communication channels from user equipment. The core network combines individual access networks, performing the functions of traffic transit between them through high-speed channels. Data centers and service management centers are network resources on the basis of which customer service is provided. Such centers can store information of two types:

- user information, i.e., information that is of direct interest to end users of the network (information resources);
- service information that helps provide services to users.

Examples of the first-type information resources are web-portals, which contain a variety of reference and news information, information from e-shops, etc. Resources of the second type are various systems of authentication and authorization allowing the operator to check the rights of users for receiving the services; billing systems, which help to manage charges for services in commercial networks; databases of user credentials that store usernames and passwords, as well as lists of services to which each user is subscribed, etc. The second type of resources should also include a centralized network management system.

A prominent example of an information and communication network in which information and communication technologies work as a single indivisible whole is the 5G network. Let us briefly consider the basic principles of this network design in order to reveal the additional purposes of data processing in 5G networks. To do this, let us analyze the white paper provided by the European organization 5G PPP (public-private partnership in the field of 5G infrastructure — a joint initiative between the European Commission and the European ICT industry (ICT manufacturers, communications operators, service providers, SMEs and research institutions). According to the documentation [9], the key paradigm of the 5th generation mobile network is its programmability. Programmability ensures flexible network adaptation at various levels, including infrastructure, network functions, services, and applications. In particular programmability in the data plane, transport network (core network) and access network are distinguished. In 5G programmability is primarily inspired by two technologies: SDN (Software Defined Network) and NFV (Network Function Virtualization) technologies.

SDN is an approach to network design, implementation, and management that separates network management (control plane) and traffic management process (data plane). This separation greatly simplifies network administration and

management, as the control plane processes only information related to the logical topology of the network. The data plane instead organizes network traffic according to the configuration set in the control plane. Unlike conventional IP networks, whose functions are decentralized, SDN follows a centralized approach [17].

The main idea of NFV technology is to replace specialized network equipment (e.g., L2 switches, routers, NAT devices (Network Address Translation), firewalls, etc.) with software — virtualized network functions — consolidated on general-purpose hardware (commodity servers) [18].

Thus, to enable these technologies communication operators must maintain the data processing infrastructure as well.

Another purpose of maintaining the data processing infrastructure for the telecom operator is a Big Data analysis. Such an analysis helps operators to improve the technical and economic parameters of the network, enable the personalization of telecom services, and ensure more efficient allocation of funds. Data processing infrastructure is used in particular for the analysis of the subscribers' activity, traffic changes, long-term network characteristics, etc. The operators report the positive impact of the Big Data analysis included in their operational workflow [19, 20].

Summarizing this analysis, we would like to highlight 3 main areas of use of data processing by a telecom operator:

- 1) ensuring the functioning of SDN and NFV technologies;
- 2) implementation of such necessary network functions as authorization and authentication of users, billing service, etc.;
- 3) analysis of Big Data in the field of communications.

Thus, the effective design, construction and operation of data processing infrastructure is an important problem for operators of modern communications networks (including 5G networks).

ONTOLOGICAL MODEL OF A DATA PROCESSING SYSTEM IN INFORMATION AND COMMUNICATION NETWORKS

Computational ontologies are the means to formally model the structure of a system, i.e., the relevant entities and relations that emerge from its observation, and which are useful to our purposes [5]. Ontologies (or ontological models) help to formalize the structure of a system in order to simplify its management, improve its design, automatize system's operation, etc. In this paper we introduce the ontological model for the data processing system of the telecom operator to simplify its management. This simplification is a consequence of considering the system as a formalized structure with explicitly exposed parameters.

Being a structured representation of the information in some subject area, every ontology is based on the raw data, stored in some kind of informational database (Fig. 1). Examples of such raw data are the subscribers' records in a billing system, statistics of the daily workload in a data processing system, nominal parameters of the data processing equipment, etc. The ontological model transforms this raw data into knowledge. Modeling the data processing system as an ontology, we may distinguish 3 separate structural parts of the system. These parts correspond to the 3 subcomponents of the ontological model respectively:

- 1) the ontology of the processing system components (servers, processing software, etc.);

- 2) the ontology of the workloads (the input workload sources and parameters, etc.);
- 3) the ontology of the system quality assessment criteria (performance, energy efficiency, etc).

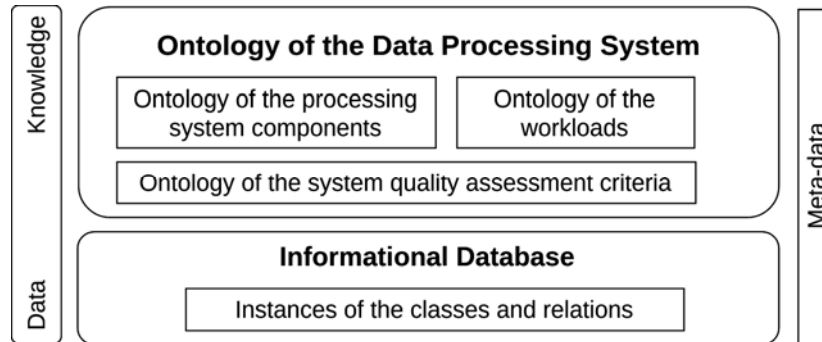


Fig. 1. General structure of the proposed ontological model of the data processing system

Dividing general ontology of Data Processing System in to such correlated components allows to describe in detailed way each subprocess that is performed be network operator while analyzing and managing data processing infrastructure with less effort.

Let us explain these subsystems. Data processing equipment and software represent the *ontology of processing system components*. The ontology of these components reflects the physical structure of a system itself. The *ontology of the workloads* encompasses the input workload and parameters of the available data processing equipment. The input workload comes from the aforementioned sources (we consider either Big Data tasks, subscribers' data or NFV processing tasks to be the sources here). Computational job is a unit of workload. Each job is characterized by its requirements which depend on the type of the workload source. The processing system consists of server clusters, consisting of N physical nodes (servers) respectively. Each server has an amount of computational resources (processing cores, RAM, etc.) and operates with specialized data processing software, aimed to manage the resources of the processing system, schedule the computational jobs, etc.

The output parameters are those criteria that are evaluated for the system. For the efficient data processing, telecom operator needs to design his data processing systems in a way to:

- fulfill the *Quality of service requirements* depending on the workload type (e.g., the allowed probability of a job loss for the internal Big Data processing tasks may be higher than the probability of loss for the billing service task, since it presumes direct cooperation with the subscriber);
- ensure sufficient *performance* of the data processing system (with “performance” here we mean the amount of processed data per time unit. I.e., the bigger performance of the data processing system is, the higher processing throughput is achieved);
- increase the *energy efficiency* of the data processing system. Since it greatly affects the operator's OPEX (Operating expense), so it is important to consider this criterion not only during system design (to purchase more energy efficient equipment), but also during operation.

These parameters are described by the *ontology of the system quality assessment criteria*.

Formally, the ontology may be specified as a set [6]:

$$O = \{C, A, R, T, F, D\},$$

where

- 1) C is the set of classes that describes the notions of a subject domain;
- 2) A is the set of attributes that describes the features of notions and relations;
- 3) R is the set of relations specified for classes:

$$R = \{R_{AS}, R_{IA}, R_n, R_{CD}\},$$

where R_{AS} is the associative relation:

$$R_{AS}(O) = \{C_i(O) \times C_j(O), M(R_{AS}) = \{str\}\},$$

where M is a type of relation meaning,

R_{IA} is the relation “is-are”, also known as a “part-whole” relation:

$$R_{IA}(O) = C_k(O) \subset C_m(O),$$

R_n is the relation of inheritance:

$$R_n(O) = a_i, r_i \vee A_{C_m}(O) \rightarrow a_i, r_i \vee A_{C_k}(O),$$

R_{CD} is the relation “class-data”:

$$R_{CD}(O) = C_j(O) \subseteq D_i(O);$$

- 4) T is the set of standard types of attribute values;
- 5) F is the set of limits for values of attribute notions and relations;
- 6) D is the set of instances for a particular class.

The proposed ontology of the telecom operator’s data processing system is represented in Fig. 2. The formal description of the ontology may be found below.

Proposed ontology is formally described as follows:

Set of ontology classes: $C = \{C_1, C_2, \dots, C_{25}\}$

C_1 — Operator’s data processing system. This class describes the concept of an operator’s data processing system as a physical entity.

C_2 — Source of the workload. This class includes concepts related to the input workload for the data processing system. As discussed in Section 1, possible sources (at least those considered in this paper) are.

C_3 — Operator’s Big Data.

C_4 — Subscribers’ data (e.g., credentials, data about the usage of services, billing information, etc.).

C_5 — NFV processing task. Within the NFV concept, the tasks usually performed by the specialized hardware (e.g., traffic routing, NAT, etc.) are performed on the commodity servers. Within this ontology we call these tasks generally “NFV processing tasks”, however this entity may be refined with respect to the concrete virtualized functions in the considered network.

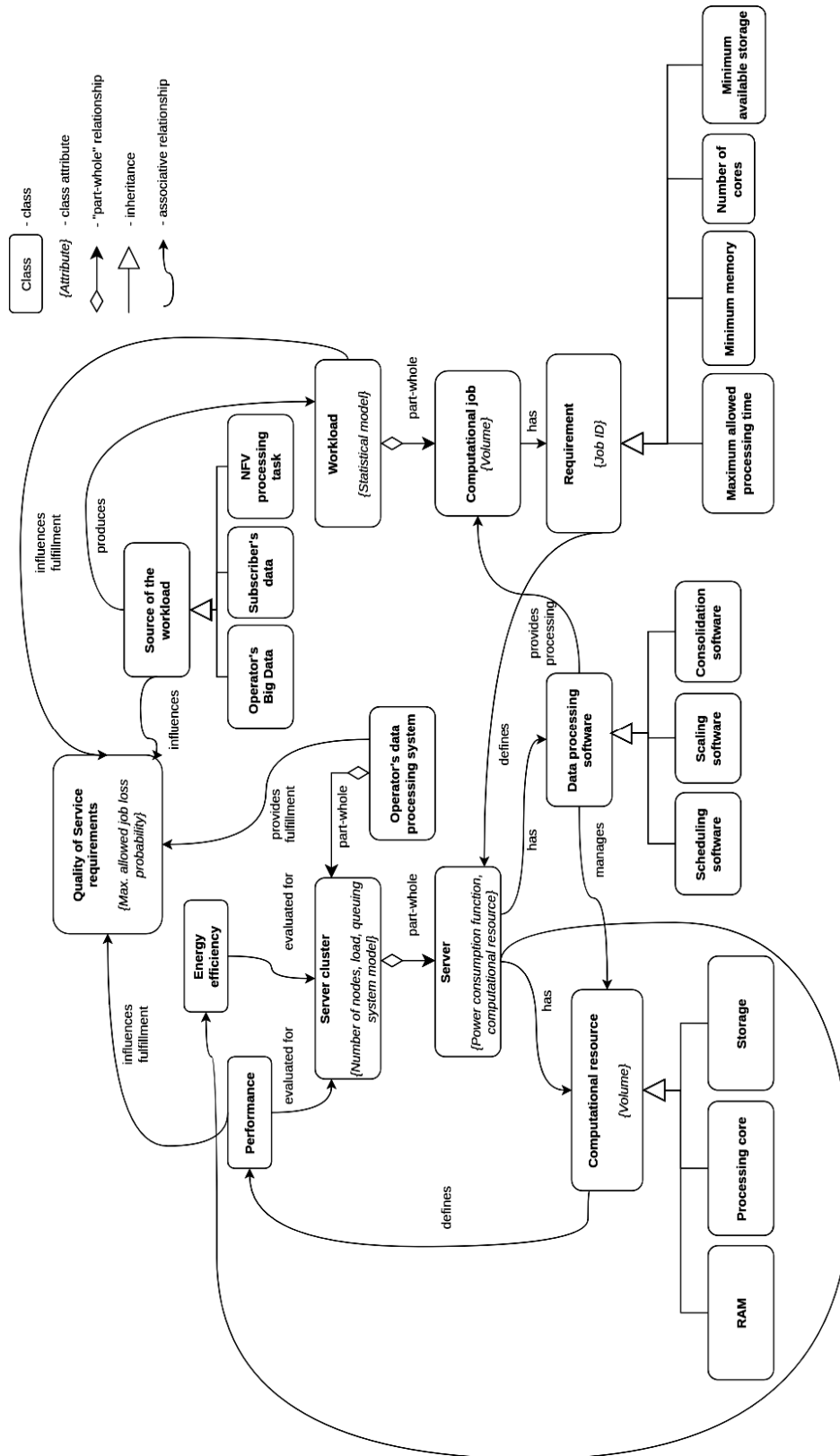


Fig. 2. The proposed ontology of the network operator's data processing system

These classes are connected with their parent class C_2 with the help of the relation of inheritance.

C_6 — Server cluster. A class describes physical objects that are part of a data processing system. The attributes of the server cluster are the number of its nodes, load and cluster model as a queuing system: $C_6 = (A_6, D_6)$.

C_7 — Server. This class describes the physical objects that are part of the server cluster. The attributes of the server are its power consumption function and computational resources: $C_7 = (A_7, D_7)$.

C_8 — Computational resource. This class includes concepts that describe the physical computing resources of the server. The attribute of the computing resource is its volume: $C_8 = (A_8, D_8)$. Inherited types of the computational resources are:

C_9 — RAM (Random Access Memory). This class reflects the concept of the physical resource of the server's RAM. The class has an inherited "volume" attribute that displays the amount of available server RAM: $C_9 = (A_9, D_9)$.

C_{10} — Processing core. The class describes the concept of the server's physical processor resource. It has an inherited "volume" attribute that displays the number of server processing cores: $C_{10} = (A_{10}, D_{10})$.

C_{11} — Data storage. This class displays the concept of a physical resource - a data storage device (hard disk, SSD, etc.). The class has an inherited attribute "volume", which reflects the available amount of server storage: $C_{11} = (A_{11}, D_{11})$.

These classes are connected with their parent class C_8 with the help of the relation of inheritance R_{n_8} .

C_{12} — Workload. This class is responsible for describing the abstract concept of input workload of a data processing system. The workload is generated by the sources of the workload (C_2). It is characterized by the attribute "statistical model", which corresponds to the concept of the statistical curve of the input load: $C_{12} = (A_{12}, D_{12})$.

C_{13} — Computational job. This class describes the concept of a computational job as a unit of input workload in the system. The class has a "volume" attribute that expresses the amount of computations required to process the job (for example, in the number of elementary operations): $C_{13} = (A_{13}, D_{13})$.

C_{14} — Requirement. This class describes the requirements of computational jobs for the physical resources of the system. The requirement has a job ID as an attribute to bind a specific requirement to the job that owns it: $C_{14} = (A_{14}, D_{14})$. The inherited requirements are:

C_{15} — Maximum allowed processing time. This is the time after which the job is removed from the processing, even if the processing was not completed successfully. The class inherits the "job ID" attribute to bind a specific job requirement to its job: $C_{15} = (A_{15}, D_{15})$.

C_{16} — Minimum memory (capacity). The class describes one of the requirements of a computational job, which reflects the minimum amount of RAM of a server, at which a job can still be processed on this server. The class inherits the "job ID" attribute to bind a specific requirement to the job that owns it: $C_{16} = (A_{16}, D_{16})$.

C_{17} — Number of cores. The class describes one of the computational job's requirements, which reflects the minimum number of free server processor cores that will be allocated for processing this job. The class inherits the "job ID" attribute to bind a specific job requirement to its job: $C_{17} = (A_{17}, D_{17})$.

C_{18} — Minimum available storage. The class describes one of the computational requirements that reflects the minimum amount of free server storage to be used during processing. The class inherits the "job ID" attribute to bind a specific job requirement to the corresponding job: $C_{18} = (A_{18}, D_{18})$.

These classes are connected with their parent class C_{14} with the help of the relation of inheritance $R_{m_{14}}$.

C_{19} — Quality of Service (QoS) requirements. The class describes the requirements to the data processing that correspond to a specific type of service (data being processed). In our interpretation the class has an attribute "maximum allowed job loss probability", which limits the probability of losing the job when processing by the system: $C_{19} = (A_{19}, D_{19})$. In general, any other attributes agreed within the QoS requirements may be shown here.

C_{20} — Energy efficiency. Describes the concept of energy efficiency of a processing system in general and each physical server in particular as the amount of electrical power consumed to perform a unit of work.

C_{21} — Performance. Describes the concept of performance of a computer system in general and each physical server in particular as the amount of work performed per unit time.

C_{22} — Data processing software. Displays the concept of software used in a distributed computing system to distribute and process the workload, as well as control the state of the system as a whole. In our interpretation it encompasses the following types of software:

C_{23} — Scheduling software. This class represents the software that is used to distribute the computational jobs between the available hardware (servers). This process is also known as jobs' scheduling.

C_{24} — Scaling software. This type of software is responsible for managing the quantity of available hardware in the system. It is in particular important for energy efficiency: fewer resources may be kept available in case of the underload of the system to save some power.

C_{25} — Consolidation software. This software manages the consolidation process of the virtual machines.

If other types of software are used by the operator, they may be added as the separate classes of the ontology as well.

These classes are connected with their parent class C_{22} with the help of the relation of inheritance $R_{n_{22}}$.

Associative relations: $R_{AS} = \{C_i X C_j\}$

"*produces*" — displays the relationship between the "Source of the workload" class and the "Workload" and shows the process of producing the workload by various sources;

"*influences*" — reproduces the relation between the "Source of the workload" and "QoS requirements" and expresses, that different types of workload sources have different QoS requirements;

“*influences fulfilment*” — connects the classes “Workload” and “QoS requirements” and shows that the amount of workload influences the fulfillment of the QoS requirements (e.g., it is more difficult to fulfill the requirements in busy hours);

“*provides fulfilment*” — this relationship describes the relationship between the “Operator’s data processing system” and “QoS requirements”. The system must operate in such a way as to ensure compliance with the QoS requirements;

“*defines*” — this relationship describes the relationship between the class “Computational resource” and “Performance”. The essence of this relationship is to reflect the impact of the quantity and quality of server computing resources on its performance. This relation connects the classes “Requirements” and “Server” as well to show the fact that the jobs’ requirements define the choice of the hardware, on which the job may be processed;

“*has*” — a connection that shows the logical affiliation of one class to another. The classes “Server” — “Computing resource”, “Server” — “Data processing software”, “Computational job” — “Requirement” have this relation;

“*provides processing*” — this relation connects the classes “Data processing software” and “Computational job” showing, that the data processing software operates with the jobs in order to let them be processed;

“*manages*” — the relation shows, that the “Computational resources” are managed (scaled, distributed, etc.) with the help of the “Data processing software”;

“*evaluated for*” — shows the relationship between the parameters of the data processing system (Energy Efficiency and Performance) and the Operator’s data processing system itself.

“Part-whole” relations: $R_{IA}(O) = C_k(O) \subset C_m(O)$

“Part-whole” relations are defined between the classes “Operator’s data processing system” and “Server cluster”, “Server cluster” and “Server”, “Workload” and “Computational job” to show that one entity is a part of another one.

The described ontology formalizes the data processing system and simplifies the analysis and management of such a system. The parameters of the system are explicitly exposed and the operator is able to see the relations between them.

AN EXAMPLE OF A SIMPLIFIED DATA PROCESSING ORGANIZATION WITH THE HELP OF THE ONTOLOGICAL MODEL

Let us consider an example of simplified data processing organization with the help of ontology. In the related research [21], we consider the problem of energy efficient data processing which is a very important topic nowadays. The problem is to ensure a minimal power consumption of a data processing system without losing the processing performance and ensuring the fulfilment of the QoS requirements. This is a complicated task which requires having an overview on a system as a whole, and taking into account multiple influencing parameters. Due to the task’s specificity, it is infeasible to analyze the separate parts of the system, since they cooperate solving the processing tasks and act together as a single distributed data processing system.

The proposed ontology explicitly shows the complex semantical dependencies between the input workload parameters of the system and assessment criteria. The operator sees that the fulfilment of the QoS requirements is directly influenced by the system performance and indirectly influenced by the computational

resources of the system. Thus, in order to fulfil the QoS requirements he should increase the volume of the computational resources. However, he also sees that the resources are managed by the data processing software and so, instead of changing the resources (which may be costly) the operator may focus on the software tuning in order to improve the resource management.

Considering the energy efficiency criterion, the operator sees that it is defined for each separate processing node (server) in the system. And the decision regarding the server to be chosen is taken based on the requirements of the input jobs. Therefore, a design decision based on this analysis would be to pay attention to the jobs' requirements analysis to ensure a more thoughtful choice of the processing server.

Based on a conducted analysis, a comprehensive energy efficient approach to workload processing was proposed in [21]. This approach takes into account individual power consumption characteristics of computing nodes, deals with dynamic workload deviations, and ensures meeting requirements to the service quality combining energy efficient scheduling and horizontal scaling. The results of the approach are largely due to the ontological model, which helped to identify and link together all the complex semantic dependencies of the system. All the details regarding the approach and its evaluation may be found in [21]. The main point that we would emphasize here is that due to the used ontological model, the complex dependencies between the parameters and the assessment criteria of the system were easily embraced and the formalized system representation was used as an input for the automatic system optimization software. The ontological model is designed and refined once for the whole system and helps to analyze it in the future due to the visualization and formalization.

CONCLUSION

In this paper we analyzed how the data processing is involved in the range of activities of the modern information and communication network operator. We defined that the main directions of data processing in modern information and communication networks are related to the NFV and SDN applications, traditional subscribers' management functions (authorization and authentication of users, billing service, etc.) and Big Data processing. In order to simplify the organization of this data processing, an ontological model of the data processing system in communications was proposed. This model formalizes the data processing system exposing its parameters and visualizing the relations between them. It simplifies the analysis of the system for the network operator and enables partial or full automatization of the system analysis and management in future. An example of an energy efficient data processing problem was considered to show how an ontological model simplifies the analysis and optimization of a complex data processing system.

The proposed ontology assumes the possibility of expansion and addition. For example, processing security can be considered as another important criterion for the quality of data processing (especially for the modern network services such as connected vehicles). This parameter and corresponding additions to the ontology should be considered as a related future work.

REFERENCES

1. *The law of Ukraine About telecommunications* [Online]. Available: <https://zakon.rada.gov.ua/laws/show/1280-15#Text>. Accessed on: 11 Oct. 2020.

2. Jan Harrington, *The Relational Data Model*, 2009. doi: 10.1016/B978-0-12-374730-3.00005-X.
3. Antonio Badia, *Relational, Object-Oriented and Object-Relational Data Models*, 2006. doi: 10.4018/9781591405603.ch088.
4. John Mariani and Tom Rodden, *A Model for Schema Evolution in Object-Oriented Database Systems*, 1994.
5. Nicola Guarino, Daniel Oberle, and Steffen Staab, *What Is an Ontology?*, 2009. doi: 10.1007/978-3-540-92673-3_0.
6. L.S. Globa, R.L. Novogrudska, and A.V. Koval, "Ontology Model of Telecom Operator Big Data", *Proceedings of IEEE International Black Sea Conference on Communication and Networking (BlackSeaCom)*, 1-5, 2018. doi:10.1109/BlackSeaCom.2018.8433710
7. C. Villalonga et al., "Mobile Ontology: Towards a Standardized Semantic Model for the Mobile Domain", *Proceedings of the 1st International Workshop on Telecom Service Oriented Architectures (TSOA 2007) at the 5th International Conference on Service-Oriented Computing, Vienna, Austria, September 17, 2007*, pp. 248–257.
8. Qiao Xiuquan, Li Xiaofeng, and Chen Junliang, *Telecommunications Service Domain Ontology: Semantic Interoperation Foundation of Intelligent Integrated Services*, 2012. doi: 10.5772/36794.
9. *View on 5G Architecture: 5G PPP Architecture Working Group* [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2019/07/5G-PPP-5G-Architecture-White-Paper_v3.0_PublicConsultation.pdf. Accessed on: 20 Sept. 2020.
10. Data Center Market and Technology Trends Power Electronics presentation [Online]. Available: https://www.slideshare.net/Yole_Developpement/data-center-market-and-technology-trends-power-electronics-presentation-held-at-apec-2016-from-yole-dveloppement. Accessed on: 18 Oct. 2020.
11. Qiu Yeliang, Jiang Congfeng, Wang Yumei, Ou Dongyang, Li Youhuizi, and Wan Jian, *Energy Aware Virtual Machine Scheduling in Data Centers*. *Energies*, 2019. doi: 12. 646. 10.3390/en12040646.
12. F. Armenta-Cano et al., "Min_c: Heterogeneous Concentration Policy for Power Aware Scheduling. Trudy ISP RAN", *Proc. ISP RAS*, vol. 27, issue 6, 2015, pp. 355–380.
13. Hosseinimotlagh Seyedmehdi, Khunjush Farshad, and Hosseinimotlagh Seyedmahyar, *A Cooperative Two-Tier Energy-Aware Scheduling for Real-Time Tasks in Computing Clouds*, 2014, pp. 178–182. doi: 10.1109/PDP.2014.91.
14. Mohammad Aldossary and Karim Djemame, *Performance and Energy-Based Cost Prediction of Virtual Machines Auto-Scaling in Clouds*, 2018.
15. P.P. Vorobienko, L.A. Nikityuk, and P.I. Reznichenko, *Communication and information networks, Approved by the Ministry of Education and Science of Ukraine as a textbook for students of higher education institutions*. Kyiv, 2010.
16. Talapko Domagoj, *Telecom datacenter power infrastructure availability comparison of DC and AC UPS*, 2012, pp. 1–5. doi: 10.1109/INTLEC.2012.6374509.
17. Benzekki Kamal, El Fergougui Abdeslam, El Belrhiti El Alaoui Abdelbaki, *Software-defined networking (SDN): A survey*. *Security and Communication Networks*, 2017. doi: 9. 10.1002/sec.1737.
18. Zhang Tianzhu, *NFV Platform Design: A Survey*, 2020.
19. *Big Data for business. Big Data from Kyivstar* [Online]. Available: <https://kyivstar.ua/uk/business/products/big-data>. Accessed on: 13 Oct. 2020.
20. "Cellular operators know everything about their subscribers", *Vedomosti*. Accessed on: 13 Oct. 2020. [Online]. Available: <https://www.vedomosti.ru/technology/articles/2015/05/26/593579-sotovie-operatori-znayut-vse-o-svoih-abonentah>
21. L. Globa and N. Gvozdetska, "Comprehensive Energy Efficient Approach to Workload Processing in Distributed Computing Environment", *2020 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, Odessa, Ukraine, 2020, pp. 1–6. doi: 10.1109/BlackSeaCom48709.2020.9235010.

Received: 07.12.2020

INFORMATION ON THE ARTICLE

Larysa S. Globa, ORCID: 0000-0003-3231-3012, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: lgloba@its.kpi.ua

Nataliia A. Gvozdetska, ORCID: 0000-0001-6549-0459, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: n.gvozdetska@gmail.com

Rina L. Novogradska, ORCID: 0000-0002-0533-5817, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: rinan@ukr.net

ОНТОЛОГІЧНА МОДЕЛЬ ДЛЯ ОРГАНІЗАЦІЇ ПРОЦЕСУ ОБРОБЛЕННЯ ДАНИХ В ІНФОРМАЦІЙНИХ ТА КОМУНІКАЦІЙНИХ МЕРЕЖАХ /

Л.С. Глоба, Н.А. Гвоздецька, Р.Л. Новогрудська

Анотація. Функціонування сучасних інформаційно-телекомунікаційних мереж неможливе без оброблення даних. З появою нових мережевих послуг кількість інформації, що потребує оброблення, зростає, при цьому ставляться дедалі жорсткіші вимоги до якості оброблення даних. Тому для оператора мережі дедалі більшої актуальності набуває проблема побудови та підтримання системи оброблення даних з можливістю гнучкого керування якістю послуг та масштабування. Такі системи оброблення даних мають комплексну внутрішню структуру з багатьма взаємопов'язаними параметрами, що ускладнює їх аналіз, керування та розширення. Запропоновано використовувати онтологічну модель для зберігання, подання та маніпулювання інформацією в системі оброблення даних оператора. Онтологічна модель дозволяє структурувати та систематизувати дані системи оброблення інформації і прозора відображати взаємозв'язки між параметрами системи для спрощення її аналізу та масштабування. Запропонована онтологія системи оброблення даних складається з трьох зв'язаних підсистем. Наведено опис запропонованої онтологічної моделі та додатково проаналізовано джерела інформації, яка потребує оброблення, в інформаційно-телекомунікаційній мережі.

Ключові слова: інформаційно-телекомунікаційна мережа, система оброблення даних, онтологія, модель, оператор мережі, аналіз, масштабування, клас, відношення.

ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ ДЛЯ ОРГАНИЗАЦИИ ПРОЦЕССА ОБРАБОТКИ ДАННЫХ В ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ СЕТЯХ / Л.С. Глоба, Н.А. Гвоздецкая, Р.Л. Новогрудская

Аннотация. Функционирование современных информационно-телекоммуникационных сетей невозможно без обработки данных. С появлением новых сетевых услуг количество информации, которая нуждается в обработке, возрастает, при этом выдвигаются все более жесткие требования к качеству обработки данных. Поэтому для оператора сети всё большую актуальность приобретает проблема построения и поддержки системы обработки данных с возможностью гибкого управления качеством услуг и масштабирования. Такие системы обработки данных имеют комплексную внутреннюю структуру со многими взаимосвязанными параметрами, что затрудняет их анализ, управление и расширение. Предложено использовать онтологическую модель для хранения, представления и манипулирования информацией в системе обработки данных оператора. Онтологическая модель позволяет структурировать и систематизировать данные системы обработки информации и прозрачно отражать взаимосвязи между параметрами системы для упрощения её анализа и масштабирования. Предложенная онтология системы обработки данных состоит из трех связанных подсистем. Приведено описание предложенной онтологической модели и дополнительно проанализированы источники информации, которая нуждается в обработке, в информационно-телекоммуникационной сети.

Ключевые слова: информационно-телекоммуникационная сеть, система обработки данных, онтология, модель, оператор сети, анализ, масштабирование, класс, отношения.

DATA SCIENCE — DEFINITION AND STRUCTURAL REPRESENTATION

P.P. MASLIANKO, Y.P. SIELSKYI

Abstract. This article is a continuation of the discussion on the existing meanings and formalization of the definition of “Data Science” as an autonomous discipline, field of knowledge, clarification of its defining components, integration, and interaction processes between them. It is noted that most scientific results trace the data-centric nature of the presentation and analysis of this discipline, i.e. the emphasis on the word Data. Analysis of the frequency of use of key terms in the definitions of Data Science shows what our colleagues focus on, which terms of the definitions of Data Science they are based on. In this paper, we make and argue certain additions to Drew Conway’s Data Science Venn Diagram, which does not reflect all the resources of the components that define the applied side of Data Science, and, moreover, does not reveal the interaction of these resources not from the point of view of the data researcher, nor in its global understanding. We also propose a unified structural representation of Data Science in the format of an updated Drew Conway’s Venn diagram based on a property/attribute that establishes correspondences that provide integration/interoperability between the elements of the sets of Drew Conway’s Venn diagram. The new definition of Data Science as an interdisciplinary science and methodology of presenting activities for analysis and extraction of data, information, and knowledge is substantiated.

Keywords: Data Science, Drew Conway’s Data Science Venn Diagram, Data Science definition, Data Science structure, data, information, knowledge.

INTRODUCTION

Starting from the 21st century, the phrase Data Science has begun to attract considerable attention from the world's academic and professional communities. Why the phrase? Despite dozens of savants trying to interpret its meaning in their own way, throughout numerous discussions about its components, this expression has not acquired the meaning of a clearly defined scientific term.

This article aims to carry out research and continue the discussion on the existing definitions and proper formalization of “Data Science” as an autonomous discipline, field of knowledge, clarification of its defining components, their characteristics of integration and interaction processes. Thus, Data Science is an object of analysis, which will be performed through in-depth study and synthesis of existing authoritative scientific results, articles and journals, blogs of well-known authors, and trusted publishers.

We systematized the information from all studied sources in the table for further analysis by the following criteria (columns of the table):

0. Definition of Data Science.
1. Keywords of the definition.
2. Semantics of a definition — list of tools on which it is based (methods, models, algorithms, processes, disciplines, etc.), as well as their interaction.

3. Features of the definition — its purpose (theoretical, practical, specialized, etc.), scope.
4. Discussion arguments and the uncertainties regarding the definition and understanding of Data Science, given in the source.
5. In total, 11 most common sources were analyzed and cited, which Data Science related key points are briefly described below.

RELATED WORK

The vast majority of scientific works on Data Science begin either with the famous expression “*Data Scientist: The Sexiest Job of the 21st Century*” of Thomas Davenport and D.J. Patil [1], or with a reference to Drew Conway’s *The Data Science Venn Diagram* [2] (Fig. 1), to which we shall return. In some cases, you can even find links to both resources at once.

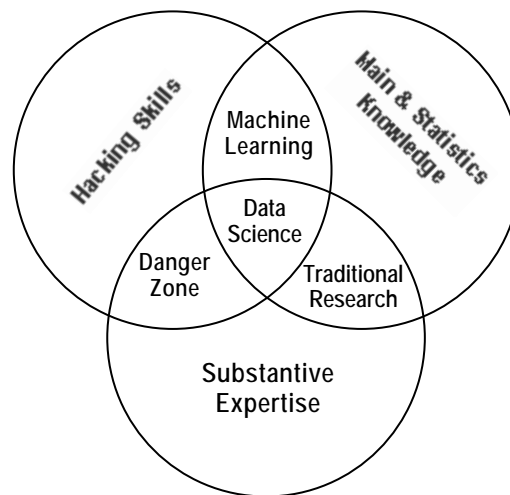


Fig. 1. The Data Science Venn Diagram [2]

It is worth noting that of all the works, dedicated to Data Science, this diagram is perhaps the only attempt at an in-depth presentation of the structure and content of the Data Science model, which nevertheless leaves a dry residue of uncertainty around some areas of the figure. What is a danger zone? What meaning does the author put into traditional research? Why did he choose machine learning as the intersection of mathematics and statistics with hacking skills? And, finally, can the last term be perceived as a scientifically justified, reasonable, and meaningful concept with an unambiguous interpretation?

Indeed, how can a job that requires hacking skills not be considered as sexy? Engineers and scientists in any definition want to see and understand a certain reasoned meaning, with a solid scientific basis, especially if the very described notion contains the word “science” itself. Otherwise, such loud statements will only provoke excitement and fruitless controversy over the newly introduced term, which, in fact, happened.

Due to the ambiguous emergence of Data Science, the debates over the interpretation of this name immediately began among the academic and professional communities. In particular, the question arises about the similarity between

Data Science and classical, well-known statistics. For example, Cathy O’Neil and Rachel Shutt in their book *Doing Data Science* [3] (where, by the way, the entire 15 pages are devoted to the first part “Introduction: What is Data Science?”), refer to numerous protests by experts in the field of statistics against the uniqueness of Data Science, calling it a new-fashioned rebranding of their alma mater. The authors themselves claim their differences, emphasizing the specific processes created by the pioneers of Data Science, allowing to work with more data — the *processes of Data Science* [3]. In general, this resource covers more the professional aspect of Data Science, explaining it from the point of view of data scientists as specialists in this field, and the skills that such positions require.

Vasant Dhar, whose testimony can be found in *Communications of the ACM’s Data Science and Prediction* article [4], also joined the defense in the case of Data Science vs. Statistics. The author focused on a whole list of differences [4]:

1. First of all, data — the main fuel of Data Science — is quickly becoming unstructured, diverse. Therefore, the analysis of “raw” data, as well as combining data of different types (*feature engineering*), demands additional interpretation and understanding, based on the foundations of multiple other disciplines (linguistics, sociology, etc.) [4].

2. Nowadays, most of the data is produced by computers to be consumed by other computers [4]. In these realities, it is computers that make decisions that encourage their operators — data scientists — to retrain: to play as well the role of risk managers, to act as a guarantor-supervisor of developed system quality instead of the more classic duty of an expert in the context of statistics.

3. Machine learning, applied for creating unfailing predictive models, is an essential Data Science component, which is more and more concerned with forecasting various values, events, phenomena [4].

“Data Science, ... , is perhaps the best label we have for the cross-disciplinary set of skills that are becoming increasingly important in many applications across industry and academia.” — this definition is given by Jake Vanderplas in the Python Data Science Handbook [5] (also with reference to Fig. 1), where he often uses the concept of “skills”, which, again, emphasizes a more professional application.

“Multifaceted discipline” — say the authors of the book *Data Science for the Layman: No Math Added* [6] Annalyn Ng and Kenneth Soo, focusing on machine learning as a key component and citing a standard algorithm of carrying out research in the field of Data Science [6]:

1. Data processing and preparation for analysis.
2. Selection of potentially effective machine learning algorithms.
3. Optimization of (hyper-) parameters of algorithms: training, validation.
4. Construction of integral models (combination of certain algorithms or their separate usage) with their further comparison and selection of the best.

In addition to applied specifics, there are definitions of a high level of abstraction, more clear and intuitively perceived by the human mind. Well-known experts in the field of Data Science, Foster Provost and Tom Fawcett formulate the key activities of data scientists: extracting useful information and knowledge from data [7]. Hence, Data Science is also compared to Data Mining: “At a high level, Data Science is a set of fundamental principles that guide the extraction of

knowledge from data. Data mining is the extraction of knowledge from data, via technologies that incorporate these principles” [7].

In parallel, the authors focus on the analysis of the structure of Data Science in the context of effective solutions to real business problems [8]. Here, one of the principal processes — data-driven decision making and its progressive automation.

Often, and certainly not without reason, Data Science is closely linked to data analysis. For example, Matthew Waller and Stanley Fawcett describe Data Science quite abstractly: “Generally, Data Science is the application of quantitative and qualitative methods to solve relevant problems and predict outcomes.” [9], but instead derive their own model of influencing data scientist’s performance by two interdependent components: domain knowledge and analytical skills.

A Ukrainian specialist, Bohdan Pavlyshenko, agrees with Waller and Fawcett, focusing on data analysis, the need for a proper understanding of the nature of data, and the specifics of a particular domain in problem-solving [10].

We are currently coming to a certain consensus on the applied essence of Data Science as a business tool, a profession. Most of the above resources trace the data-centric nature of the representation and analysis of the discipline, i.e., the emphasis on the word Data. And what about Science? What about the academic side of the coin?

Jeff Leek answers these questions, listing a number of arguments in defense of science and the complexity of solving scientific problems [11], citing, in particular, a quote of John Tukey, a pioneer in data analysis: “The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.” [12]. The author also accentuates the main reason for the outbreak of excitement around Data Science — the focus on data, proclaiming: “The long term impact of data science will be measured by the scientific questions we can answer with the data.” [11].

Based on the aforementioned arguments, Fig. 2 shows the results of frequency analysis of the most commonly used key terms present in the various definitions of Data Science. This analysis is an example of one of the operations of semantic decomposition, carried out on the basis of the constructed table, the criteria of which are described in the introduction.

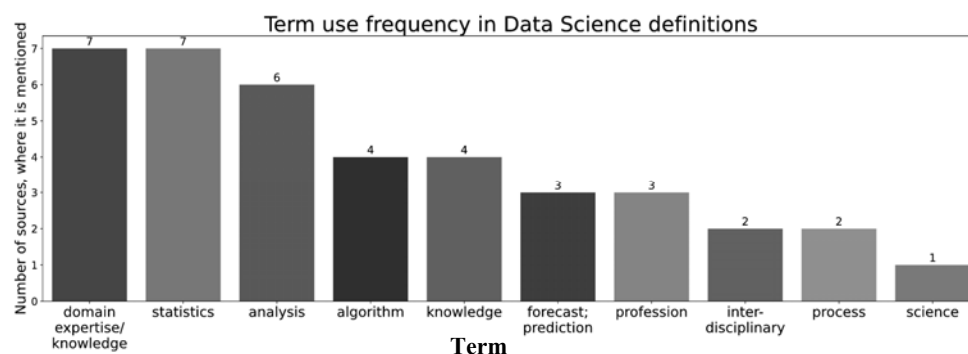


Fig. 2. Histogram of frequency analysis of terms used in the considered definitions of Data Science

These indicators help to better understand what our colleagues are focusing on, what terms the definitions of Data Science are based on: for example, the

most commonly used terms are “*domain expertise/knowledge*” and “*statistics*”, which do not fully reflect the components of the object of our research. Instead, the word “*science*” is mentioned only once, confirming the mostly data-centric nature of existing definitions.

Such a simple but quite clear way of comparative analysis of the term use frequency approximately reflects the overall vision of examined authors on the structure and content of Data Science.

Thus, on the basis of the results of even these brief studies of the publications of authoritative experts, an ambiguous, incomplete picture of the defining elements of Data Science is formed. Moreover, the obvious problem of a lack of compromise and a clear link between Science and Data is highlighted.

SYNTHESIS

Based on the preliminary conclusions, on the above scientific results, we will make some clarifications of the interaction of entities and formalization of Data Science.

Studies of the Data Science representation, analysis of the results of the selection and justification of its attributes, provide grounds for making adjustments to the definition and structure of Data Science.

These rectifications imply some additions to the repeatedly mentioned Drew Conway’s Data Science Venn diagram (Fig. 1) [2], which does not reflect all the resources of the components that form the applied side of Data Science, and, moreover, does not reveal the interaction of these resources from the point of view of the data scientist, nor in its global sense.

In this article, we propose an updated, refined version of Drew Conway’s Venn Data Science diagram and try to explain and justify not only the essence of its components but also the principles of their integration and interoperability (Fig. 3).

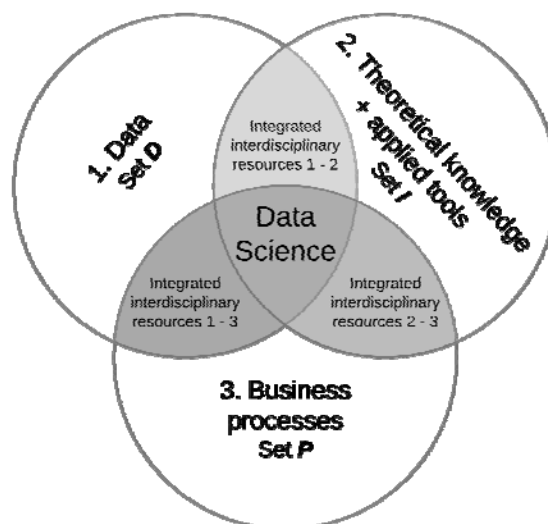


Fig. 3. Structural representation of Data Science in the format of an updated Venn Data Science diagram

FORMAL REPRESENTATION OF ELEMENTS OF DATA SCIENCE

RESOURCE SETS

The set of formalized data resources — D (data), d is an element of set D , $d \in D$.

The set of theoretical and applied data processing tools — I (instrument), i is an element of set I , $i \in I$.

The set of business processes for data, information, and knowledge acquisition by means of theoretical and applied data processing tools — P (process), p is an element of set P , $p \in P$.

Sets of integrated interdisciplinary resources 1–2, 1–3, and 2–3.

Let's explain each entity separately:

- data — raw materials, a key research resource that determines the features and models of the particular domains;
- theoretical knowledge and applied tools — instruments, through which the process of extracting information and knowledge from data takes place. These include both exact sciences (mathematics, statistics, computer science, machine learning, data analysis, etc.) and software applications (programming languages, their libraries, frameworks, development environments, visualization tools, etc.);
- business processes: here we are talking about the organization of research at the meta-level, setting goals and objectives, determining the main stages of work, their sequence, the nature and features of the evaluation of results, etc. Essentially, all the above actions will vary from one problem to another. Why “business”? To emphasize the need to optimize this component in order to maximize the benefits of the carried out researches/developed systems;
- integrated interdisciplinary resources 1–2, 1–3, and 2–3: the essence of the corresponding intersections is not fully compatible with the classical definition of the intersection operation in Set Theory, given the obvious fact that the nature of the elements of different sets is different. Therefore, we are talking about the existence of functional relationships between different types of resources: for example, theoretically, zone 1–2 includes existing data, known to mankind, that can be processed using existing theoretical knowledge and applied tools, but such interaction is not a subject to any existing business process.

DECOMPOSITION OF ELEMENTS OF DATA SCIENCE RESOURCE SETS

1. Data (set D). Plays the role of a kind of fuel for Data Science instruments and processes. Obviously, there is a lot of data in the world and every second more is created. So let's try to bring order to the ocean of this natural chaos, defining the main elementary component of the entirety of data — a formalized data resource d . Thus, zone 1, shown in Fig. 3, consists of formalized data resources — information, collected, stored in a certain form, which can be classified, for example, by the following criteria:

- by type of storage: distinguish digital types of data storage (hard drives, SSDs, USB-drives, etc.) and in contrast to them — more classic — rock, wall inscriptions, carvings, books, magazines, newspapers, etc. We do not forget the

immaterial data — unrecorded real-time speech, thoughts, movement of any objects;

- by structure, data can be divided into structured (mainly numbers and numerical arrays, standard types of programming languages) and unstructured: audio, images, video, text;
- by availability: confidential and public data.
- In any case, in the context of area 1 we are talking about the set D of formalized data resources d :

$$\exists d, d \in D. D = \{d\}.$$

Some examples of data as formalized resources are listed below:

- numeric data arrays: the simplest (at least for a computer) formalized representation of information. Arrays can be of different shapes, sizes and contain any number of elements. Note also that the numbers (scalars) themselves can be seen as formalized data resources;
- digital images: in their structure — the same, in some sort of way, ordered numerical arrays (pixel values), but at the level of human perception of information (visualization), play a special, more significant role, so they can also be considered as formalized resources;
- audio files are another type of information that humans perceive by ear. Physically, an audio file is a specific set of frequencies — numbers that follow a strict order. Therefore, it is about data in the form of sequences, series, which are reproduced in time;
- video files — a more complex case, which includes not only a set of images but also audio. That is, we define the presence of two different types of sequences, as well as the mechanism of their synchronization as integral elements of correct video playback;

As a part of Data Science, all of the above and many other types of data are summarized in one, more extensive formalized data resource — a so-called *dataset*.

In *Data Science for Business*, such terms as database *table*, *worksheet* (for example, a sheet of an .xsl file), and *dataset* are equated to each other; a more specific decomposition of the latter is presented [7]:

- data sets consist of so-called *examples* (samples) or *instances* (table rows) [7];
- each instance, in turn, is comprised of a fixed (in the classical representation) number of *features* (*columns* of the table), the values of which uniquely identify instances [7].

2. Theoretical knowledge and applied tools (set I). Any instruments, models, algorithms, human skills, formalized or materially implemented, aimed at carrying out certain operations on data for their better understanding. For the purpose of formalization we will define an elementary component of this set as an instrument i :

$$\exists i, i \in I. I = \{i\}.$$

Such elements can acquire different levels of abstraction and different scales: for instance, individual clusters of knowledge can, in turn, be combined

into whole areas of knowledge. Here are the most interesting theoretical aspects for Data Science:

- Statistics: let's start with it to pay tribute to fellow statisticians. Undoubtedly, this is a vast science that includes many *i*'s, but only a certain part of them are used in Data Science, in particular, elements of descriptive statistics at the stages of exploratory data analysis (*EDA*) [13]: mode, median, mean, standard deviation. *EDA* is also a part of *Data Analysis*.

- Data analysis: in order not to invent anything superfluous, we go back to the definition of the most reliable source — John Tukey [14], who provides a comprehensive list of components of the discipline: data analysis procedures, methods of interpreting their results, simplification and improvement of data analysis on the earliest stages of data collection, as well as all the techniques of statistics, that are applicable to data analysis [14]. That is, Data Analysis is closely related to the Statistics domain.

- Artificial Intelligence (AI): since the inception of this term, the constant debate around its essence has never subsided. AI should be considered as a separate section of computer science, designed to program machines for human behavior, thinking, and independent decision-making [15]. In the case of Data Science applications, decision-making is often based on predicting certain results.

- Machine learning: this term originates from Arthur Samuel's article "*Some studies in machine learning using the game of checkers*" [16], where the author uses this phrase literally — the process of learning machines — programming computers for behavior as such that includes the learning process if it were inherent in humans or animals [16]. More specifically, it is about automated optimization of computer performance, based on experience.

- Deep learning [17]: the problems of applying machine learning techniques on unstructured data: texts, music, images, etc. are becoming more and more popular. Informative (for computers) representation of such raw data requires their automatic interpretation through step-by-step processing of numerical input arrays throughout many stages of data projection onto spaces of the higher levels of abstraction. Such a procedure is a key aspect of Deep Learning [17], i.e. learning the layers (stages) of neural networks on the data via the generalized learning process instead of explicitly developing the necessary projections by hand.

- Big Data Analytics: this term should be taken literally — the field of knowledge about the application of advanced analytical methods on the big amounts of data, according to Philip Russom [18]. The presence of big data is manifested not only by their volumes but also by such characteristics as data variety and velocity (*3 Vs*).

- Data Mining: recall the definition of Provost and Fawcett that Data Mining — the extraction of knowledge from data using technologies that embody the principles of Data Science [7]. This example allows us to trace the direct connection of Data Mining with Data Science as an integral part of it.

- Data visualization — techniques for presenting data of different nature and dimensions in the most understandable and human-readable form — graphic [19]. In this set of tools, in addition to countless frameworks and software that implement the full range of possible charts and graphs (in Cartesian, polar coordi-

nates, scatter, line plots, histograms, bar and pie charts, 3D images, etc.), more complex machine learning methods of dimensionality reduction can be highlighted here as well. A good example is the Principal Component Analysis (PCA).

Let's also notice the application tools — instruments that allow implementing algorithms and methods of the above theoretical knowledge in the form of (open-source or private) software applications, platforms, frameworks, libraries, systems, and so on. These include:

- programming languages widely used in the domain of Data Science: here we can consider both: such programming languages as *R* [20], *Python* [5], which are used directly for the development of Data Science systems, for the implementation of the higher-level interfaces and components of such systems; as well as *C*-family programming languages, used for the development of lower-level APIs (Application Programming Interfaces) in order to optimize and parallelize basic computations. As an example of such a hierarchy — *TensorFlow* [21] machine learning system from *Google Brain*;

- whole systems of computer mathematics and algebra (*MATLAB* [22], *MathCad* [23]), statistics (*STATISTICA* [24]); machine and deep learning systems, big data systems and environments, which are distinctive by the presence of interfaces for different programming languages (*TensorFlow* [21], *Torch* [25], *Spark* [26]) or even by embedded graphical user interfaces (eg *Orange* and *KNIME* [27]);

- separate add-ons of the aforementioned systems of the highest level of abstraction (*Keras* [28] for *TensorFlow*); specialized programming language libraries, modules, packages that provide ready-made software solutions for machine learning (*Scikit-Learn* [29]), data processing (e.g. *NLTK* [30] for text data), their visualization and interactive calculations (*matplotlib* [31, 32], *pygal*, *Plotly*, *Pyvot* [31], *pandas*, *seaborn* [32], etc.), and many others.

This list is not exhaustive and can be extended with many other theoretical and applied instruments.

3. Business processes (set P). The set of business processes for data, information, and knowledge acquisition by means of theoretical and applied data processing tools.

In order to formally represent the interaction of the two previous sets D and I , a third set P is introduced. To optimally extract knowledge and new information from the formalized data resource d using the instrument i , we subordinate the whole entirety of work that needs to be done to a certain process p :

$$\exists p, p \in P. P = \{p\}.$$

The relationship between d , i , and p will be demonstrated in more detail further, in the context of the sets of integrated interdisciplinary resources. So far, a basic example of the Data Science process is shown in Fig. 4, suggested by Cathy O'Neil and Rachel Shutt [3].

In this representation, 8 main stages of the process are identified:

1. Collecting raw data from any real-world resources.
2. Data Processing.
3. Their cleaning.
4. Exploratory Data Analysis (EDA).

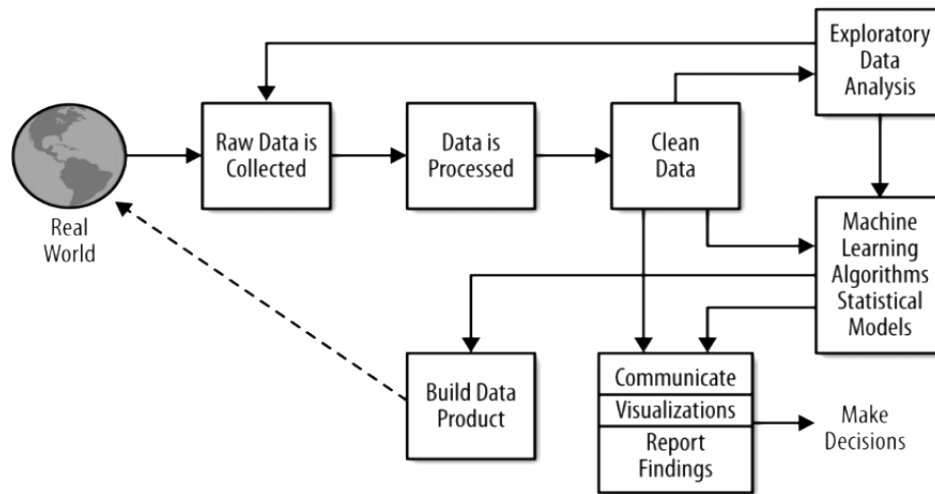


Fig. 4. Data Science Process [3]

5. Construction of statistical models and training of Machine Learning algorithms using collected, processed, cleaned data.

6. Stage of internal communication. Presentation of results to members of the development team of a specific Data Science system, as well as its stakeholders.

7. Creation/production of new data used in the real world.

8. Decision-making based on the obtained results.

It is worth noting that there are direct links between certain stages: for example, EDA may reveal a lack of data that needs to be collected, or cleaned data can be visualized to explain its nature to colleagues.

We compare the above-described Data Science process with the *Cross Industry Standard Process for Data Mining* (CRISP-DM), analyzed by Foster Provost and Tom Fawcett [7] and presented in Fig. 5.

Let us pay attention to its circular iterative-incremental nature, as well as the presence of two rather abstract, generalizing stages — *Business Understanding* and *Data Understanding* [7]:

- the first embodies the need for a clear problem statement in accordance with the given task, the search for creative methods to achieve the goal, its optimal formalization, which would allow the application of already existing methods as effectively as possible;

- while the second stage focuses on the strategic analysis of the main raw material Data Mining — data. Here it is essential to understand the basic structure, pros, and cons of the involved data. The proper assessment of the potential sources of additional information, the necessary investment (both time and financial) in their research and use, is also important.

Whereafter is an integral *Data Preparation* procedure, which, by analogy with the process in Fig. 4, combines data processing, cleaning [7], and EDA, which in themselves can be a multi-iterative subprocess.

The next stage of CRISP-DM — *Modeling* — also has a direct correspondence in the presentation of the Data Science process, where, again, more specific names are given.

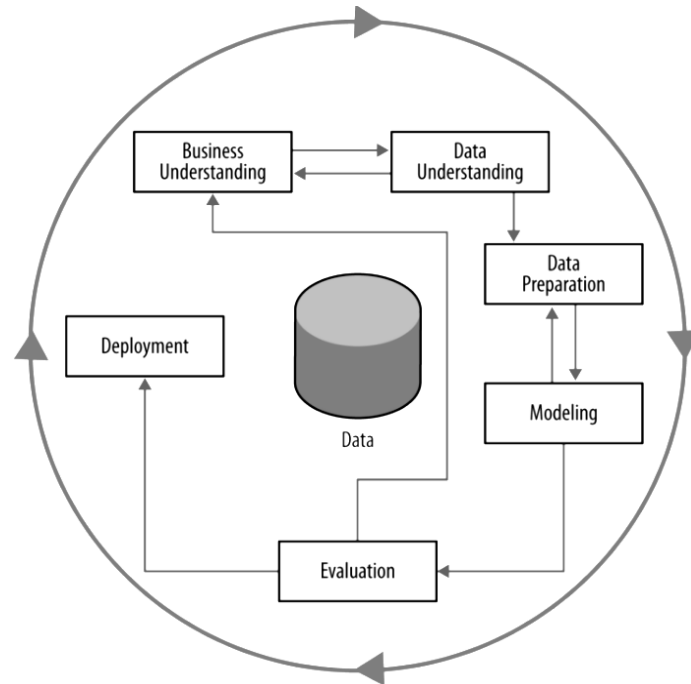


Fig. 5. CRISP-DM [7]

Any development should be subject to quality control through regular verification and validation of the built models and the entire system. The authors also implicitly emphasize the need for communication, presentation of results, as well as the concepts of their simple and clear explanation to key stakeholders (investors) [7], who are responsible for making major business decisions.

After all, in the context of business, the decisive factor is the successful *Deployment* of each system approved by management [7]. In essence, the solution to business problems is directly related to obtaining a certain material benefit.

4. Integrated interdisciplinary resources 1–2, 1–3, and 2–3.

Here and later in this paper, integrated interdisciplinary resources 1–2, 1–3, and 2–3 are subsets formed on the basis of the presence of integration/interoperability properties between elements of sets D , I , and P . Integration/interoperability of elements of sets D , I , and P is the ability to process a certain data resource by means of a certain subset of instruments, following certain processes.

Such a decomposition and generalized systematic representation of the elements of resource sets of Data Science shows and justifies both the complexity and the need for comprehensive research and ongoing discussion on existing definitions and formalizations of Data Science as an autonomous discipline and field of knowledge, clarification of its defining components and characteristics, their integration and interaction processes.

DATA SCIENCE AS A SET OF INTERDISCIPLINARY RESOURCES OF SETS D , I , AND P

To generalize the formal representation of interdisciplinary sets, we define, for example, an arbitrary entity A as a finite resource set A , where $a \in A$ — ele-

ments of the set A , and an arbitrary entity B as a finite resource set B , where $b \in B$ — elements of the set B .

Let us establish the rules of correspondence [33] C_{ab} and C_{ba} between the elements of the resource set A and the elements of the resource set B .

That is, if $\exists a \in A, \exists b \in B, (a, b) \in C_{ab}$, then we say that the element b of the set B corresponds to the element a of the set A , given correspondence C_{ab} .

And if $\exists b \in B, \exists a \in A, (b, a) \in C_{ba}$, then we say that the element a of the set A corresponds to the element b of the set B , given correspondence C_{ba} .

Hereinafter, the term “ C_{xy} correspondence” must be understood as the fundamental concept of set theory, which establishes, explains, and formalizes the relationships between the elements of the sets $X = \{x\}$ and $Y = \{y\}$ [33].

Next, we consider the presence of “correspondence” as a common property/feature of pairs of elements of sets A and B .

Let’s now construct some finite set of interdisciplinary elements M from all pairs of elements of the resource set A and elements of the resource set B , which have common properties/features and established on the basis of these features correspondences C_{ab} and C_{ba} .

Definition 1. Set $M, \forall m \in M, m = \{a, b \mid a \in A, b \in B, (a, b) \in C_{ab}\} \mid m = \{b, a \mid b \in B, a \in A, (b, a) \in C_{ba}\}$, of interdisciplinary pairs of of arbitrary finite sets A and B — a set of pairs of elements formed by elements elements of set A , and elements of set B , having a common property/feature that establishes the correspondences C_{ab} and C_{ba} between these elements.

We substantiate Definition 1 as a simple way to form a set M by combining pairs (a, b) and $(b, a) \forall a \in A, \forall b \in B$, selected by a common property/feature that establishes the correspondence C_{ab} and C_{ba} between these elements. To do so, we define and apply a common property/feature that establishes the correspondences C_{ab} and C_{ba} between the elements of sets A and B . Since the types of properties/features that determine the correspondence C_{xy} can be set quite a lot, in this case, we introduce restrictions and specify the correspondence C_{xy} between elements of sets A and B .

In particular, we define such a necessary property/feature for us that provides integration/interoperability between the elements $a \in A, b \in B$.

According to this property/feature, a certain subset of pairs of elements of sets A and B can be distinguished, which has established correspondences C_{ab} and C_{ba} .

Let the set of pairs of elements $\{a, b \mid \forall a \in A, \forall b \in B, (a, b) \in C_{ab}\}$ and the set of pairs of elements $\{b, a \mid \forall b \in B, \forall a \in A, (b, a) \in C_{ba}\}$ be defined as such, that have the property/feature, which establishes correspondences C_{ab} and C_{ba} and provides integration/interoperability between elements a and b .

Then the set M is defined as the union of pairs of elements of the set $\{a, b \mid \forall a \in A, \forall b \in B, (a, b) \in C_{ab}\}$ and the set $\{b, a \mid \forall b \in B, \forall a \in A, (b, a) \in C_{ba}\}$,

selected as having a common property/feature that determines the correspondences C_{ab} and C_{ba} . More formally:

$$M = \{a, b \mid \forall a \in A, \forall b \in B, (a, b) \in C_{ab}\} \cup \\ \cup \{b, a \mid \forall b \in B, \forall a \in A, (b, a) \in C_{ba}\}.$$

Similarly, it is possible to form a set of interdisciplinary pairs of elements of any number of sequentially combined arbitrary finite sets A, B, D, \dots on the basis of a common property/feature defined for them, which provides integration/interoperability between pairs of elements $a \in A, b \in B, d \in D, \dots$ and establishes the correspondences of C_{xy} between the elements of adjacent sets.

Definition 2. The set of interdisciplinary pairs of elements of an arbitrary number of finite sets A, B, D, \dots, X, L is the set of elements of successive pairs $(a, b), (b, d), \dots, (x, l), \forall a \in A, \forall b \in B, \forall d \in D, \dots, \forall x \in X, \forall l \in L$ such that they have a common property/feature that provides integration/interoperability between the elements a, b, d, \dots, x, l and establishes C_{xy} correspondences between the elements of adjacent sets:

$$M = [\{a, b \mid \forall a \in A, \forall b \in B, (a, b) \in C_{ab} \cup \\ \cup \{b, a \mid \forall b \in B, \forall a \in A, (b, a) \in C_{ba}\}] \cup \\ \cup [\{b, d \mid \forall b \in B, \forall d \in D, (b, d) \in C_{bd}\} \cup \\ \cup \{d, b \mid \forall d \in D, \forall b \in B, (d, b) \in C_{db}\}] \cup \dots \\ \dots \cup [\{x, l \mid \forall x \in X, \forall l \in L, (x, l) \in C_{xl}\} \cup \\ \cup \{l, x \mid \forall l \in L, \forall x \in X, (l, x) \in C_{lx}\}].$$

The practical application aims to solve the problem of forming common pairs, triples, quadruples, etc. of elements of any number of arbitrary finite sets A, B, D, \dots, X, L on the basis of their defined common property/feature, which provides integration/interoperability between the elements $a \in A, b \in B, d \in D, \dots, x \in X, l \in L$ and establishes the correspondences $C_{abd\dots xla}$ and $C_{alx\dots dba}$ between the elements of these sets.

Definition 3. The set M , such that

$$\forall m \in M, m = \{a, b, d, \dots, x, l, a \mid a \in A, b \in B, d \in D, \dots, x \in X, l \in L, \\ (a, b, d, \dots, x, l, a) \in C_{abd\dots xla}\} \mid m = \{a, l, x, \dots, d, b, a \mid a \in A, \forall l \in L, \\ x \in X, \dots, d \in D, b \in B, (a, l, x, \dots, d, b, a) \in C_{alx\dots dba}\}$$

of interdisciplinary pairs, triples, quadruples, etc. of elements — a set of pairs, triples, quadruples, etc. of elements that can be formed by elements of arbitrary finite sets A, B, D, \dots, X, L having a common property/feature that determines the correspondences $C_{abd\dots xla}$ and $C_{alx\dots dba}$ between these elements. Hence:

$$\begin{aligned}
 M = & \{a,b,d,\dots,x,l,a \mid \forall a \in A, \forall b \in B, \forall d \in D, \dots, \forall x \in X, \forall l \in L, \\
 & (a,b,d,\dots,x,l,a) \in C_{abd\dots xla}\} \cup \{a,l,x,\dots,d,b,a \mid \forall a \in A, \forall l \in L, \\
 & \forall x \in X, \dots, \forall d \in D, \forall b \in B, (a,l,x,\dots,d,b,a) \in C_{alx\dots dba}\}. \quad (1)
 \end{aligned}$$

In equation (1), subset $\{a,b,d,\dots,x,l,a \mid \forall a \in A, \forall b \in B, \forall d \in D, \dots, \forall x \in X, \forall l \in L, (a,b,d,\dots,x,l,a) \in C_{abd\dots xla}\} \subseteq A \times B \times D \times \dots \times X \times L \times A$ given correspondence $C_{abd\dots xla}$, and subset $\{a,l,x,\dots,d,b,a \mid \forall a \in A, \forall l \in L, \forall x \in X, \dots, \forall d \in D, \forall b \in B, (a,l,x,\dots,d,b,a) \in C_{alx\dots dba}\} \subseteq A \times L \times X \times \dots \times D \times B \times A$ given correspondence $C_{alx\dots dba}$.

Thus, formula (1) will be rewritten:

$$\begin{aligned}
 M = & C_{abd\dots xla} \cup C_{alx\dots dba} = \{\text{tuple}(A,B,D,\dots,X,L,A,G_{abd\dots xla}) \\
 & \cup \text{tuple}(A,L,X,\dots,D,B,A,G_{alx\dots dba})\}, \quad (2)
 \end{aligned}$$

where $G_{abd\dots xla}$ and $G_{alx\dots dba}$ — graphs/diagrams/matrices of correspondences $C_{abd\dots xla}$ and $C_{alx\dots dba}$ respectively.

Then, for three sets — components of Data Science: formalized data resources D ; theoretical and applied data processing tools I ; business processes of data, information and knowledge extraction by means of theoretical and applied data processing instruments P , we formalize the definition of Data Science on the basis of the updated Venn diagram (Fig. 3).

Definition 4. The definition “Data Science — interdisciplinary science and methodology of representing activities for analysis and extraction of data, information, and knowledge” can be formalized as a set of triples of elements of interdisciplinary resources from three resource sets: Data D , Instruments I , and Processes P , such that having a common property/feature that provides integration/interoperability between the elements $d \in D, i \in I, p \in P$, and establishes the correspondences C_{dipd} and C_{dpid} between the elements of these sets. That is:

$$\begin{aligned}
 DS = & \{d,i,p,d \mid \forall d \in D, \forall i \in I, \forall p \in P, (d,i,p,d) \in C_{dipd}\} \\
 & \cup \{d,p,i,d \mid \forall d \in D, \forall p \in P, \forall i \in I, (d,p,i,d) \in C_{dpid}\}.
 \end{aligned}$$

And expression (2) in the context of Data Science will look like:

$$M = C_{dipd} \cup C_{dpid} = \{\text{tuple}(D,I,P,D,G_{dipd}) \cup \text{tuple}(D,P,I,D,G_{dpid})\},$$

where G_{dipd} i G_{dpid} — graphs/diagrams/matrices of correspondences C_{dipd} and C_{dpid} respectively.

Consequently, we formalize the structural representation of Data Science of the updated Venn diagram by the presence of a property/feature, that provides integration/interoperability between elements $d \in D, i \in I, p \in P$, and establishes the correspondences C_{dipd} and C_{dpid} between the elements of resource sets D , I , and P .

DYNAMICS OF DEVELOPMENT OF STRUCTURAL REPRESENTATION OF DATA SCIENCE IN THE FORMAT OF THE UPDATED VENN DIAGRAM

Figure 6 depicts a part of the updated Venn Data Science diagram with sets of integrated interdisciplinary resources in the form of intersecting triangles.

We will show how the areas of integrated interdisciplinary resources can be narrowed in favor of Data Science on the example of two extreme cases:

1. Figure 6 — structural representation of Data Science as a partial intersection of integrated interdisciplinary resources (excluding areas of non-integrated resources 1, 2, and 3). For simplicity of visualization, triangles 1–2, 1–3, and 2–3 are equal, but, of course, in practice, the cardinalities of the corresponding sets may differ. Generally, the comparison of the sets of pairs of elements with correspondences of different nature is incorrect.

Central area — Data Science can be expanded in one of three possible directions by moving one of the three sides of the central triangle outward (see arrows on Fig. 6). The following transformations may take place:

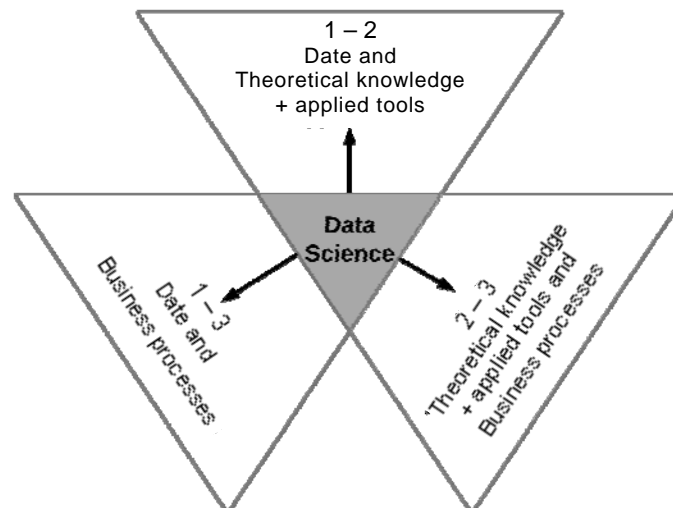


Fig. 6. Data Science at the intersection of integrated interdisciplinary resources

- or in the case of the emergence of a new third element that will cover the existing pair of interdisciplinary resources, belonging to the set adjacent to Data Science (an innovative business process, that allows to organize at the meta-level processing of existing data with existing theoretical knowledge and applied tools, has been discovered) — Data Science domain expansion with the advent of new resources;

- or in the case of the appearance of a new pair of integrated resources that can be covered by an existing third (method of processing a certain new type of data, that can be subordinated to existing business processes, has been invented) — Data Science domain expansion with the emergence of new pairs of resources, links between them.

2. Theoretically, the option of a complete expansion of the Data Science domain with full correspondence and imposition of integrated interdisciplinary resource areas is possible as well (Fig. 7).

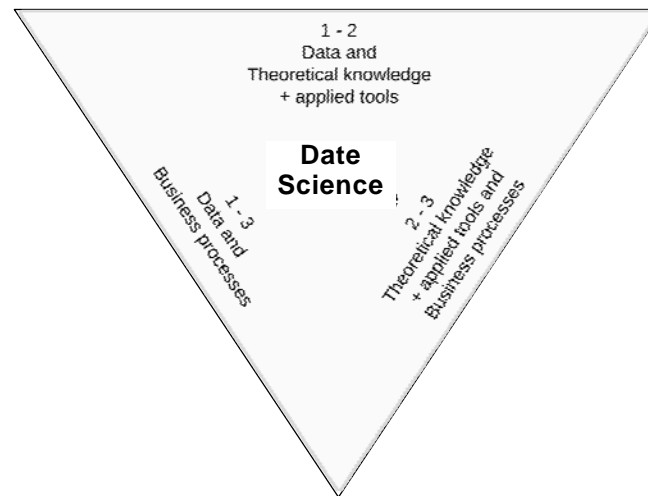


Fig. 7. Idealistic example of the complete expansion of Data Science

Based on the cited scientific results of authoritative authors, as well as on the detailed decomposition and justification of the structural representation of Data Science, we propose the following definition of Data Science:

Data Science — interdisciplinary science and methodology of representing activities for analysis and extraction of data, information, and knowledge.

This definition of Data Science, from our point of view, more closely unites both Science and Data in the methodology of scientific and practical activities for the analysis and extraction of data, information, and knowledge.

CONCLUSION

1. During this study, we examined the main scientific results on Data Science, the numerous debates about its right to exist as a separate field. The ambiguity of the existing definitions of Data Science has been established, in particular the incompleteness of individual elements of Drew Conway's Data Science Venn diagram [2] and the vague meaning of its components, which does not fully reflect the required set of skills for data scientists and engineers of Data Science systems.

2. We propose a unified structural representation of Data Science in the format of an updated Venn diagram based on a property/feature that establishes correspondences that provide integration/interoperability between the elements of the sets of the Venn diagram.

3. A unified diagram of the Data Science domain at the intersection of triangles of integrated interdisciplinary resources is presented and the potential for expansion of this domain is demonstrated.

4. The new definition of Data Science as an interdisciplinary science and methodology of representing activities for analysis and extraction of data, information, and knowledge is substantiated.

REFERENCES

1. Thomas Davenport and D.J. Patil, "Data Scientist: The Sexiest Job of the 21st Century", *Harvard Business Review*, October 2012.
2. Drew Conway, "The Data Science Venn Diagram", *Personal blog*, September 30, 2010.

3. Cathy O’Neil and Rachel Schutt, *Doing data science: Straight talk from the front-line*. O’Reilly Media, Inc., 2013.
4. Vasant Dhar, “Data science and prediction”, *Communications of the ACM*, 56.12, pp. 64–73, 2013.
5. Jake Vanderplas, *Python data science handbook: Essential tools for working with data*. O’Reilly Media, Inc., 2016.
6. Annalyn Ng and Kenneth Soo, “Data Science for the Layman: No Math Added”, *Numsense!*, 2017.
7. Provost Foster and Tom Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O’Reilly Media, Inc., 2013.
8. Provost Foster and Tom Fawcett, “Data science and its relationship to big data and data-driven decision making”, *Big data*, 1.1, pp. 51–59, 2013.
9. Matthew A. Waller and Stanley E. Fawcett, “Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management”, *Journal of Business Logistics*, 34.2, pp. 77–84, 2013.
10. Bohdan Pavlyshenko, “Subjective view on Data Science in Ukraine”, *dou.ua article*, January 9, 2017.
11. Jeff Leek, “The key word in “Data Science” is not Data, it is Science”, *Simply Statistics*, December 12, 2013.
12. J.W. Tukey, “Sunset salvo”, *The American Statistician*, 40(1), pp. 72–76, 1986.
13. J.W. Tukey, *Exploratory data analysis*, 1977.
14. J.W. Tukey, “The future of data analysis”, *The annals of mathematical statistics*, 33(1), pp. 1–67, 1962.
15. N.J. Nilsson, *The quest for artificial intelligence*. Cambridge University Press, 2009.
16. A.L. Samuel, “Some studies in machine learning using the game of checkers”, *IBM Journal of research and development*, 3(3), pp. 210–229, 1959.
17. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning”, *Nature*, 521(7553), pp. 436–444, 2015.
18. P. Russom, “Big data analytics”, *TDWI best practices report, fourth quarter*, 19(4), pp. 1–34, 2011.
19. C.H. Chen, W.K. Härdle, and A. Unwin (Eds.), *Handbook of data visualization*. Springer Science & Business Media, 2007.
20. R. Ihaka and R. Gentleman, “R: a language for data analysis and graphics”, *Journal of computational and graphical statistics*, 5(3), pp. 299–314, 1996.
21. M. Abadi et al., “Tensorflow: A system for large-scale machine learning”, in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pp. 265–283, 2016.
22. D.J. Higham and N.J. Higham, *MATLAB guide*. Society for Industrial and Applied Mathematics, 2016.
23. B. Maxfield, *Essential PTC® Mathcad Prime® 3.0: A guide for new and current users*. Academic Press, 2013.
24. J.P.M. De Sá, *Applied statistics using SPSS, Statistica, MatLab and R*. Springer Science & Business Media, 2007.
25. R. Collobert, S. Bengio, and J. Mariéthoz, *Torch: a modular machine learning software library* (No. REP_WORK). Idiap, 2002.
26. X. Meng et al., “Mllib: Machine learning in apache spark”, *The Journal of Machine Learning Research*, 17(1), pp. 1235–1241, 2016.
27. H. Wimmer and L.M. Powell, “A comparison of open source tools for data science”, *Journal of Information Systems Applied Research*, 9(2), pp. 4, 2016.
28. A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing Ltd., 2017.
29. F. Pedregosa et al., “Scikit-learn: Machine learning in Python”, *Journal of machine Learning research*, 12, pp. 2825–2830, 2011.
30. E. Loper and S. Bird, “Nltk: The natural language toolkit”, *arXiv preprint cs/0205028*, 2002.
31. C. Adams, *Learning Python data visualization*. Packt Publishing Ltd., 2014.

32. C. Rossant, *Learning IPython for interactive computing and data visualization*. Packt Publishing Ltd., 2013.
33. A.N. Kolmogorov and S.V. Fomin, *Introductory real analysis*. Courier Corporation, 1975.

Received 01.03.2021

INFORMATION ON THE ARTICLE

Pavlo P. Maslianko, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: mppdom@i.ua

Yevhenii P. Sielskyi, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: youdjin.sel15@gmail.com

DATA SCIENCE — ДЕФІНІЦІЯ ТА СТРУКТУРНЕ ПОДАННЯ / П.П. Маслянко, Є.П. Сельський

Анотація. Робота є продовженням дискусії щодо існуючих означень та формалізації дефініції «наука про дані» (Data Science) як автономної дисципліни, галузі знань, уточнення її визначальних складових, інтеграції і процесів взаємодії між ними. Зазначено, що в більшості наукових результатів прослідковується датацентричний характер подання й аналізу цієї дисципліни, тобто акцентування на слові Data. Аналіз частоти вживання ключових термінів в означеннях науки про дані (Data Science) показує, на що саме робиться основний акцент та на які терміни означень науки про дані спираються. Внесено й аргументовано певні доповнення до діаграми Венна Дрю Конвея, яка не відображає всіх ресурсів складових, що характеризують прикладний характер науки про дані і не розкриває взаємодію цих ресурсів ані з точки зору дослідника даних, ані в її глобальному розумінні. Запропоновано уніфіковане структурне подання Data Science у форматі оновленої діаграми Венна Дрю Конвея на основі властивості/ознаки, яка встановлює відповідності, що забезпечують інтеграцію/інтероперабельність між елементами множин діаграми Венна Дрю Конвея. Обґрунтовано нову дефініцію «наука про дані» як міждисциплінарної науки і методології подання діяльності з аналізу і добування даних, інформації та знань.

Ключові слова: наука про дані, діаграма Венна Дрю Конвея, означення науки про дані, структура науки про дані, дані, інформація, знання.

DATA SCIENCE — ДЕФИНИЦИЯ И СТРУКТУРНОЕ ПРЕДСТАВЛЕНИЕ / П.П. Маслянко, Є.П. Сельський

Аннотация. Работа является продолжением дискуссии о существующих определениях и формализации дефиниции «наука о данных» (Data Science) как автономной дисциплины, области знаний, уточнении ее определяющих составляющих, интеграции и процессов взаимодействия между ними. Отмечено, что в большинстве научных результатов прослеживается датацентрический характер представления и анализа этой дисциплины, т.е. акцентирование на слове Data. Анализ частоты употребления ключевых терминов в определениях науки о данных (Data Science) показывает, на что именно делается основной акцент и на какие термины определений науки о данные опираются. Внесены и аргументированы определенные дополнения к диаграмме Венна Дрю Конвея, которая не отражает всех ресурсов составляющих, характеризующих прикладной характер науки о данных и не раскрывает взаимодействие этих ресурсов ни с точки зрения исследователя данных, ни в ее глобальном понимании. Предложено унифицированное структурное представление Data Science в формате обновленной диаграммы Венна Дрю Конвея на основе свойства/признака, устанавливающего соответствия, которые обеспечивают интеграцию/интероперабельность между элементами множеств диаграммы Венна Дрю Конвея. Обоснована новая дефиниция «наука о данных» как междисциплинарной науки и методологии представления деятельности по анализу и извлечению данных, информации и знаний.

Ключевые слова: наука о данных, диаграмма Венна Дрю Конвея, определение науки о данных, структура науки о данных, данные, информация, знания.



**ЗАГАЛЬНА МЕТОДИКА ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ
НЕСТАЦІОНАРНИХ ПРОЦЕСІВ НА ОСНОВІ
МАТЕМАТИЧНИХ МОДЕЛЕЙ З ВИКОРИСТАННЯМ
СТАТИСТИЧНИХ ДАНИХ**

О.М. БЕЛАС, А.О. БЕЛАС

Анотація. Розглянуто проблематику прогнозування нелінійних нестационарних процесів, поданих у вигляді часових рядів, що можуть собою описувати динаміку процесів як у технічних, так і в економічних системах. Детально описано загальну методику аналізу таких даних і побудови відповідних математичних моделей на базі авторегресійних моделей та рекурентних нейронних мереж. Методику застосовано на практичних прикладах — виконано порівняльний аналіз моделей прогнозування кількості каналів обслуговування абонентів стільникового зв'язку для конкретної базової станції, виявлено переваги та недоліки кожного з методів. Сформульовано необхідність удосконалення існуючої методики та розробленні нового підходу.

Ключові слова: математичне моделювання, оброблення сигналів, нестационарні процеси, авторегресійні моделі, нейронні мережі, рекурентні нейронні мережі.

ВСТУП

Прогнозування на основі моделей, побудованих за експериментальними (статистичними) даними — один з найпопулярніших підходів до прогнозування динаміки процесів у технічних, соціально-економічних та фінансових системах. Завдання прогнозування нелінійних нестационарних процесів на тепер є дуже актуальним.

Розвиток інформаційних технологій і розширення обсягів інформаційних послуг значною мірою ґрунтується на науково-технологічних розробках у галузі телекомунікаційних мереж. Дослідження показують, що сучасні мережеві технології своїм зростанням випереджають теоретичне та аналітичне розуміння мережевих взаємодій. Методи розрахунку характеристик телекомунікаційної мережі (пропускної здатності каналів, ємності буферів та ін.), що засновані на класичних моделях, не відповідають необхідним вимогам і не дозволяють адекватно оцінювати навантаження в мережі.

У праці [1] встановлено, що завдання прогнозування процесів у технічних системах ґрунтовно проаналізовано із застосуванням класичних регресійних підходів, які доволі просто використовуються як у теоретичному аспекті, так і в обчислювальному. Проте такий підхід має певні недоліки.

Тому у праці [1] запропоновано розглянути нейронні мережі типу LSTM, які також вирішують завдання моделювання послідовностей, до того ж ураховують нелінійний або комбінований вплив зовнішніх факторів, але мають деякі недоліки отриманих моделей (їх складність).

У цьому дослідженні сформульовано й описано загальну методику прогнозування нелінійних нестационарних процесів на основі математичних моделей, застосовано методологію на практичних прикладах — виконано порівняльний аналіз моделей прогнозування кількості каналів обслуговування абонентів стільникового зв'язку для конкретної базової станції.

ОПИС МЕТОДИКИ

Основи першої методики побудови моделей часових рядів запропонували Бокс і Дженкінс у праці [2]. Модифікована авторами методика побудови математичної моделі процесу, що складається з декількох кроків, поклала основу процесу моделювання та прогнозування часових рядів. Із розвитком математичного моделювання, появою нових завдань і методів їх розв'язання методика модифікувалась і поліпшувалась.

Існує кілька методик побудови таких математичних моделей: KDD, SEMMA, CRISP-DM [3].

У цій роботі пропонується використовувати сучасну аналітичну методологію SEMMA (Sampling, Exploring, Modifying, Modeling, Assessing) [4]. Методологія добре орієнтована на побудову моделей процесів, для яких, як правило, набагато складніше поставити експеримент та отримати інформативні експериментальні дані в достатньому обсязі. Методологія більш орієнтована саме на процес дослідження, більше фокусуючись на побудові моделей, а не на бізнес-розумінні або інтеграції з технічними системами.

1. Sampling — відбір і завантаження даних у проект, попереднє оброблення експериментальних даних. На цьому етапі виконуються ретельний відбір даних, фільтрація, видалення пропусків та пошкоджених або некоректних даних, зведення до найбільш зручного вигляду, підготовка остаточної «вітрини даних» до роботи. Формально можна виокремити такі операції на цьому етапі:

- коригування даних — заповнення пропусків та зменшення викидів (екстремальних імпульсних значень), що виходять за основний діапазон значень змінних. Некоректні виміри замінюються інтерпольованими або усередненими даними;

- ортогональні перетворення та цифрова фільтрація даних з метою вилучення шумових складових (за необхідності);

- нормування даних: їх логарифмування або зведення до зручного діапазону їх зміни, наприклад, від 0 до 1; від -1 до +1; від +10 до -10 і т.ін.

2. Exploring — розуміння суті даних, пошук трендів, аномалій та взаємозв'язків. Для цього використовуються візуалізація, кластеризація та асоціація даних. Виконується аналіз даних на можливу наявність нелінійностей за допомогою множини статистичних критеріїв. Використовуються тести на тренд та на гетероскедастичність для розуміння належності досліджуваного процесу до певного класу нестационарності [5].

3. Modifying — модифікування даних шляхом створення, обрання та перетворення змінних; остаточне перетворення вихідних даних, розбиття

вибірки на навчальну і тестову. Правильне обрання тестової вибірки є важливим завданням та запорукою коректного навчання моделі.

4. Modeling — складається з аналітичних інструментів, призначених для побудови обраних математичних моделей, що дають бажаний результат. У праці [1] виконано огляд сучасних підходів до прогнозування процесів, подано статистичні дані у вигляді часових рядів. Розглянуто модель авторегресії з інтегрованим ковзним середнім (ARIMA):

$$y'(k) = a_0 + \sum_{i=1}^p a_i y'(k-i) + \sum_{j=1}^q b_j \varepsilon(k-j) + \varepsilon(k).$$

Методику автоматичної побудови такої моделі для обраного ряду досліджено та описано Хайндманом у праці [6].

Розглянуто нейронні мережі типу LSTM. Мережі з довгочотроковою пам'яттю (Long Short Term Memory) — спрощено LSTM — особливий вид рекурентних нейронних мереж, здатних до навчання довгочотрокових залежностей. Їх запропонували Хохрейтер і Шмідхубер [7] і доопрацювали та популяризували інші дослідники [8, 9]. Вони дають змогу отримати високоякісні результати на великій різноманітності проблем і нині широко застосовуються. Однак поки не існує загальної методики побудови такого типу нейронних мереж, зокрема методики вибору початкових ваг, алгоритму оптимізації, функції активації, архітектури мережі тощо. Через складність даних моделей створення такої методики є надважливим завданням.

5. Assessing — побудова графіків та використання критеріїв для оцінювання якості прогнозів, отриманих у процесі моделювання та дослідження; Обрання кращої з оцінених моделей-кандидатів. Застосування моделі до розв'язання основного завдання — прогнозування, керування, поглиблене дослідження процесу та остаточне встановлення її придатності.

МЕТРИКИ ДЛЯ ОЦІНЮВАННЯ ЯКОСТІ ПРОГНОЗІВ

Щоб визначити найбільш придатну модель для розв'язання певної задачі, використовують різні метрики для оцінювання якості прогнозів.

Для оцінювання якості прогнозів зазвичай використовують множину взаємно доповнювальних статистичних критеріїв. Наприклад, значення середньоквадратичної похибки залежить від масштабу даних, а тому недостатньо використовувати тільки цей статистичний параметр для аналізу якості прогнозу. Поглиблене оцінювання якості прогнозів досягається з використанням критеріїв, які дають відносні оцінки у відсотках. Найпоширеніші статистичні критерії якості прогнозу та їх призначення [10]:

— середньоквадратична похибка моделі

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^N [\hat{y}(k) - y(k)]^2}{N}};$$

— середня абсолютна похибка моделі

$$\text{MAE} = \frac{\sum_{k=1}^N |\hat{y}(k) - y(k)|}{N};$$

— середня абсолютна похибка моделі у відсотках

$$MAPE = \frac{100}{N} \sum_{k=1}^N \left| \frac{\hat{y}(k) - y(k)}{y(k)} \right|,$$

де $\hat{y}(k)$ — прогнозоване за моделлю значення; $y(k)$ — реальне вимірювання; N — довжина вибірки.

Із можливих кандидатів необхідно обирати ту модель, для якої RMSE, MAE і MAPE набувають мінімального значення.

Однак трапляються випадки, коли деяка модель є пріоритетною за одними критеріями, а інша модель — за іншими. У таких випадках потрібно обрати пріоритетний критерій або побудувати інтегральний критерій оцінювання якості прогнозів, що наразі є дуже актуальним і досі невіршеним завданням.

ПРИКЛАДИ ЗАСТОСУВАННЯ МЕТОДИКИ НА СТАТИСТИЧНИХ ДАНИХ

Побудовані моделі для прогнозування необхідної кількості каналів обслуговування абонентів стільникового зв'язку деякого українського оператора у формі регресії та LSTM нейронних мереж.

Реалізація алгоритмів спирається на праці [11, 12].

Основні показники точності даних моделей та якості оцінювання RMSE, MAE, MAPE подано у табл. 1–3.

Побудовані графіки порівняння реальних та прогнозованих значень різними моделями показано на рис. 1–5.

Таблиця 1. Порівняльна характеристика моделей прогнозування кількості каналів обслуговування абонентів стільникового зв'язку (базова станція 1)

Тип моделі	Оцінки якості прогнозу		
	RMSE	MAE	MAPE
АРИКС(2,1,2)	0,5392346	0,404728	9,289125
Нейронна мережа типу LSTM	0,3718906	0,288897	1,988698

Таблиця 2. Порівняльна характеристика моделей прогнозування кількості каналів обслуговування абонентів стільникового зв'язку (базова станція 2)

Тип моделі	Оцінки якості прогнозу		
	RMSE	MAE	MAPE
АРИКС(2,1,2)	0,6349195	0,474474	1,587532
Нейронна мережа типу LSTM	0,4224225	0,315366	0,961578

Таблиця 3. Порівняльна характеристика моделей прогнозування кількості каналів обслуговування абонентів стільникового зв'язку (базова станція 3)

Тип моделі	Оцінки якості прогнозу		
	RMSE	MAE	MAPE
АРИКС(3,1,2)	1,19619	1,019926	2,341559
Нейронна мережа типу LSTM	0,7199281	0,613519	1,314227

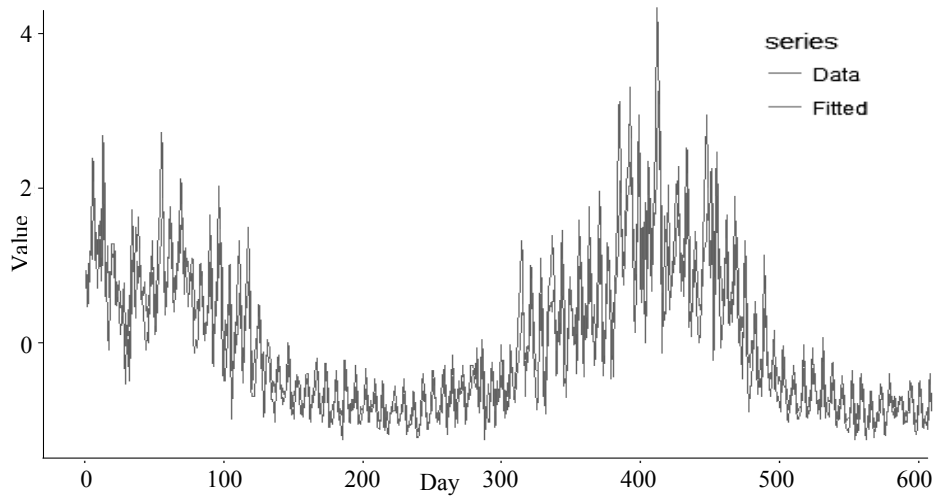


Рис. 1. Порівняльний графік реальних та прогнозованих значень регресійної моделі для базової станції 1 стільникового зв'язку

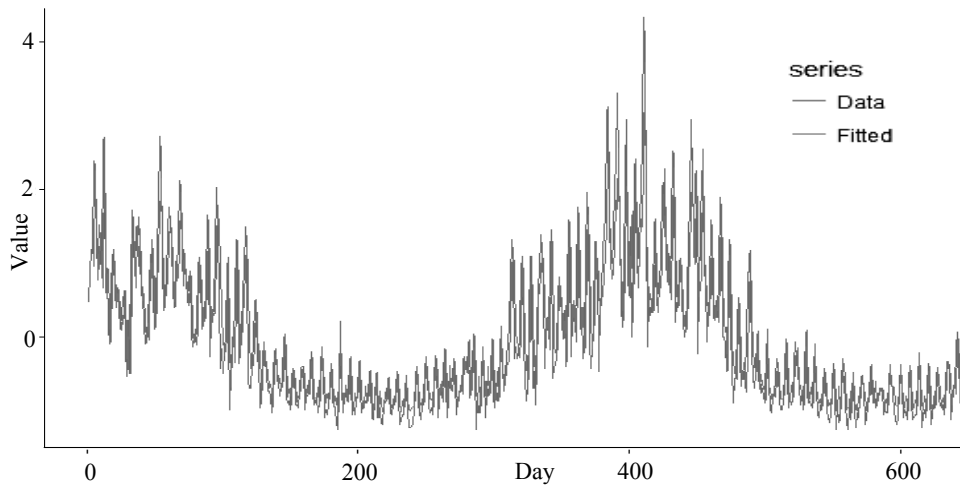


Рис. 2. Порівняльний графік реальних та прогнозованих значень нейронної мережі типу LSTM для базової станції 1 стільникового зв'язку

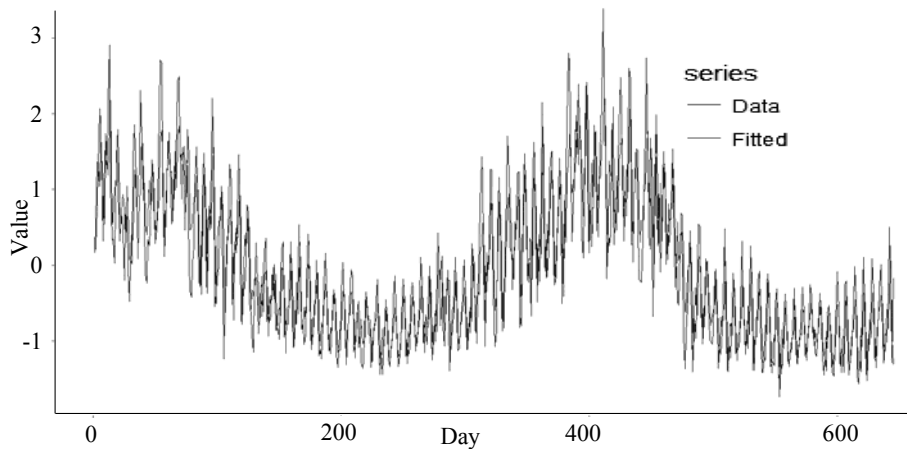


Рис. 3. Порівняльний графік реальних та прогнозованих значень нейронної мережі типу LSTM для базової станції 2 стільникового зв'язку

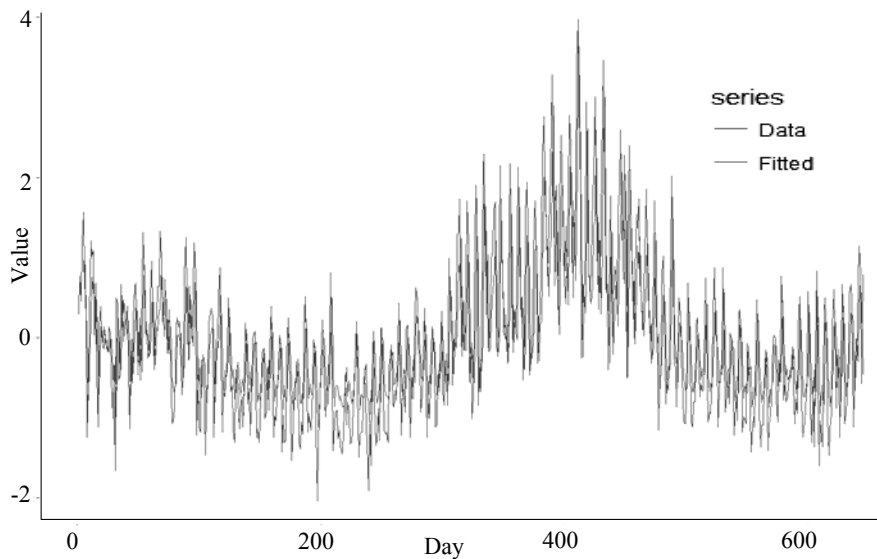


Рис. 4. Порівняльний графік реальних та прогнозованих значень регресійної моделі для базової станції 3 стільникового зв'язку

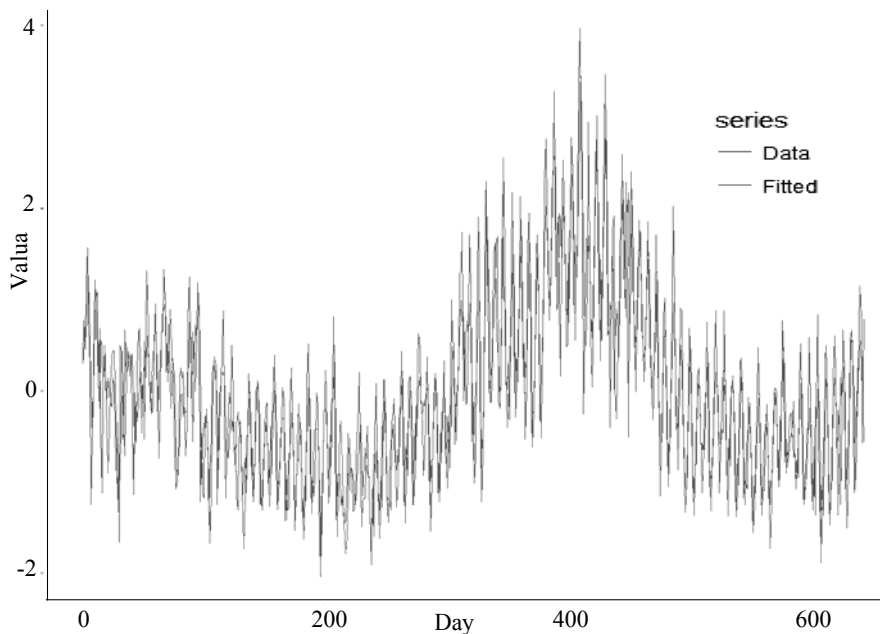


Рис. 5. Порівняльний графік реальних та прогнозованих значень нейронної мережі типу LSTM для базової станції 3 стільникового зв'язку

АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ

У ході дослідження кращою виявилася модель з використанням нейронних мереж (МАРЕ=1,98). Регресійні методи виявились менш точними (МАРЕ=9,29), однак швидшими для побудови і використання. Серед регресійних моделей найкращою виявилась модель АРІКС(2,1,2). Обрання відповідної структури моделі для автоматизованої роботи в системі на час дослідження спирається саме на важливість швидкості або точності обчислень, що підтверджує проблематику, описану у праці [1].

ВИСНОВКИ

Сформульовано й описано основні етапи загальної методики прогнозування нелінійних нестационарних процесів на основі сучасної аналітичної методології SEMMA. Показано досліджену методику для автоматичної побудови регресійних моделей, сформульовано проблеми існуючої методики для побудови моделей на основі нейронних мереж. Сформульовано проблему єдиного критерію якості прогнозів. Побудовано нові моделі для прогнозування вибраних процесів і у формі регресії та нейронних мереж типу LSTM, показано їх переваги та недоліки.

ЛІТЕРАТУРА

1. O.M. Belas, P.I. Bidyuk, and A.O. Belas, “Comparative analysis of autoregressive approaches and recurrent neural networks for modeling and forecasting of nonlinear nonstationary processes”, *Information Technology and Security*, vol. 7, no. 1, pp. 91–99, 2019.
2. J. Box and G. Jenkins, *Time Series Analysis*. Moscow, USSR: Mir, 1974.
3. A. Azevedo and M. Santos, “KDD, SEMMA and CRISP-DM: a parallel overview”, in *Proc. of the IADIS European Conference on Data Mining*, pp. 182–185, 2008.
4. *SEMMA in SAS Enterprise Miner* [Online]. Available: <https://web.archive.org/web/20120308165638/http://www.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html/>. Accessed on: Sep. 12, 2020.
5. R. Shumway and D. Stoffer, *Time Series Analysis and Its Applications*. New York, USA: Springer, 2011. doi: 10.1007/978-1-4757-3261-0.
6. R. Hyndman and Y. Khandakar, “Automatic time series forecasting: the forecast package for R”, *Journal of Statistical Software*, vol. 27, no. 3, 2008.
7. S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory”, *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
8. F.A. Gers, D. Eck, and J. Schmidhuber, “Applying LSTM to Time Series Predictable Through Time-Window Approaches”, in *Proc. of International Conference on Artificial Neural Networks*, pp. 669–676, 2001.
9. S. Hochreiter, Y. Bengio, and J. Schmidhuber, *Gradient flow in recurrent nets: the difficulty of learning long-term dependencies* [Online]. Available: <http://www.bioinf.jku.at/publications/older/ch7.pdf>. Accessed on: Dec. 12, 2018.
10. P.I. Bidyuk, V.D. Romanenko, and O.L. Timoschuk, *Analysis of time series*. Kyiv, Ukraine: Polytechnic, 2010.
11. R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Melbourne, Australia: OTexts, 2013.
12. F. Chollet, *Deep Learning with R*. New York, USA: Manning, 2017.

Надійшла 28.11.2020

INFORMATION ON THE ARTICLE

Oleg M. Belas, ORCID: 0000-0002-1595-3029, Institute of Special Communication and Information Security of National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: belas@ukr.net

Andrii O. Belas, ORCID: 0000-0001-7883-2489, Educational and Scientific Complex “Institute for Applied System Analysis” of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: andrii.belas@gmail.com

GENERAL METHODS OF FORECASTING NONLINEAR NONSTATIONARY PROCESSES BASED ON MATHEMATICAL MODELS USING STATISTICAL DATA / O.M. Belas, A.O. Belas

Abstract. The article considers the problem of forecasting nonlinear nonstationary processes, presented in the form of time series, which can describe the dynamics of processes in both technical and economic systems. The general technique of analysis of such data and construction of corresponding mathematical models based on autoregressive models and recurrent neural networks is described in detail. The technique is applied on practical examples while performing the comparative analysis of models of forecasting of quantity of channels of service of cellular subscribers for a given station and revealing advantages and disadvantages of each method. The need to improve the existing methodology and develop a new approach is formulated.

Keywords: mathematical modeling, signal processing, nonstationary processes, autoregressive models, neural networks, recurrent neural networks.

ОБЩАЯ МЕТОДИКА ПРОГНОЗИРОВАНИЯ НЕЛИНЕЙНЫХ НЕСТАЦИОНАРНЫХ ПРОЦЕССОВ НА ОСНОВЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ СТАТИСТИЧЕСКИХ ДАННЫХ / О.Н. Белас, А.О. Белас

Аннотация. Рассмотрена проблематика прогнозирования нелинейных нестационарных процессов, представленных в виде временных рядов, которые могут собой описывать динамику процессов как в технических, так и в экономических системах. Подробно описана общая методика анализа таких данных и построения соответствующих математических моделей на базе авторегрессионных моделей и рекуррентных нейронных сетей. Методика применена на практических примерах — выполнен сравнительный анализ моделей прогнозирования количества каналов обслуживания абонентов сотовой связи для конкретной базовой станции, выявлены преимущества и недостатки каждого из методов. Сформулирована необходимость усовершенствования существующей методики и разработки нового подхода.

Ключевые слова: математическое моделирование, обработка сигналов, нестационарные процессы, авторегрессионные модели, нейронные сети, рекуррентные нейронные сети.

ПРОГНОЗУВАННЯ ОБСЯГІВ ВИРОБНИЦТВА СІЛЬСЬКОГОСПОДАРСЬКОЇ ПРОДУКЦІЇ ЗА ДОПОМОГОЮ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

**С.М. КОНЦЕБА, Р.І. ЛІЩУК, С.Д. СКУРТОЛ, Г.Ю. РОДАЩУК,
І.П. ВАСИЛЬЧЕНКО**

Анотація. Спрогнозовано майбутні значення показників виробництва зернових і зернобобових культур у сільськогосподарських підприємствах Черкаської області на основі часових рядів динаміки, рівні яких виражено в натуральних одиницях вимірювання. Під час дослідження використано одну з методик інтелектуального аналізу даних — аналіз часових рядів, що дозволяє на основі моделі ряду динаміки, побудованої на базі фактичних значень попередніх показників минулих років, спрогнозувати виробництво на наступні роки. Методика містить такі етапи: графічний аналіз (для підбору відповідного рівняння моделі), виділення і аналіз детермінованих складових ряду, згладжування і фільтрація часового ряду, дослідження випадкової компоненти, побудова і перевірка на адекватність моделі часового ряду, прогнозування поведінки часового ряду на підставі проведених досліджень.

Ключові слова: інтелектуальний аналіз даних, часовий ряд, прогнозування, тренд, сільськогосподарське підприємство, модель.

ВСТУП

Питання прогнозування економічних процесів стоять досить гостро в періоди нестійкого економічного стану виробництва сільськогосподарської продукції. Застосування методик інтелектуального аналізу даних дозволяє підвищити ефективність виробництва продукції в сільськогосподарських підприємствах за рахунок оптимального управління запасами, доходами, витратами. Зростання ролі прогнозування розвитку виробництва сільськогосподарської продукції зумовлено ускладненням управлінських функцій в умовах динамічної і нестабільної ситуації як у зовнішньому середовищі, так і на самому ринку (поведінка фермерських господарств, які суттєво впливають на ринкову ситуацію, експорт-імпорт, особливо в умовах відкритості ринку, тінізація економіки тощо). Поряд з тим виробництво і реалізація сільськогосподарської продукції в Україні займає значну частку в економіці країни, тому використання методів інтелектуального аналізу для прогнозування показників діяльності підприємств агропромислового комплексу є актуальним.

Проблеми прийняття рішень в економічних системах на основі прогнозування показників з використанням методів інтелектуального аналізу даних розглянуто у працях вітчизняних авторів Р.О. Петрова, О.Я. Кучерука [1]. Прогнозуючи терміни продажу товарів, звертають увагу на те, що для прогнозування продажів найчастіше використовуються класичні методи аналізу

часових рядів та дерева рішень. П. І. Бідюк, С.М. Савченко, А.С. Савченко [2] визначили переваги методів інтелектуального аналізу та їх комбінацій (гібридні методи інтелектуального аналізу даних) для прийняття раціональних рішень у системі управління та прогнозування конкурентоспроможності вітчизняних підприємств. Група вчених Г. Чорноус, С. Рибальченко [3], П.І. Бідюк, А.В. Федоров [4], Л.О. Коршевнюк [5], О.Ю. Берзлев, М.М. Маляр, В.В. Ніколенко [6, 7] застосовують методи інтелектуального аналізу даних для прогнозування процесів ціноутворення, біржових показників, оцінювання фінансових ризиків.

Значно менше уваги звертають учені на використання методів інтелектуального аналізу даних для прогнозування показників діяльності підприємств агропромислового комплексу. В.Д. Кишенько, М.А. Сич [8] зазначають, що інтелектуальний аналіз даних, а саме часові ряди, дозволяє визначити стан технологічного комплексу цукрового заводу, зробити висновки про якість його функціонування, дати рекомендації щодо пошуку й усунення проблемних ситуацій. І.І. Глаголева, А.Ю. Берко [9] вважають, метод інтелектуального аналізу даних на основі прогнозування часових рядів для земельного кадастру дає змогу аналізувати дані та прогнозувати майбутні значення показників використання земельних ресурсів.

Питання прогнозування виробництва продукції в сільськогосподарських підприємствах з використанням часових рядів є недостатньо вивченим, залишається дискусійним і потребує подальших досліджень.

Мета роботи — спрогнозувати майбутні значення показників виробництва зернових і зернобобових культур у сільськогосподарських підприємствах Черкащини на 2020–2021 рр. на основі часових рядів динаміки, рівні яких виражені в натуральних одиницях вимірювання.

МЕТОДИКА ДОСЛІДЖЕНЬ

Під час дослідження використовувалась одна з методик інтелектуального аналізу даних, а саме: аналіз часових рядів для того, щоб на основі моделі ряду динаміки, побудованої на базі фактичних значень показника за 2000–2019 рр., виконати прогноз виробництва на 2020–2021 рр.

Методика містить такі етапи: графічний аналіз, виділення й аналіз детермінованих складових ряду, згладжування і фільтрація часового ряду, дослідження випадкової компоненти, побудова і перевірка на адекватність моделі часового ряду, прогнозування поведінки часового ряду на підставі проведених досліджень.

Для прикладу використано значення показників виробництва зернових і зернобобових культур у сільськогосподарських підприємствах Черкащини. Аналогічну методику досліджень можна використати для будь-яких видів продукції сільськогосподарського виробництва як у галузі рослинництва, так і в галузі тваринництва.

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

Показники багатьох економічних явищ і процесів змінюються в часі, характерними особливостями яких є те, що рівень показників у наступному часо-

вому періоді значною мірою залежить від їх рівня в минулому. Тенденцію зміни в часі кількісної міри досліджуваного показника можна подати аналітичною формулою, а у вибраній системі координат — плавною траєкторією або трендом. Послідовність спостережень одного показника (ознаки), упорядкована залежно від послідовно зростаючих або спадних значень часу, називають часовим рядом, який подано обсягами виробництва зернових і зернобобових культур у Черкаській області за 2006–2019 рр. (табл. 1).

Таблиця 1. Обсяги виробництва зернових і зернобобових культур у Черкаській області, тис. ц

Рік	Обсяги виробництва	Рік	Обсяги виробництва
2006	21086,0	2013	40684,8
2007	18424,2	2014	36996,5
2008	29780,3	2015	37454,9
2009	31939,6	2016	40917,0
2010	25310,7	2017	29265,5
2011	37618,8	2018	46440,1
2012	33105,5	2019	45598,3

Джерело: Рослинництво (1995–2019 рр.). Головне управління статистики у Черкаській області. URL: http://www.ck.ukrstat.gov.ua/source/arch/2020/roslinnictvo_19.pdf

Для моделювання тренду потрібно перевірити однорідність та порівнянність рівнів часового ряду y_t , а також визначити можливості їх зіставлення і стійкості. Попередній аналіз даних табл. 1 показує, що обсяги виробництва зернових і зернобобових культур в області за період 2006–2009 рр. нестабільні, варіюють у великому діапазоні, і тому для побудови часового ряду візьмемо дані за 2010–2019 рр., попередньо замінивши рівні часового ряду за 2010, 2011, 2013 та 2017 роки середньоарифметичними попередніх і наступних років (табл. 2).

Таблиця 2. Скориговані обсяги виробництва зернових і зернобобових культур для побудови тренду, тис. ц

Рік	Обсяги виробництва y_t	Рік	Обсяги виробництва y_t
2010	4698,0	2015	5665,0
2011	4739,0	2016	6073,0
2012	5194,0	2017	6034,0
2013	5051,0	2018	6117,0
2014	5257,0	2019	5928,0

Джерело: Розраховано авторами з використанням джерел: Рослинництво (1995–2019 рр.). Головне управління статистики у Черкаській області. URL: http://www.ck.ukrstat.gov.ua/source/arch/2020/roslinnictvo_19.pdf

Перевірені за методом Ірвіна [11, с. 409] рівні y_t скоригованого часового ряду не є аномальними ($\lambda_p < \lambda_{0,05} = 1,3$, якщо $n = 10$).

Для перевірки гіпотез про однорідність дисперсій та наявність тренду застосуємо метод перевірки різниць середніх рівнів.

Для дослідження однорідності дисперсій початковий часовий ряд $y_1, y_2, \dots, y_t, \dots, y_n$ розбивається на дві частини приблизно однієї довжини $n_1 = n_2 = 10/2 = 5$ і для кожної частини розраховуються [11, с. 409] (табл. 3):

- середні значення $y_{\text{сеп}_1} = \frac{\sum_t^{n_1} y_{t_1}}{n_1}$ і $y_{\text{сеп}_2} = \frac{\sum_t^{n_2} y_{t_2}}{n_2}$;
- дисперсії $\delta_1^2 = \frac{\sum_t^{n_1} (y_t - y_{\text{сеп}_1})^2}{(n_1 - 1)}$ і $\delta_2^2 = \frac{\sum_t^{n_2} (y_t - y_{\text{сеп}_2})^2}{(n_2 - 1)}$.

Таблиця 3. Розрахунок середніх значень та дисперсії часового ряду

t	y_t	$y_t - y_{\text{сеп}_1}$	$(y_t - y_{\text{сеп}_1})^2$
1	34779,2	951,1	904667,3
2	29208,1	-4620,0	21344030,4
3	33105,5	-722,6	522093,0
4	35051,0	1222,9	1495582,2
5	36996,5	3168,4	10039012,0
Разом	169140,3	-	34305384,9
Середнє значення $y_{\text{сеп}_1}$	33828,1	Дисперсія δ_1^2	8576346,2
6	37454,9	-5362,9	28760481,9
7	40917,0	-1900,8	3612964,6
8	43678,6	860,8	741011,1
9	46440,1	3622,3	13121202,2
10	45598,3	2780,5	7731291,5
Разом	214088,9	-	53966951,2
Середнє значення $y_{\text{сеп}_2}$	42817,8	Дисперсія δ_2^2	13491738

Джерело. Розраховано авторами з використанням праці [11, с. 409].

Однорідність дисперсій обох частин перевіряється порівнянням розрахованого критерію Фішера [11, с. 410] $F_p = \frac{\delta_2^2}{\delta_1^2} = \frac{13491738,0}{8576346,2} = 1,573$ з теоретичним значенням $F_{(0,05;n_1-1;n_2-1)} = 6,388$. Розраховане значення F_p менше за теоретичне, тому з імовірністю 95% гіпотеза про однорідність дисперсій приймається.

Гіпотеза про наявність тренду перевіряється за допомогою t -критерію Стьюдента, який розраховується за формулою [11, с. 410]

$$t_p = \frac{-|y_{\text{сеп}_1} - y_{\text{сеп}_2}|}{\delta \sqrt{1/n_1 + 1/n_2}} = \frac{-(33808,1 - 42817,8)}{3321,8 \sqrt{1/5 + 1/5}} = 4,279,$$

де δ — середньоквадратичне відхилення різниці середніх, яке розраховується за формулою

$$\delta = \sqrt{\frac{[(n_1 - 1)\delta_1^2 + (n_2 - 1)\delta_2^2]}{n_1 + n_2 - 2}} =$$

$$= \sqrt{\frac{[(5 - 1) * 8576346,3 + (5 - 1) * 13491738,0]}{5 + 5 - 2}} = 3321,8.$$

Розраховане значення $t_p = 4,279$ більше за теоретичне $t_{(0,05; 7)} = 2,365$, тому з імовірністю 95% гіпотеза про існування тренда приймається.

Аналітичне вирівнювання часового ряду виконаємо за методом експоненціального згладжування (табл. 4), використовуючи формулу

$$x_t = \gamma y_t + (1 - \gamma)x_{t-1},$$

де $\gamma = 0,03$ — параметр згладжування;

$$x_0 = x_1 \frac{34779,2 + 29208,1 + 33105,5}{3} = 32364,3 ;$$

x_{t-1} — рівні згладженого часового ряду.

Тенденції зміни часового ряду можуть бути встановлені добиранням емпіричної функції, яка має назву кривої зростання. Вибір форми кривої виконаємо за згладженим рядом на основі графічного зображення часового ряду (на ЕОМ – *Excel – Мастер диаграмм*). Візуально найкращою формою залежності буде поліноміальна крива третього порядку (див. рисунок):

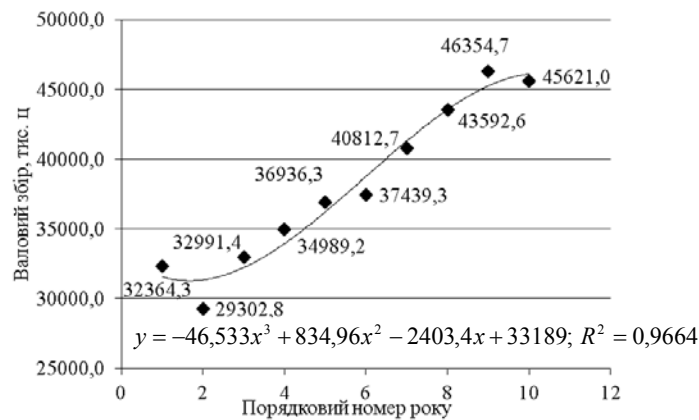
Таблиця 4. Згладжений часовий ряд

n	x_t
1	32364,3
2	29302,8
3	32991,4
4	34989,2
5	36936,3
6	37439,3
7	40812,7
8	43592,6
9	46354,7
10	45621,0

Джерело. Розраховано авторами.

$$y = 33189 - 2103,4t + 834,96t^2 - 46,533t^3.$$

Перевірка адекватності моделі тренду ґрунтується на перевірці виконання в залишках u_t таких властивостей:



Динамічний ряд виробництва зернових і зернобобових культур у Черкаській області за досліджуваний період

1. Незалежність значень рівнів залишків або відсутність між ними автокореляції досліджується за критерієм Дарбіна–Уотсона [11, с. 433] (табл. 5).

Таблиця 5. Розрахунки фактичного значення критерію Дарбіна–Уотсона

y_t	y_{tm}	u_t	u_t^2	u_{t-1}	$u_t - u_{t-1}$	$(u_t - u_{t-1})^2$
34779,2	31574,0	3205,2	10273134,0	–	–	–
29208,1	31349,8	–2141,7	4586776,1	3205,2	–5346,8	28588794,2
33105,5	32237,0	868,5	754207,1	–2141,7	3010,1	9060864,6
35051,0	33956,6	1094,4	1197606,3	868,5	225,9	51031,3
36996,5	36229,4	767,1	588480,8	1094,4	–327,2	107077,5
37454,9	38776,0	–1321,1	1745389,8	767,1	–2088,3	4360817,3
40917,0	41317,4	–400,4	160337,0	–1321,1	920,7	847708,7
43678,6	43574,3	104,3	10869,3	–400,4	504,7	254698,9
46440,1	45267,6	1172,5	1374749,2	104,3	1068,2	1141138,8
45598,3	46118,0	–519,7	270088,1	1172,5	–1692,2	2863530,7
Разом	–	–	20961637,6	–	–	47275662,0

Джерело. Розраховано авторами.

Фактичне значення критерію Дарбіна–Уотсона

$$DW_p = \frac{\sum_t^n (u_t - u_{t-1})^2}{\sum_t^n u_t^2} = \frac{47275662,0}{20961637,6} = 2,255.$$

Розраховане значення критерію Дарбіна–Уотсона $DW_p = 2,181$ міститься в межах $1,5 < DW_p < 2,5$, що свідчить про відсутність автокореляції в моделі тренду.

2. Дослідження відповідності розподілу ймовірностей рівнів залишків нормальному закону виконаємо за RS -методом [11, с. 429]: розраховується величина розмаху між максимальним та мінімальним рівнями ряду залишків

$$R = u_{\max} - u_{\min} = 3205,2 - (-2141,7) = 5346,9$$

та їх стандартне відхилення

$$S = \sqrt{\frac{\sum u_t^2}{n-1}} = \sqrt{\frac{20961637,6}{10-1}} = 1926,13.$$

Розраховане значення

$$RS_p = \frac{R}{S} = \frac{5346,9}{1926,13} = 2,775$$

потрапляє в інтервал між табличними межами, якщо $n = 10$, від 2,67 до 3,18, тому гіпотеза про нормальний закон випадкової складової приймається.

3. Рівність нулю математичного сподівання залишків — якщо ряд залишків підпорядкований нормальному закону розподілу ($S_n = 0$), то розраховується критерій Стюдента за формулою [11, с. 430]

$$t_p = \frac{u_{\text{сеп}}}{\delta_y} \cdot \sqrt{n} = \frac{282,9}{1618,7} \cdot 3,162 = 0,553,$$

де $u_{\text{сеп}}$ — середнє значення залишків; δ_y — середнє квадратичне відхилення; n — довжина часового ряду.

Розраховане значення $t_p = 0,553$ менше за теоретичне $t_{(0,05;7)} = 2,365$, тому гіпотеза про рівність нулю математичного сподівання залишків приймається.

4. Для дослідження випадковості коливань залишкових рівнів часового ряду застосуємо критерій піків (поворотних точок) [11, с. 428].

Крок 1. У послідовності залишків обираються поворотні точки (табл. 6).

Таблиця 6. Визначення піків або поворотних точок

t	y_t	y_{tm}	u_t	П
1	34779,2	31574,0	3205,2	–
2	29208,1	31349,8	–2141,7	1
3	33105,5	32237,0	868,5	0
4	35051,0	33956,6	1094,4	1
5	36996,5	36229,4	767,1	0
6	37454,9	38776,0	–1321,1	1
7	40917,0	41317,4	–400,4	0
8	43678,6	43574,3	104,3	0
9	46440,1	45267,6	1172,5	1
10	45598,3	46118,0	–519,7	–
Разом				4

Джерело. Розраховано авторами.

Крок 2. Розраховується загальна кількість піків (поворотних точок), яка дорівнює $\Pi = 4$.

Крок 3. Обчислюється математичне сподівання поворотних точок Π^* та їх дисперсія δ_n^2 за умови випадкової вибірки залишків:

$$\Pi^* = (2/3)(n - 2) = (2/3)(10 - 2) = 5,333$$

та

$$\delta_n^2 = \frac{16n - 29}{90} = \frac{16 \cdot 10 - 29}{90} = 1,456.$$

Крок 4. Критерієм випадковості залишків з 5%-м рівнем значущості має бути виконання нерівності

$$\Pi = 5 > [\Pi^* - 1,96\sqrt{\delta n^2}] = [5,333 - 1,96\sqrt{1,456}] = 2,969,$$

де квадратні дужки означають цілу частину числа. Оскільки $\Pi = 4 > [\Pi^* - 1,96\sqrt{\delta n^2}] = 2$, то випадковість коливань залишків з імовірністю 95% підтверджується.

Таким чином, побудована модель тренду (1) адекватна.

Точковий прогноз Y_{n+L} визначається показником, коли в рівняння моделі тренду підставляється значення часу $t = n + L$, що відповідає періоду упередження або прогнозованому періоду $t = n + 1 = 10 + 1 = 11$ (2020 р.) або

$t = n + 2 = 10 + 2 = 12$ (2021 р.) і становить:

- для 2020 р.

$$y_{11} = 33189 - 2103,4 \cdot 11 + 834,96 \cdot 11^2 - 46,533 \cdot 11^3 = 45846,3 \text{ тис. ц.}$$

- для 2021 р.

$$y_{12} = 33189 - 2103,4 \cdot 12 + 834,96 \cdot 12^2 - 46,533 \cdot 12^3 = 44173,4 \text{ тис. ц.}$$

Інтервальний прогноз для тренду, що має вид полінома третього порядку, розраховується за формулою [11, с. 437]

$$U_y = Y_{n+L} \mp t_\alpha S_y \sqrt{1 + \frac{1}{n} + \frac{t_L}{\sum t^2} + \frac{\sum t^4 - 2t_L^2 \sum t^2 + nt_L^4}{n \sum t^4 - (\sum t^2)^2}},$$

де Y_{n+L} — точковий прогноз за моделлю на $(n + L)$ -й період часу; t_α — теоретичне значення критерію Стьюдента для рівня значущості α ;

$$S_y = \sqrt{\frac{\sum u_t^2}{n - m}} = \sqrt{\frac{20961637,6}{10 - 3}} = 1730,5;$$

S_y — середнє квадратичне відхилення; t — порядковий номер рівня ряду ($t = 1, 2, \dots, n$); t_L — відповідає $(n + L)$ -му періоду часу, для якого виконується прогноз; N — довжина часового ряду; m — кількість параметрів моделі тренду.

Інтервальний прогноз виробництва зернових і зернобобових культур сільськогосподарськими підприємствами в Черкаській області становить:

- для 2020 р.

$$U_{11} = 45846,3 \mp 2,365 \cdot 1730,5 \times$$

$$\times \sqrt{1 + \frac{1}{10} + \frac{11}{385} + \frac{25333 - 2 \cdot 121 \cdot 385 + 10 \cdot 14641}{10 \cdot 25333 - 148225}} = 45846,3 \mp 5604,8;$$

- для 2021 р.

$$U_{12} = 44173,4 \mp 2,365 \cdot 1730,5 \times$$

$$\times \sqrt{1 + \frac{1}{10} + \frac{12}{385} + \frac{25333 - 2 \cdot 144 \cdot 385 + 10 \cdot 20736}{10 \cdot 25333 - 148225}} = 44173,4 \mp 6192,4.$$

Показники точності моделі тренда розраховуються за формулами [11, с. 430]:

- 1) середнє квадратичне відхилення

$$S_y = \sqrt{\frac{\sum u_t^2}{n - m}} = 1730,5;$$

- 2) середня відносна помилка апроксимації

$$\varepsilon_t = \frac{\frac{1}{n} \sum |u_t|}{y_t} \cdot 100 = \frac{32,77}{10} \cdot 100 = 3,3,$$

яка не перевищує 5%;

3) коефіцієнт збіжності

$$\varphi^2 = \frac{\sum_{t=1}^n u_t^2}{\sum_{t=1}^n (y_t - y_{cp})^2} = \frac{20961637,6}{290310000} = 0,07,$$

4) коефіцієнт детермінації

$$R^2 = 1 - \varphi^2 = 1 - 0,07 = 0,93.$$

Розраховані показники свідчать про високу точність прогнозу, оскільки середня відносна помилка дорівнює 3,3%, коефіцієнт детермінації вказує, що отримані спостереження на 93% підтверджують модель.

ВИСНОВКИ

Для дослідження прогнозованих значень показників виробництва зернових і зернобобових культур у сільськогосподарських підприємствах Черкаської області використано одну з методик інтелектуального аналізу даних: аналіз часових рядів для того, щоб на основі моделі ряду динаміки, побудованої на базі фактичних значень попередніх показників минулих років, виконати прогноз виробництва на наступні роки. Застосована методика включає в себе такі етапи: графічний аналіз (дає змогу підібрати рівняння моделі), виділення і аналіз детермінованих складових ряду, згладжування і фільтрація часового ряду, дослідження випадкової компоненти, побудова і перевірка на адекватність моделі часового ряду, прогнозування поведінки часового ряду на підставі результатів досліджень. Практична апробація моделі дозволила визначити, що зі збереженням тенденцій за попередні роки прогнозне значення обсягу виробництва зернових і зернобобових культур у Черкаській області на 2021 р. буде коливатися від 4091,4 до 5551,8 тис. ц, а на 2021 р. – від 4587,7,4 до 6317,9 тис. ц.

ЛІТЕРАТУРА

1. Р.О. Петров та О.Я. Кучерук, *Прогнозування термінів продажу товарів методами інтелектуального аналізу даних. Актуальні проблеми комп'ютерних наук*. 2019. Доступно: http://elar.khnu.km.ua/jspui/bitstream/123456789/7933/1/APKN-2019_%28v_2_0%29-156-158.pdf (дата звернення 20.10.2020).
2. П.І. Бідюк, С.М. Савченко, та А.С. Савченко, *Методи інтелектуального аналізу даних в прогнозуванні конкурентоспроможності підприємств*. Доступно: <http://www.ei-journal.in.ua/index.php/journal/article/view/61/48> (дата звернення 20.10.2020).
3. Г. Черноус та С. Рибальченко, “Оптимізація ціноутворення на основі моделей інтелектуального аналізу даних”, *Вісник Київського національного університету імені Тараса Шевченка*, №7 (172), с. 52–58, 2015.
4. П.І. Бідюк та А.В. Федоров, “Ймовірнісне прогнозування процесів ціноутворення на фондових ринках”, *Системні дослідження та інформаційні технології*, № 1, с. 65–73, 2009.
5. Л.О. Коршевнюк та Бідюк П.І., “Інформаційно-аналітична система для адаптивного прогнозування фінансових процесів та оцінювання ризиків”, *Наукові праці. Комп'ютерні технології*, вип. 201, т. 213, с. 59–62, 2013.
6. О.Ю. Берзлев, М.М. Маляр, та В.В. Ніколенко, “Адаптивні комбіновані моделі прогнозування біржових показників”, *Вісник Черкаського держ. технолог. ун-ту. Серія: технічні науки*, № 1, с. 50–54, 2011.

7. А.Ю. Берзлев, “Оценка эффективности прогнозирования и принятия решений на финансовом рынке”, *Problems of Computer Intellectualization*, V.M. Glushkov Institute of Cybernetics of NAS of Ukraine. Kyiv-Sofia: ITHEA, 2012, с. 249–257.
8. В.Д. Кишенько та М.А. Сич, “Інтелектуальний аналіз в задачах прогнозування тенденцій розвитку технологічного комплексу цукрового заводу”, *Цукор України*, № 6–7 (126–127), с. 36–40, 2016.
9. І.І. Глаголева та А.Ю. Берко, *Застосування методів інтелектуального аналізу даних для прогнозування використання земельних ресурсів*. Доступно: file:///C:/Users/user/Downloads/VNULPICM_2013_770_20.pdf (дата звернення 20.10.2020).
10. О.І. Черняк та П.В. Захарченко, *Інтелектуальний аналіз даних: підручник*. Київ, 2014, 599 с.
11. С.І. Наконечний, Т.О. Терещенко, та Т.П. Романюк, *Економетрія: підручник*, 2-ге вид., допов. та перероб. Київ: КНЕУ, 2000, 296 с.
12. О.В. Ульянченко, *Дослідження операцій в економіці : підручник для студентів вузів*. Харків: Гриф, 2002, 580 с.
13. К. Холден, Д.А. Піл, та Дж. Л. Томпсон, *Економічне прогнозування. Вступ*. Київ: Інформтехніка, 1996, 216 с.
14. С. Chatfield, *The Analysis of Time Series: An Introduction*. London, 1996, 283 p.

Надійшла 27.10.2020

INFORMATION ON THE ARTICLE

Serhii M. Kontseba, ORCID: 0000-0003-4161-5581, Uman National University of Horticulture, Ukraine, e-mail: kontseba@meta.ua

Roman I. Lishchuk, ORCID: 0000-0002-2051-5365, Uman National University of Horticulture, Ukraine, e-mail: roma0lir@gmail.com

Svitlana D. Skurtol, ORCID: 0000-0003-4517-2466, Uman National University of Horticulture, Ukraine, e-mail: skurtol@i.ua

Halyna Yu. Rodashchuk, ORCID: 0000-0002-0336-3435, Uman National University of Horticulture, Ukraine, e-mail: galya.rodashchuk@gmail.com

Ivan P. Vasylychenko, ORCID: 0000-0002-9341-539X, Uman National University of Horticulture, Ukraine, e-mail: ipvasylchenko@gmail.com

ПРОГНОЗИРОВАНИЕ ОБЪЕМОВ ПРОИЗВОДСТВА СЕЛЬСКОХОЗЯЙСТВЕННОЙ ПРОДУКЦИИ С ПОМОЩЬЮ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ / С.М. Концеба, Р.И. Лишук, С.Д. Скуртол, Г.Ю. Родащук, И.П. Васильченко

Аннотация. Спрогнозированы будущие значения показателей производства зерновых и зернобобовых культур в сельскохозяйственных предприятиях Черкасской области на основе временных рядов динамики, уровни которых выражены в натуральных единицах измерения. При исследовании использовалась одна из методик интеллектуального анализа данных — анализ временных рядов, что позволяет на основе модели ряда динамики, построенной на базе фактических значений показателей прошлых лет, спрогнозировать производство на будущие годы. Методика содержит следующие этапы: графический анализ (для подбора подходящего уравнения модели), выделение и анализ детерминированных составляющих ряда, сглаживание и фильтрация временного ряда, исследования случайной компоненты, построение и проверка на адекватность модели временного ряда, прогнозирование поведения временного ряда на основе проведенных исследований.

Ключевые слова: интеллектуальный анализ данных, временной ряд, прогнозирование, тренд, сельскохозяйственное предприятие, модель.

FORECASTING OF AGRICULTURAL PRODUCTION VOLUMES USING METHODS OF DATA MINING / S.M. Kontseba, R.I. Lishchuk, S.D. Skurtol, H.Yu. Rodashchuk, I.P. Vasylenko

Abstract. In this article, the future values of indicators were forecasted for production of grains and legumes on farms in Cherkasy region based on the time series expressed in physical units. Time series analysis as one of the data mining techniques was used during the research in order to make a forecast of production using the data (based on the model of dynamic series) from past years to predict the future production volumes. This method contains the following steps: a graphical analysis (allows you to choose the model equation in the best way), separation and analysis of deterministic components of the series, smoothing and filtering of time series, study of random components, construction and testing for the adequacy of the time series model, forecasting the behavior of the time series based on the conducted research.

Keywords: data mining, time series, forecasting, trend, agricultural enterprise, model.

REFERENCES

1. R.O. Petrov and O.Y. Kucheruk, "Forecasting the terms of sale of goods by the methods of intellectual data analysis", *Current problems of computer science*, pp. 156–158, 2019. Accessed on: October 20, 2020. [Online]. Available: http://elar.khnu.km.ua/jspui/bitstream/123456789/7933/1/APKN-2019_%28v_2_0%29-156-158.pdf
2. P. Bidyuk, A. Savchenko, and S. Savchenko, *Methods of intellectual data analysis in forecasting competitiveness of enterprises*. [Online]. Available: <http://www.ei-journal.in.ua/index.php/journal/article/view/61/48> [Accessed on: October 20, 2020].
3. H. Chornous and S. Rybalchenko, "Pricing optimization based on data mining models", *Bulletin of Taras Shevchenko National University of Kyiv*, no. 7 (172), pp. 52–58, 2015.
4. P. Bidyuk and A. Fedorov, "Probabilistic forecasting of pricing processes in stock markets", *System Research and Information Technologies*, no. 1, pp. 65–73, 2009.
5. L. Korshevnyuk and P. Bidyuk, "Information-analytical system for adaptive forecasting of financial processes and risk estimation", *Scientific works. Computer Technology*, vol. 201, no. 213, pp. 59–62, 2013.
6. O. Berzlev, M. Malyar, and V. Nikolenko, "Adaptive combined models of forecasting stock market indicators", *Bulletin of Cherkasy State Technological University*, pp. 50–54, 2011.
7. A. Berzlev, "Assessment of the effectiveness of forecasting and decision-making in the financial market", *Problems of Computer Intellectualization, V.M. Glushkov Institute of Cybernetics of NAS of Ukraine*. Kyiv-Sofia: ITHEA, pp. 249–257, 2012.
8. V. Kyshenko and M. Sych, "Intellectual analysis in the tasks of forecasting the trends in the technological complex of the sugar factory development", *Sugar of Ukraine*, no. 6-7 (126-127), pp. 36–40, 2016.
9. I. Hlaholeva and A. Berko, *Application of data mining methods for forecasting land use*. [Online]. Available: <http://science.lpnu.ua/sisn/all-volumes-and-issues/volume-770-2013/zastosuvannya-zasobiv-intelektualnogo-analizu-danih-dlya> [Accessed on: October 20, 2020].
10. O. Chernyak and P. Zakharchenko, *Data mining*. Kyiv, 2014, 599 p.
11. S. Nakonechnyy, T. Tereshchenko, and T. Romanyuk, *Econometric*, 2nd edition. Kyiv National Economic University, 2000, 296 p.
12. O. Ulyanchenko, *Research of operations in economics: a textbook for university students*. Kharkiv: Gryf, 2002, 580 p.
13. K. Holden, D.A. Peel, and J.L. Thompson, *Economic forecasting: an introduction*. Kyiv: Information technology, 1996, 216 p.
14. C. Chatfield, *The Analysis of Time Series: An Introduction*. London, 1996, 283p.

REVIEW METHODS FOR BREAST CANCER DETECTION USING ARTIFICIAL INTELLIGENCE AND DEEP LEARNING METHODS

M. NADERAN

Abstract. Nowadays, there are many related works and methods that use Neural Networks to detect the breast cancer. However, usually they do not take into account the training time and the result of False Negative (FN) while training the model. The main idea of this paper is to compare already existing methods for detecting the breast cancer using Deep Learning Algorithms. Moreover, since the breast cancer is one of the most common lethal cancers and early detection helps prevent complications, we propose a new approach and the use of the convolutional autoencoder. This proposed model has shown high performance with sensitivity, precision, and accuracy of 93,50%, 91,60% and 93% respectively.

Keywords: convolutional autoencoder, breast cancer detection, machine learning algorithms, convolutional neural networks, medical image classification.

INTRODUCTION

Breast Cancer is the second most deadly cancer mostly in women in age 20–59 worldwide. There are two classes for Breast Cancer: benign and malignant. The benign tumor cells are non–cancerous cells and grow only locally and cannot spread by invasion. Where as malignant tumor is cancerous, and the cells can grow and spread to other parts of the body. However, early detection of breast cancer could prevent cancer growth. For this reason, it is essential to train neural networks with more accurate performance and with low false negative results. According to [1], one in every eight women in the United States of America has breast cancer over their lifetime. Due to early detection and increased breast cancer awareness, there were an estimated 375,900 fewer breast cancer deaths in 2017.

REVIEW OF PREVIOUS WORKS

Authors in [2] proposed modified deep learning model where they compared two neural networks VGG and Resnet 50. In addition, they demonstrated how the number of epochs could change the performance of the model. In their experiment, the accuracy of Resnet 50 was 97% and VGG 84% with number of epoch 198 and 25, respectively. The advantage of this work is that authors worked with a whole image, which could combine two steps of preprocessing data. The first step is sliding images to recognize local patches for generating a grid of probabilistic outputs. Second step is processing to summarize the patch classifier's outputs. The limitation of this work is that authors did not include sensitivity of the model.

In [3] authors compared different methods for pre-processing data. According to the experiments, using a resample filter on input data could increase the performance of the model. Besides that, authors compared three different classifier Decision Tree (J48), NB and Sequential Minimal Optimization (SMO) on two different datasets and classifier J48 has shown the best result.

Selection of the most informative features as a classifier input is an essential task. Authors in [4] used Principle Component Analysis (PCA) method to decrease the dimensions of the feature map that will be sent to the classifier. However, there are some limitations in using PCA as a feature reduction. First, PCA is a linear relationship between features. Second, it is sensitive to the scale of the features. Third, it assumes a correlation between features and if features are not correlated, then PCA is unable to complete the principal components. To solve this problem, the Convolutional Autoencoder can be used for reducing the dimensionality of feature maps.

Dalal Bardou and other authors in [5] proposed CNN model where they compared the result of the accuracy for different magnification factors 40X, 100X, 200X and 400X. They demonstrated the result of the accuracy on 40X zoom that shows better result comparing other magnification zoom where it was 94,65%, 94,07%, 94,54% and 93,77% for 40X, 100X, 200X and 400X, respectively. Moreover, they demonstrate how pre-processing data is an essential task that must be considered while training convolutional neural networks.

Dataset BIRADS was used in experiments [6], where Support Vector Machine (SVM) method was considered for classifying breast cancer. The BIRADS dataset contain 516 benign and 445 malignant cases. During experiments authors improved the SVM model to increase false positive rate to only 13%. Moreover, SVM was compared with other methods like Decision tree, Naïve bayes, neural network and auto multilayer perceptron (MLP). Comparison result showed that sensitivity of decision tree is higher than other methods and it is 86,44%. However, the accuracy for decision tree was 78,79%. As a result, it could be concluded that the low result of accuracy does not mean a low result of sensitivity. Thus, for sensitive tasks like breast cancer detection, it is essential to consider both sensitivity and precision.

Even though SVM shows high performance [7], it cannot be used directly on images unless some pre-processing were applied on the images to extracting informative features. For example, in [8] authors used VGG as a feature extraction and applied SVM on the output of the VGG19 for classification.

Authors in [9] consider two essential tasks for training CNN: feature extraction, and fine tuning. During the experiment, CNN VGG-16 was used for feature extraction and ANFIS used for classification as well as SVM for compressing reason. Result of the experiment shows the combination of the CNN VGG-16 and ANFIST demonstrate high performance with precision 88,39%. Whereas precision of the architecture VGG + SVM was 87,10%.

DATASET

The BreakHis dataset was used during the experiments. The dataset includes two classes Benign and Malignant. There are 2,480 cases belonging to the benign category and 5,429 scans belonging to the malignant category. There were 5,536

training scans, 709 validation scans, and 1,581 test scans. Experiments showed even though the dataset is imbalanced, it will not affect the performance of the model.

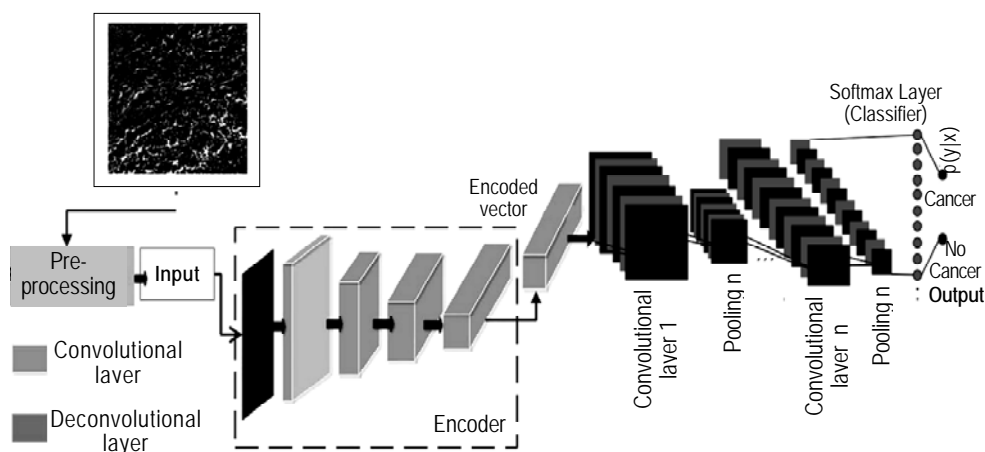
PRE-PROCESSING

There were a few steps that were considered before using the data. First, on the input data normalization was applied. Then, all data were resized using lossless compression from 700 by 460 pixels to 176 by 116 pixels. The use of lossless compression was used to reduce the input size of the data without destroying any critical information and while also preserving the aspect ratio thus maintaining the integrity of the data.

METHOD

In the current paper, a new architecture of convolutional autoencoder was proposed, where sensitivity of the model was 93,50%. Moreover, the training time was considerably lower than in the reviewed existing models.

Increasing performance of the model, required extracting the most important features. The proposed hybrid convolutional network consists of two models of convolutional autoencoder as a selection of informative features and convolutional network as a classifier. Only after achieving the most essential features, the fully connected layer was applied on the output of the convolutional layers for classifying input images as benign or malignant. The figure below demonstrates the architecture of the hybrid convolutional neural network.



Architecture of the proposed hybrid convolutional neural network

CONCLUSION

In this paper several works by various authors were reviewed. These works focused on the topic of image classification in the classification of cancer. These works tend to rely on accuracy, while sometimes neglecting the importance of sensitivity. The reviewed papers contain informative topics and ideas such as the authors in [3] use of resampling of input data as a pre-processing technique to

improve the results of the J48 network method. The analysis of accuracy in regard to different zoom levels, of the images in the dataset was proposed by authors in [5]. Authors in [8] show the use of feature extraction prior to classification using the SVM method.

According to the research and comparing related works, some of them used methods that showed high performance. However, those models are complex, and the training time is considerably high. In this paper, the proposed model showed higher performance with less training time. Faster training time is because as feature extraction, modified convolution autoencoder was used. Moreover, extraction of only informative features helped to increase the performance of the model. The sensitivity of the model has shown 93,50%.

REFERENCES

1. *Breast Cancer Facts*. Available: <https://www.nationalbreastcancer.org/breast-cancer-facts>
2. L. Shen et al., “Deep Learning to Improve Breast Cancer Detection on Screening Mammography”, *Sci Rep*, 9, 12495, 2019. Available: <https://doi.org/10.1038/s41598-019-48995-4>
3. S.A. Mohammed, S. Darrab, S.A. Noaman, and G. Saake, “Analysis of Breast Cancer Detection Using Different Machine Learning Techniques”, in *Data Mining and Big Data. DMBD 2020. Communications in Computer and Information Science*, vol. 1234, 2020. Springer, Singapore. Available: https://doi.org/10.1007/978-981-15-7205-0_10
4. Habib Dhahri, Eslam Al Maghayreh, Awais Mahmood, Wail Elkilani, and Mohammed Faisal Nagi, “Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms”, *Journal of Healthcare Engineering*, vol. 2019, article ID 4253641, 11 pages, 2019. Available: <https://doi.org/10.1155/2019/4253641>.
5. Dalal Bardou, Kun Zhang and Sayed Mohammad Ahmad, “Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks”, *2018 IEEE*, vol. 6, 2018.
6. Sandeep Chaurasia, Prasun Chakrabarti, and Neha Chourasia, “Prediction of Breast Cancer Biopsy Outcomes – An Approach using Machine Learning Perspectives”, *International Journal of Computer Applications*, vol. 100, no.9, 2014.
7. Md. Milon Islam, Hasib Iqbal, Md. Rezwanaul Haque, and Md. Kamrul Hasan, “Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors”, *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, 21–23 Dec 2017, Dhaka, Bangladesh.
8. Md. Kamrul Hasan and Tajwar Abrar Aleef, “Automatic Mass Detection in Breast Using Deep Convolutional Neural Network and SVM Classifier”, *arXiv preprint arXiv:1907.04424v1 [cs.CV]*, 9 Jul 2019.
9. Yuriy P. Zaychenko, Kostiantyn A. Zdor, and Galib Hamidov, “Brain tumor diagnostics with application of hybrid fuzzy convolutional neural networks”, *System Research and Information Technologies*, no. 1, 2020. Available: <https://doi.org/10.20535/SRIT.2308-8893.2020.1.06>.

Received 26.01.2021

INFORMATION ON THE ARTICLE

Maryam Naderan, Educational and Scientific Complex “Institute for Applied System Analysis” of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: ma.naderan@gmail.com

ОГЛЯД МЕТОДІВ ВИЯВЛЕННЯ РАКУ МОЛОЧНОЇ ЗАЛОЗИ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ І МЕТОДІВ ПОГЛИБЛЕНОГО НАВЧАННЯ / М. Надеран

Анотація. Існує безліч схожих праць і методів, у яких використовуються нейронні мережі для виявлення раку молочної залози. Утім зазвичай у них не враховуються час навчання і результати помилкового негативного FN під час навчання моделі. Основною ідеєю цієї роботи є порівняння вже наявних методів виявлення раку молочної залози з використанням глибоких алгоритмів навчання. Оскільки рак молочної залози є одним з найбільш поширених смертельних видів раку, а його раннє виявлення запобігає ускладненням, запропоновано новий підхід і використання згорткового автоенкодера. Запропонована модель показала високу продуктивність із чутливістю (recall), точністю (precision) і точністю (accuracy) 93,50%, 91,60% і 93% відповідно.

Ключові слова: згортковий автоенкодер, розпізнавання раку молочної залози, алгоритми машинного навчання, згорткові нейронні мережі, класифікація медичних зображень.

ОБЗОР МЕТОДОВ ВЫЯВЛЕНИЯ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ / М. Надеран

Аннотация. Существует множество похожих работ и методов, в которых используются нейронные сети для обнаружения рака молочной железы. Однако обычно в них не учитываются время обучения и результаты ложного отрицательного FN при обучении модели. Основной идеей этой работы является сравнение уже существующих методов обнаружения рака молочного железы с использованием глубоких алгоритмов обучения. Поскольку рак молочной железы является одним из наиболее распространенных смертельных видов рака, а его раннее обнаружение предотвращает осложнения, предложен новый подход и использование сверточного автоэнкодера. Предложенная модель показала высокую производительность с чувствительностью (recall), точностью (precision) и точностью (accuracy) 93,50%, 91,60% и 93% соответственно.

Ключевые слова: сверточный автоэнкодер, распознавание рака молочной железы, алгоритмы машинного обучения, сверточные нейронные сети, классификация медицинских изображений.

**СЦЕНАРНОЕ КОГНИТИВНОЕ МОДЕЛИРОВАНИЕ
ТЕНДЕНЦИЙ РАЗВИТИЯ СЛОЖНОЙ СИСТЕМЫ
«МОЛОДЕЖЬ, РЫНОК ТРУДА, КАЧЕСТВО ЖИЗНИ»**

Г.В. ГОРЕЛОВА, А.А. СААК

Аннотация. Проведение исследований качества жизни молодежи и влияния на нее такого существенного фактора, как молодежный рынок труда, а также связанных с этими исследованиями проблем государственной поддержки молодежи, требуют использования методов имитационного моделирования в связи с неполнотой информации и слабоструктурированностью проблем таких сложных систем, как качество жизни молодежи и рынок труда. Дан краткий обзор категорий «рынок труда», «качество жизни», «молодежь», необходимый для определения содержания когнитивного моделирования сложной системы «молодежь, рынок труда, качество жизни». Кратко представлена технология когнитивного моделирования. Приведены результаты применения когнитивного имитационного моделирования к изучению качества жизни молодежи во взаимосвязи с рынком труда. Пример моделирования подан в виде соответствующей когнитивной карты и ряда сценариев возможного развития ситуаций на ней, которые могут возникнуть под предполагаемым воздействием различных социальных и экономических факторов, а также под воздействием регулирующих и управляющих систем.

Ключевые слова: молодежь, рынок труда, имитационное моделирование, когнитивная карта, сценарий, возможность, ситуация.

ВВЕДЕНИЕ

В настоящее время повышение качества жизни населения во многих странах мира декларируется как важнейшее условие их устойчивого развития и одно из главных направлений социально-экономической политики государства. Качество жизни населения, молодежь, рынок труда давно являются объектом исследования многих ученых, например [1, 3, 9, 10, 12, 13, 17, 20, 25–28]. Существует большое количество подходов к изучению этих категорий, как и методов исследования состояния качества жизни молодежи, молодежного рынка труда. В данной работе использован когнитивный подход и когнитивное имитационное моделирование для исследования сложной социально-экономической системы, названной нами «молодежь, рынок труда, качество жизни». Такой подход к исследованию был обусловлен как свойствами социально-экономических систем, имеющих сложную структуру, функциони-

рующих в условиях разного рода неопределенности, и сложностью получения точной количественной информации о них, так и возможностями когнитивного моделирования [2, 3, 5–8, 11, 14, 16, 18, 19, 21–24], позволяющего имитировать структуру и поведение сложной системы.

На основании существующих теоретических и практических знаний о системе при когнитивном моделировании сложных систем строится когнитивная модель системы, с помощью которой возможно имитировать процессы распространения возмущений в системе под предполагаемом воздействии различных социально-экономических факторов и анализировать возникающие при этом сценарии развития ситуаций в системе [14,16,18,23]. При этом реализуется одно из важных достоинств когнитивного моделирования – порождение таким образом новых знаний о системе на основании исходных не только количественных, но и качественных данных, а также научное предвидение вариантов возможных путей развития сложной системы.

ПОДХОДЫ К ПОНИМАНИЮ И ИЗУЧЕНИЮ ПОНЯТИЙ «КАЧЕСТВО ЖИЗНИ», «РЫНОК ТРУДА», «МОЛОДЕЖЬ»

Обзор и анализ работ, относящихся к изучению категорий качества жизни, молодежи, молодежного рынка труда, в том числе [4, 9, 10, 12, 13, 15, 17, 20, 25–28], позволили выделить основные понятия и элементы системы «молодежь, рынок труда, качество жизни», которые необходимо учитывать при ее моделировании. Рассмотрим кратко основные широко известные теоретические сведения по каждому из включенных в исследование понятий, что необходимо для обоснования разрабатываемой когнитивной модели «молодежь, рынок труда, качество жизни».

Рынок труда. Рынку труда, являющегося одной из основ различных экономических теорий, посвящено огромное количество работ. В изучении рынка труда можно выделить имена представителей *кейнсианской школы* (таких, как ее основатель Дж. Кейнс, кейнсианцы А. Лейонхуфвуд, Дж. Робинсон, П. Сраффа, Э. Хансен, Д. Хикс, Й. Шумпетер и др.), *монетаристской школы* (Д. Юм, М. Фридман, Э.Фелпс, Ф.Хайек и др.), *институциональной социологической школы*, противопоставляющей себя классической школе как с точки зрения предмета, так и метода (Т. Веблен, М. Вебер, Дж. Гелбрайт, Дж. Данлоп, У. Гамильтон, Коммонс, У. Митчелл, Л. Ульман Фурастье и др.). В современном постиндустриальном обществе существует *школа нового институционализма* (Д.Белл, Т. Веблен, Дж. Гелбрейт, К. Кларк, П. Друкер, Э.Тоффлер, Ж. Фукрастье, М. Кастельс, А. Турен и др.). Среди имен украинских и российских современных исследователей рынка труда можно обозначить имена А.С.Булатова, И.Ю. Бочарова, Л. Башука, В.С. Буланова, Н.А. Волгина, Б.Г. Збышко, В.В. Ишина, С.А. Карташова, А.В. Кашепова, Е.И. Колотюк, Т.С. Калашниковой, Л.А. Костина, А.З. Котляра, Л.Г. Матвеевой, Е.В. Михалкиной, Т.А. Макареши, Б.Ч. Муртозаевой, А.А. Никифоровой, А.Ю. Никитаевой, Ю.Г. Одегова, В.С. Осанкина, В.А. Павленкова, В.О. Романшина, А.И. Рофе, Г.Г. Руденко, Л.А. Рязанвой, Э.Р. Саруханова, Е.Л. Спиридоновой, О.А.Черновой и др.

Основной *идеей кейнсианства* является необходимость государственного монополистического регулирования при развитии капиталистической

экономики на базе анализа макроэкономических показателей; один только рыночный механизм не в состоянии решать многообразные экономические проблемы; только активная финансовая политика государства в состоянии регулировать спрос и справиться с массовой безработицей. Вынужденная безработица, гибкая политика заработной платы, инвестиции оптимальных размеров, результативная организация общественных работ — косвенные методы вмешательства в рынок.

Как известно, кейнсианский подход к экономическому регулированию — это стимулирование спроса через достижение полной занятости путем государственного манипулирования налогами и расходами достаточно актуален в наше время.

Основная идея монетаристского подхода состоит в том, что характер экономических отношений определяет всю систему экономических институтов (фирмы, профсоюзы, государство, законы, указы, постановления, обычаи, навыки, традиции и т.д.), а не рынок сам по себе. Рыночная экономика внутренне устойчива, но ее неустойчивость может возникнуть из-за вмешательства государства, скорость обращения денег относительно постоянна, прямое регулирование денежной массы стимулирует экономику, инфляция — денежный феномен, зависит от роста денежной массы, деньги решают всё. Представители монетаристской школы исследовали взаимосвязь безработицы с динамикой реальной заработной платы, с инфляцией. Они исходят из жесткой структуры цен на рабочую силу (как и кейнсианцы) из предпосылки их однонаправленного, повышательного движения. Для рынка труда негативными факторами, усиливающими рыночное неравновесие, являются установление государством минимального уровня заработной платы, сильные позиции профсоюзов, отсутствие всей необходимой информации о наличии вакансий и резервной рабочей силы. Для уравнивания рынка предлагают использовать инструменты денежно-кредитной политики. Денежная политика должна быть направлена на достижение естественного уровня безработицы, отражающего структурные диспропорции на рынке труда и не связанного с циклической конъюнктурой в экономике.

Анализируя работы современных ученых и суммируя совокупность существующих определений рынка труда и его элементов, можно считать, что признание наличия системы социально-экономических отношений между субъектами рынка труда является одним из важнейших, определяющих сущность этого рынка. И именно социально-экономические отношения являются важным фактором качества жизни населения. Исходя из этого на данном этапе исследования нами принята следующая трактовка рынка труда: *рынок труда — это, прежде всего, система социально-экономических отношений, связанных с наймом и предложением труда (с его куплей и продажей); это также экономическое и географическое пространство — сфера трудоустройства, в которой взаимодействуют покупатели и продавцы специфического товара — труда; наконец, это механизм, обеспечивающий согласование цены и условий труда между работодателями и наемными работниками.* Под *региональным рынком труда* будем понимать сложную систему обмена индивидуальных способностей к труду на фонд жизненных средств с учётом локальной специфики региона.

Качество жизни. В условиях быстрореформируемого современного мира исследование социальных и экономических проблем качества человеческой жизни является чрезвычайно актуальным. Категория «качество жизни» — это обобщающая социально-экономическая характеристика жизни людей в любом обществе, в любой исторический период, раскрывающая не только жизнедеятельность, жизнеобеспечение, но и жизнеспособность общества как целостного социального организма. Качество жизни — основная характеристика благосостояния народа. Повышение благосостояния народа является важнейшей задачей экономической и социальной политики любого государства в современном мире и при оценке развитости стран качество жизни его населения становится всё более значимой категорией.

Анализ генезиса категории «качество жизни» показал, что существует несколько подходов к ее изучению в зависимости от научной дисциплины, которая его использует. Можно выделить основные подходы: философский, социально-экономический, экологический и альтернативной цивилизации.

Философский подход можно рассматривать, начиная с древнегреческого периода, далее — с эпохи просвещения, XIX и XX ст.; развитие представлений о качестве жизни в рамках философского подхода характеризуется большим разнообразием трактовок самого понятия, различным пониманием сущности высокого качества жизни, которое отождествлялось с духовной свободой человека, его моральным совершенством, с состоянием счастья, с добродетелью и др.

Социально-экономический подход начал развиваться в период первой промышленной революции (П. Буагильбер, А. Тюрго и др., XIX ст., И. Бентам, А. Смит, Д. Рикардо, К. Маркс и др., XX ст.). Существуют оптимистический и пессимистический разновидности социально-экономического подхода. Представители оптимистической позиции: Р. Арон, Д. Белла, Дж. Гэлбрейт, П. Дракер, Р. Инглегарт, А. Пигу, Ж. Фурастье и др., российские исследователи В.Г. Афанасьев, Б.Д. Парыгин, И.В. Блауберг, Т.А. Ильина, Г. Кунц, О.Е. Лебедев, Л.Г. Логинова, В.Н. Садовский, Г.Н. Сериков и др. К представителям пессимистической позиции относятся Т. Адорно, Г. Маркузе, Э. Мишан, Л. Мэмфорд, Ф. Розак, Б. Скинер, Ю. Хамбермас, члены Римского клуба, Э. Фромм, К. Ясперс и др.

Термин «качество жизни» впервые был введен Доном Гелбрейтом («Общество изобилия», 1958 г.). Был осознан факт связи между уровнем экономического развития страны и удовлетворенностью жизнью. Что важно, в это время пришло понимание того, *что экономические, культурные и политические преобразования причинно связаны в своем развитии, причем эти связи имеют взаимный характер. Поэтому, зная один компонент общества, можно прогнозировать взаимосвязь в развитии других компонентов; эта взаимосвязь позволяет прогнозировать характер их воздействия на общество.*

Такое понимание существенно при разработке когнитивной модели причинно-следственных связей сложной системы «молодежь, рынок труда, качество жизни».

Экологический подход начал развиваться с середины XX ст.; его представители, зарубежные и российские: У. Бек, Дж. Гэлбрейт, Д.М. Гвишиани,

Э. Гиденс, Л.Г. Гусякова, С.И. Григорьев, В.И. Данилов-Данильян, Г. Дилигенский, В.И. Жуков, Н.Н. Моисеев, Д. Маркович, Р. Супек, А.И. Субетто, Э. Тоффлер и др.

В рамках экологической парадигмы существует концепция *альтернативной цивилизации* (Л. Барун, И.В. Бестужев-Лада, Г. Гендерсон, Г.Г. Дилигенский, В.А. Коптюг, В.М. Матросов, Ж. Робен и др.). Сторонники альтернативистики предлагают самые разные пути улучшения качества жизни. В конце XX ст. была выдвинута концепция «sustainable development» — устойчивого развития цивилизации (позиция российских ученых представлена в энциклопедической монографии [21]) и предложены цели устойчивого развития. Этих целей будем придерживаться в исследовании.

Итак, для проводимого исследования важно констатировать следующее: экономические, культурные и политические преобразования в обществе тесно связаны в своем развитии; экономика определяет культуру и политику, и наоборот, экологические проблемы начинают детерминировать экономику, политику, культуру; причинно-следственные связи между всеми обозначенными компонентами имеют взаимный характер. Поэтому, зная один компонент общества, можно предвидеть взаимосвязь других компонентов, эта взаимосвязь позволяет прогнозировать характер их воздействия на общество.

Молодежь. Теоретическое осмысление категории «молодежь» может происходить в разных контекстах. Молодежь как социально-демографическая группа и как особая составная часть современного общества исследуется различными социально-гуманитарными науками, такими, как демография, политология, философия, педагогика, психология, социология и др. В настоящее время существуют подходы к исследованию, которые позиционируются как *демографический подход* (возрастная социализация *подросток*→*молодежь*→*молодые взрослые*; В.Т. Лисовский, Б.А. Ручкин, В.Я. Суртаев); *социологический подход* (Д.М. Гвишиани, И.С. Кон, Н.И. Лапин и др.); *социологический и демографический подходы* (универсальная дефиниция этого понятия, которая синтезирует социологический и демографический подходы), *ресурсный подход*. Существует *тезаурусная концепция* как совокупность всех сведений, которыми обладает субъект, происходит соотнесение слова «молодежь» с другими понятиями и их группами (К.В. Рубчевский, Л.А. Окольская и др.). Ряд исследователей связывают этапы возраста молодежи с трудовой деятельностью (Э. Котляр). Приведем необходимые в дальнейших исследованиях определения концепта «молодежь» в официальных документах Российской Федерации: «Молодежь — это социально-демографическая группа населения, выделяемая на основе возрастных особенностей и положения в обществе, в отношении которой на территории Российской Федерации гарантируется осуществление особой политики как самостоятельного и системного направления деятельности государства, важнейшего фактора обеспечения его национальной безопасности, стратегического и инновационного развития».

Проведенный анализ понятий «качество жизни», «рынок труда», «молодежь» позволил определить концептуальную основу настоящего исследования — это понимание и представление качества жизни населения, его молодежной составляющей и рынка труда как единой сложной системы, в которой компоненты качества жизни и рынка труда взаимосвязаны и нахо-

дятся во взаимодействии. Такое представление требует учета общих признаков сложных систем и закономерностей, которым сложные системы подчиняются, требует привлечения знаний общей теории систем и реализации идей системного подхода к изучению явлений и процессов в этой сложной системе. Основная черта — это системный подход, который содержательно отражает группу методов, с помощью которых реальный объект описывается как совокупность взаимодействующих компонентов, строящих структуру системы.

КОГНИТИВНАЯ МОДЕЛЬ «МОЛОДЕЖЬ, РЫНОК ТРУДА, КАЧЕСТВО ЖИЗНИ» И СЦЕНАРНОЕ МОДЕЛИРОВАНИЕ

Когнитивное имитационное моделирование сложных систем — это многоэтапный процесс [6–8, 14, 23], в котором три этапа являются основными. Это: 1) разработка и отображение структуры системы в виде когнитивной модели той или иной сложности (математически «простейшая» модель — когнитивная карта в виде знакового ориентированного графа); 2) анализ свойств когнитивной модели (связности, сложности, устойчивости и др.); 3) моделирование возможных сценариев развития процессов в системе на когнитивной модели (научное предвидение возможного будущего системы).

Модель системы «молодежь, рынок труда, качество жизни», в виде когнитивной карты $G = \langle V, E \rangle$, разработанной на основе изучения теоретической и практической информации об этой предметной области и исследований, изображена на рис. 1 [6–8, 24].

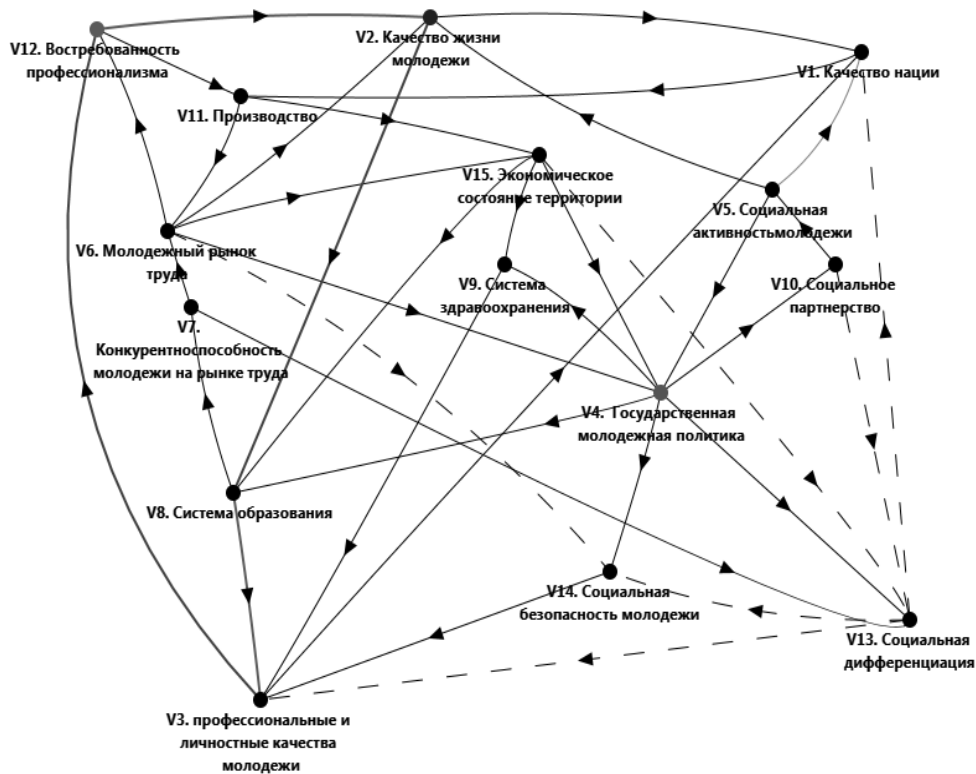


Рис. 1. Когнитивная карта G «молодежь, рынок труда, качество жизни»

В результате приведенного выше анализа были выбраны основные концепты — множество $V = \{V_i\}$ вершин знакового ориентированного графа G (табл. 1) и множество отношений (дуг) $E = \{e_{ij}\}$, $i, j = 1, 2, \dots, k$ между вершинами.

При разработке когнитивной карты удобно предварительно представлять данные о ее вершинах в форме табл. 1, в которой обосновывается выбор вершины и ее тип (назначение).

Таблица 1. Вершины когнитивной карты G , фрагмент

Код	Наименование вершины	Определение вершины	Тип вершины
V_1	Качество жизни населения	«Качество жизни» является категорией, обобщающей социально-экономические характеристики жизни людей в любом обществе и в любой исторический период, раскрывающей не только жизнедеятельность, жизнеобеспечение, но и жизнеспособность общества, как целостного социального организма. Качество жизни является основной характеристикой благосостояния народа. Повышение благосостояния народа считается важнейшей задачей экономической и социальной политики любого государства в современном мире и при оценке развитости стран качество жизни его населения становится всё более значимой категорией.	Целевая
...			
V_6	Рынок труда	«Рынок труда» — это система социально-экономических отношений, связанных с наймом и предложением труда (с его куплей и продажей); это также экономическое и географическое пространство — сфера трудоустройства, в которой взаимодействуют покупатели и продавцы специфического товара — труда; наконец, это механизм, обеспечивающий согласование цены и условий труда между работодателями и наемными работниками. Основными элементами рынка труда в любой экономической теории являются работодатель и наемный работник, от которых зависят спрос и предложение рабочей силы, ее количество, которое характеризуется либо безработицей, либо занятостью, что зависит, в том числе, от зарплаты.	Базисная, регулируемая
V_7	Уровень развития производства	«Уровень развития производства» определяется уровнем развития промышленности, сельского хозяйства и других отраслей. В условиях глобализации экономики каждая страна формирует и проводит государственную промышленно-экономическую политику, направленную на развитие определенных отраслей, проявляются тенденции интегрированной отраслевой направленности экономики, в том числе, отраслевой интегрированности различных стран. Уровень развития производства зависит от показателей деятельности предприятий, влияет на ВВП, на уровень доходов населения, качество жизни, объемы внешней торговли и др.	Базисная, управляющая
...			

Продолжение табл. 1

Код	Наименование вершины	Определение вершины	Тип вершины
V_9	Качество системы образования	«Качество системы образования» может определяться результатом — подготовкой всесторонне образованных (в смысле образование — созидание) людей, способных к саморазвитию и востребованных обществом. Качество образования — интегральная характеристика образовательного процесса и его результатов, выражающая меру их соответствия распространенным в обществе представлениям о том, каким должен быть образовательный процесс и каким целям он должен служить, степень соответствия реальных достигаемых образовательных результатов нормативным требованиям, потребностям, социальным и личностным ожиданиям, это уровень развития образования как государственно-общественной системы (в том числе любого образовательного учреждения) и его место в международном рейтинге образовательных систем. Качество образования — это важнейшая составляющая качества жизни и важнейшая характеристика страны в мировых рейтинговых системах.	Базисная, управляемая
...			
V_{15}	Государственная политика в области повышения качества жизни населения	«Государственная политика в области регулирования уровня и качества жизни населения», государственное регулирование реализуется через систему социальной политики государства и посредством проведения государственной экономической политики. Социальная государственная политика обеспечивает «прямое» воздействие на уровень и качество отдельных (слабозащищенных) слоев населения; государственная экономическая политика оказывает «косвенное» воздействие на благосостояние населения.	Базисная, управляющая
		Социальная политика — это взаимоотношение социальных групп по поводу сохранения и изменения социального положения населения в целом и составляющих его классов, слоев, социально-демографических, социально-профессиональных групп, социальных общностей (семьи, народы, население города, поселка, региона и т.п.). В практическом смысле под социальной политикой понимают совокупность (систему) конкретных мер и мероприятий, направленных на жизнеобеспечение населения. Объект и предмет социальной политики совпадают с главными элементами, блоками и структурами, входящими в крупный единый комплекс — социально-трудовую сферу — систему взаимосвязанных компонентов и частей.	
...			
V_{18}	Состояние экономики страны	«Состояние экономики страны» характеризуется ее ВВП и его динамикой, доходами населения, индексом конкурентоспособности, уровнем безработицы, строительством жилья, реализацией недвижимости, национальным богатством, политической обстановкой в мире и др.	Базисная, индикативная

Когнитивная карта (рис. 1) построена с помощью программной системы когнитивного моделирования CMCS (Cognitive Modeling Complexity System) [23].

После проверки структурных свойств когнитивной карты и свойств ее устойчивости [14, 23], показавших не противоречие модели реальной системе, был разработан план эксперимента и проведено импульсное моделирование различных сценариев развития ситуаций на данной модели (процесс

распространения возмущения по графу при известных начальных значениях $X(0)$ во всех вершинах) определяется по правилам:

$$x_i(n+1) = x_i(n) + \sum_{j=1}^{k-1} f_{ij} P_j(n) + Q_i(n),$$

где $x_i(n)$ и $x_i(n+1)$ — значения параметра в вершине V_i в моменты $n, n+1, \dots$, $P_j(t)$ — изменение в вершине V_j в момент времени n ; f_{ij} — функция отношений между вершинами; $Q(n)$ — вектор возмущений, инициирующий импульсный процесс.

Моделирование сценариев происходило с помощью приведенной модели импульсных процессов [5, 17, 18, 22] при внесении возмущающих воздействий — положительных $q_i = +1$ или отрицательных $q_i = -1$ импульсов в одну, две и более вершин; множество импульсов возмущающих воздействий (вектор возмущений) в каждом сценарии определяется как $Q = \{q_1, \dots, q_i, \dots, q_n\}$.

Количество шагов моделирования n в вычислительном эксперименте не ограничивается; импульсное моделирование может продолжаться до тех пор, пока не проявится характер тенденций изменений (постоянная величина, нарастание, убывание, колебание) во всех вершинах.

Сценарий 1. Предположим, что в системе G начала реализовываться продуманная государственная политика (вершина V_4); вводим единичный управляющий импульс $q_4 = +1$ в одну вершину; все вершины в начальном состоянии не инициированы, $q_i = 0, i = 2, 3, \dots, 18$; вектор возмущений $Q_1 = \{q_1 = 0, \dots, q_4 = +1, \dots, q_{15} = 0\}$.

Результаты импульсного моделирования приведены в табл. 2 и на рис. 2. Проиллюстрировано 6 шагов моделирования, на которых уже проявились основные тенденции. В целях облегчения визуального анализа результатов изображения импульсных процессов разбиты на 2 блока вершин: рис. 2,а — это ($V_1, V_2, V_3, V_4, V_5, V_6, V_7$); рис. 2,б — это ($V_8, V_9, V_{10}, V_{11}, V_{12}, V_{13}, V_{14}$).

Таблица 2. Результаты расчетов импульсных процессов, сценарий 1

Вершина	Шаг						
	0	1	2	3	4	5	6
V_1 . Качество нации	0	0	0	-1	4	4	16
V_2 . Качество жизни молодежи	0	0	0	0	1	14	15
V_3 . Профессиональные и личностные качества молодежи	0	0	0	3	3	3	8
V_4 . Государственная молодежная политика	0	1	1	1	2	3	2
V_5 . Социальная активность молодежи	0	0	0	1	1	1	2
V_6 . Молодежный рынок труда	0	0	0	0	1	0	11
V_7 . Конкурентоспособность молодежи на рынке труда	0	0	0	1	1	1	4
V_8 . Система образования	0	0	1	1	1	4	31
V_9 . Система здравоохранения	0	0	1	1	1	2	3
V_{10} . Социальное партнерство	0	0	1	1	1	2	3
V_{11} . Производство	0	0	0	0	-1	10	11
V_{12} . Востребованность профессионализма	0	0	0	0	6	7	6
V_{13} . Социальная дифференциация	0	0	1	0	1	2	2
V_{14} . Социальная безопасность молодежи	0	0	1	0	1	0	1
V_{15} . Экономическое состояние территории	0	0	0	0	0	0	10

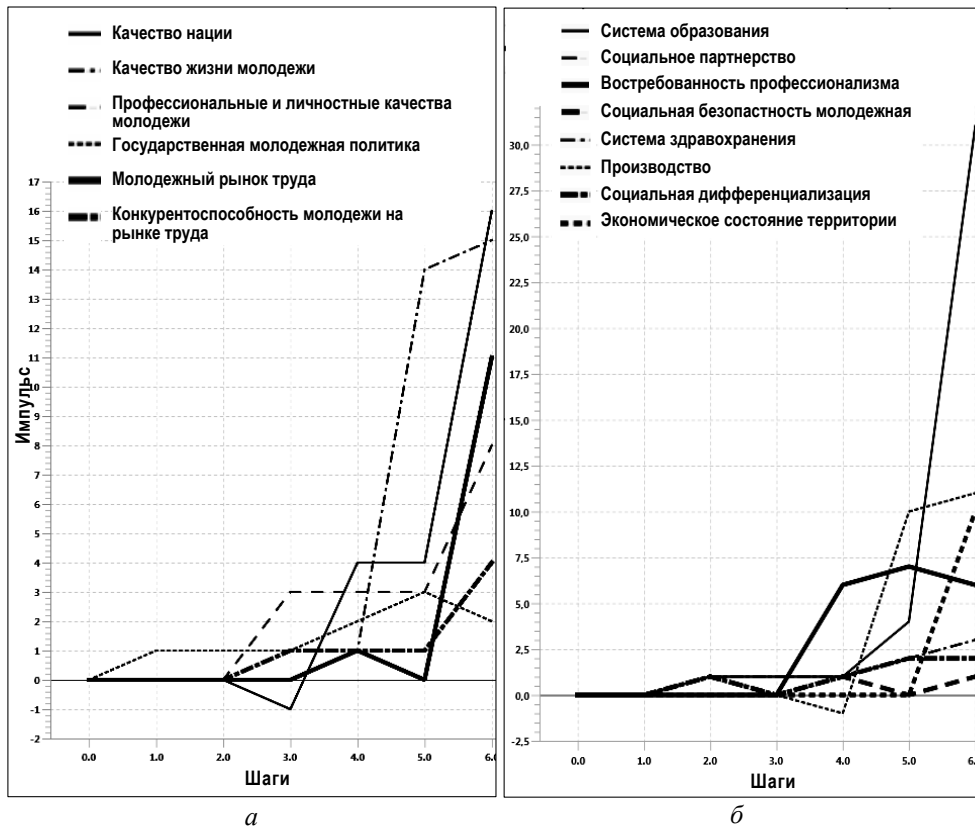


Рис. 2. Графики импульсных процессов, сценарий 1

Как видно из рис. 2, продуманная государственная политика может положительно сказаться на всех показателях системы G «молодежь, рынок труда, качество жизни». Графики импульсных процессов во всех вершинах в общем имеют тенденцию нарастания, главное при этом может повышаться качество жизни молодежи. Эти тенденции сохраняются и на большем количестве шагов моделирования.

Сценарий 2. Предположим, что состояние на молодежном рынке труда (вершина V_6) ухудшается; вводим единичный возмущающий импульс $q_6 = -1$ в одну вершину, остальные вершины в начальном состоянии не инициированы; вектор возмущений $Q_2 = \{q_1 = 0, \dots, q_6 = -1, \dots, q_{15} = 0\}$.

Результаты импульсного моделирования приведены в табл. 3 и на рис. 3.

Таблица 3. Результаты расчетов импульсных процессов, сценарий 2

Вершина	Шаг						
	0	1	2	3	4	5	6
V_1 . Качество нации	0	0	0	-1	-2	-14	-23
V_2 . Качество жизни молодежи	0	0	-1	-3	-3	-1	-53
V_3 . Профессиональные и личные качества молодежи	0	0	0	1	-10	-22	-22
V_4 . Государственная молодежная политика	0	0	-1	-2	-2	-5	-13
V_5 . Социальная активность молодежи	0	0	0	0	-1	-2	-2
V_6 . Молодежный рынок труда	0	-1	-1	-1	-2	-7	-11

Продолжение табл. 3

Вершина	Шаг						
	0	1	2	3	4	5	6
V_7 . Конкурентоспособность молодежи на рынке труда	0	0	0	0	-4	-9	-10
V_8 . Система образования	0	0	0	-4	-9	-10	-11
V_9 . Система здравоохранения	0	0	0	-2	-3	-4	-9
V_{10} . Социальное партнерство	0	0	0	-1	-2	-2	-5
V_{11} . Производство	0	0	0	-1	-2	-1	-36
V_{12} . Востребованность профессионализма	0	0	-1	-1	1	-22	-51
V_{13} . Социальная дифференциация	0	0	0	0	0	-2	-8
V_{14} . Социальная безопасность молодежи	0	0	1	0	-1	0	4
V_{15} . Экономическое состояние территории	0	0	-1	-1	-2	-4	-8

Как видно из графиков импульсных процессов (рис. 3), ухудшение положения на молодежном рынке труда может крайне негативно повлиять на всю систему «молодежь, рынок труда, качество жизни»; тенденции развития ситуаций ухудшаются уже со второго, третьего шагов моделирования. Качество жизни падает, производство ухудшается.

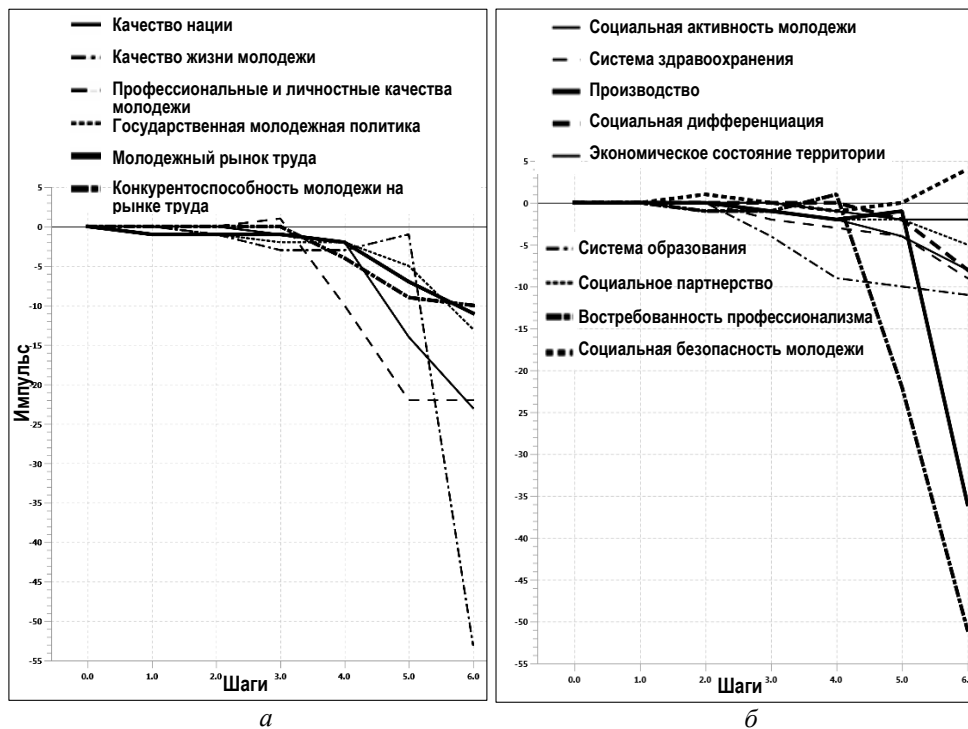


Рис. 3. Графики импульсных процессов, сценарий 2

Рассмотрим более сложные сценарии, когда возникают возмущения в нескольких вершинах.

Сценарий 3. Пусть производство будет находиться в плохом состоянии ($q_{11} = -1$), на молодежном рынке труда предложение превышает спрос на рабочую силу ($q_6 = -1$), но система образования улучшается ($q_5 = +1$), рас-

тет социальное партнерство ($q_{10} = +1$) и входит в действие продуманная государственная молодежная политика ($q_4 = +1$); вектор возмущений

$$Q_3 = \{q_1 = 0, \dots, q_4 = +1, \dots, q_6 = -1, \dots, q_8 = +1, \dots, q_{10} = +1, q_{11} = -1, \dots, q_{15} = 0\}.$$

Результаты импульсного моделирования приведены в табл. 4 и на рис. 4.

Таблица 4. Результаты расчетов импульсных процессов, сценарий 3

Вершина	Шаг						
	0	1	2	3	4	5	6
V_1 . Качество нации	0	0	0	2	5	-3	-4
V_2 . Качество жизни молодежи	0	0	-1	-3	5	26	-32
V_3 . Профессиональные и личные качества молодежи	0	0	2	7	-7	-29	8
V_4 . Государственная молодежная политика	0	1	0	-2	-1	-2	1
V_5 . Социальная активность молодежи	0	0	1	2	1	-1	0
V_6 . Молодежный рынок труда	0	-1	-2	-1	-1	-1	6
V_7 . Конкурентноспособность молодежи на рынке труда	0	0	1	2	-3	-10	7
V_8 . Система образования	0	1	2	-3	-10	7	53
V_9 . Система здравоохранения	0	0	1	-2	-5	-4	0
V_{10} . Социальное партнерство	0	1	2	1	-1	0	-1
V_{11} . Производство	0	-1	-1	-2	3	17	-19
V_{12} . Востребованность профессионализма	0	0	-1	2	13	-15	-59
V_{13} . Социальная дифференциация	0	0	0	1	2	0	-14
V_{14} . Социальная безопасность молодежи	0	0	2	2	-2	-2	-1
V_{15} . Экономическое состояние территории	0	0	-2	-3	-3	2	16

Результаты импульсного моделирования по сценарию 3 приведены в табл. 4 и на рис. 4.

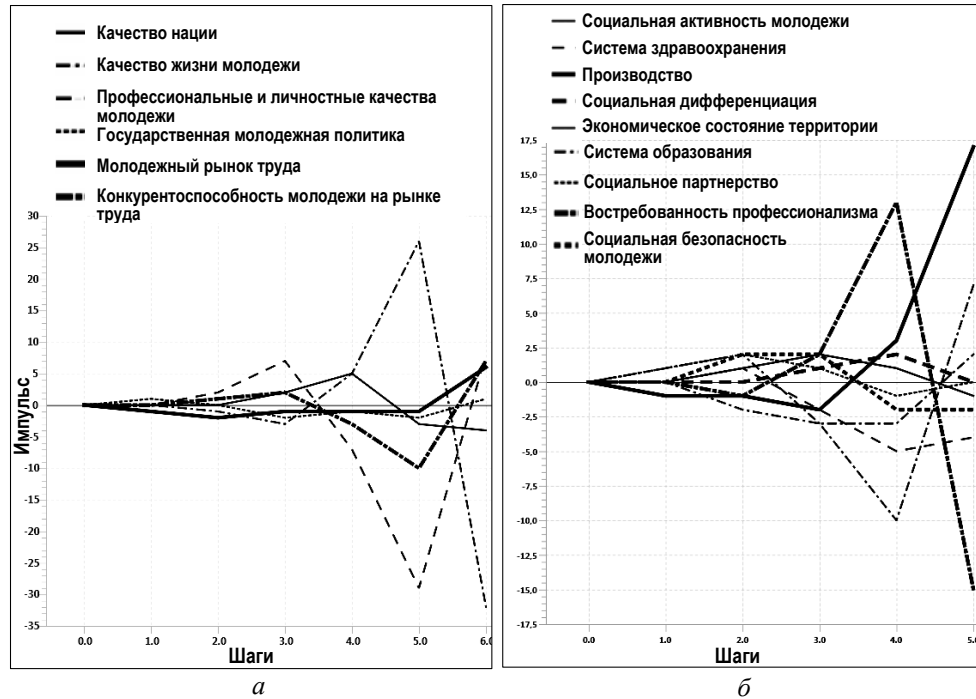


Рис. 4. Графики импульсных процессов, сценарий 3

Как видно по графикам импульсных процессов, соответствующих сценарию 3 (рис. 4), при воздействии на систему пяти разнонаправленных импульсов, среди которых есть отрицательные (производство падает, молодежный рынок труда в плохом состоянии), в ней возникает колебательный режим (ухудшение, улучшение состояний), что в общем плохо сказывается на качестве жизни и этому не могут сразу противостоять ни государственная молодежная политика, ни социальное партнерство, ни система образования.

Такой сценарий развития ситуаций можно отнести к пессимистическим, нежелательным.

Рассмотрим еще один из возможных сценариев.

Сценарий 4. Пусть начинает положительно действовать государственная молодежная политика ($q_4 = +1$), система образования улучшается ($q_8 = +1$), совершенствуется социальное партнерство ($q_{10} = +1$), социальная дифференциация падает ($q_{13} = -1$); вектор возмущений

$$Q_4 = \{q_1 = 0, \dots, q_4 = +1, \dots, q_8 = +1, \dots, q_{10} = +1, \dots, q_{13} = -1, \dots, q_{15} = +1\}.$$

Результаты импульсного моделирования показаны в табл. 5 и на рис. 5.

Таблица 5. Результаты расчетов импульсных процессов, сценарий 4

Вершина	Шаг						
	0	1	2	3	4	5	6
V_1 . Качество нации	0	0	1	6	16	33	80
V_2 . Качество жизни молодежи	0	0	0	1	15	57	92
V_3 . Профессиональные и личностные качества молодежи	0	0	3	12	16	20	85
V_4 . Государственная молодежная политика	0	1	2	3	5	11	38
V_5 . Социальная активность молодежи	0	0	1	2	3	4	6
V_6 . Молодежный рынок труда	0	0	0	1	4	16	48
V_7 . Конкурентоспособность молодежи на рынке труда	0	0	1	3	4	7	39
V_8 . Система образования	0	1	3	4	7	39	143
V_9 . Система здравоохранения	0	0	2	3	4	8	28
V_{10} . Социальное партнерство	0	1	2	3	4	6	12
V_{11} . Производство	0	0	0	1	12	41	69
V_{12} . Востребованность профессионализма	0	0	0	6	25	36	56
V_{13} . Социальная дифференциация	0	-1	-2	-1	1	1	-6
V_{14} . Социальная безопасность молодежи	0	0	2	4	3	0	-6
V_{15} . Экономическое состояние территории	0	1	1	1	3	17	58

Графики импульсных процессов сценария 4 иллюстрируют возможность положительных тенденций развития ситуаций в системе G «молодежь, рынок труда, качество жизни».

Таким образом, даже по результатам приведенных сценариев, имитирующих возможные процессы в исследуемой системе, можно сделать ряд выводов на качественном уровне, которые не противоречат теоретическим предположениям о поведении сложной системы «молодежь, рынок труда, качество жизни».

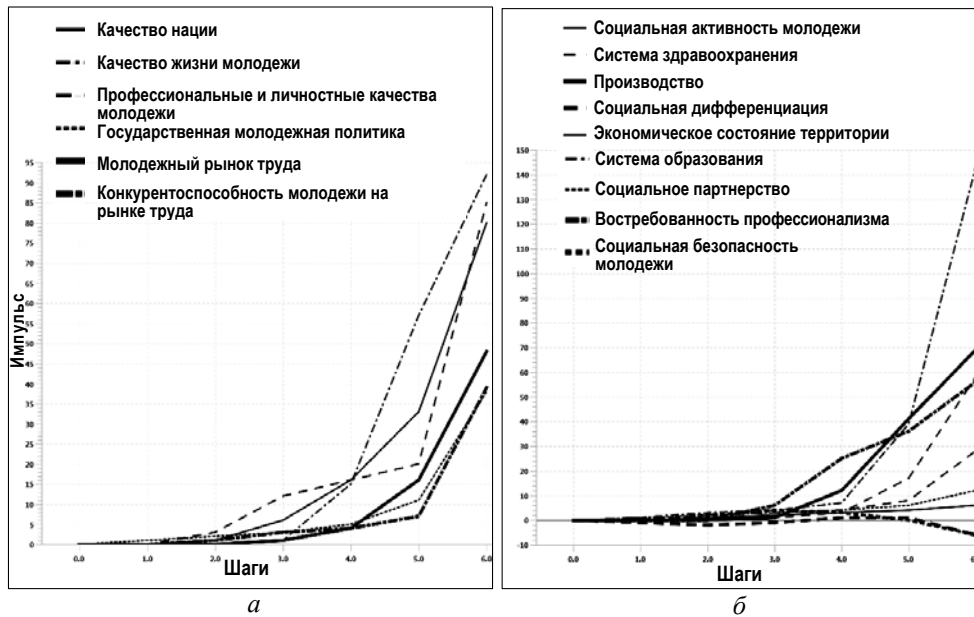


Рис. 5. Графики импульсных процессов, сценарий 4

Во-первых, продуманная государственная молодежная политика может способствовать развитию молодежного рынка труда и повышению качества жизни (сценарий 1), но гораздо в меньшей степени, чем при положительном воздействии на систему G совокупности факторов (сценарий 4), что видно по величине импульсов в вершинах V_1 и V_2 (на шаге 6 моделирования импульс в вершине качество жизни молодежи V_2 по сценарию 1 равен 15, а по сценарию 2 равен 92).

Во-вторых, состояние молодежного рынка труда может оказывать заметное влияние на качество жизни как молодежи, так и всего населения страны (сценарии 1 и 2).

В-третьих, при планировании и реализации различных сценариев на когнитивной карте следует учитывать возможность действия на изучаемую систему не только единичных акторов, но и их совокупности. При этом необходимо анализировать различные варианты «оптимистических» и «пессимистических» сценариев.

ЗАКЛЮЧЕНИЕ

Применение когнитивного подхода и инструментария имитационного когнитивного моделирования сложных систем к изучению и моделированию сценариев развития системы, названной в данном исследовании «молодежь, рынок труда, качество жизни», позволил предвидеть различные процессы развития ситуаций в этой системе, которые могут возникнуть в ней под предполагаемым воздействием различных социальных и экономических факторов, а также под воздействием регулирующих и управляющих систем. Имитационное моделирование на модели типа «когнитивная карта», как в данном исследовании, дают результаты в качественном, обобщенном виде, определяя возможные тенденции развития процессов в системе. Для получения, при необходимости, количественных результатов возможен после-

дующий переход к более сложной когнитивной модели, например, к модели типа «векторный ориентированный функциональный граф», в котором заданы параметры вершин, а некоторые отношения между вершинами определены как функции или весовые коэффициенты.

ЛИТЕРАТУРА

1. С.А. Айвазян, *Межстрановой анализ интегральных категорий качества жизни населения (эконометрический подход)*. М: ЦЭМИ РАН, 2001, 61с.
2. R. Axelrod, *The Structure of Decision: Cognitive Maps of Political Elites*. Princeton, University Press, 1976.
3. Н.А. Абрамова и З.К. Авдеева, “Когнитивный анализ и управление развитием ситуаций: проблемы методологии, теории и практики”, *Проблемы управления*, № 3, с. 85–87, 2008.
4. И.В. Бестужев-Лада, Г.С. Батыгин, и Н.П. Гришаева, “Категория «качество жизни» в современной западной социологии”, *Современные концепции уровня, качества и образа жизни*; отв. ред. И.В. Бестужев-Лада, Н.М. Блинов. М., 1978, с. 142–158.
5. J. Casti, *Connectivity, Complexity, and Catastrophe in Large-scale Systems*. A Wiley–Interscience Publication International Institute for Applied Systems Analysis. Chichester–New York–Brisbane–Toronto: JOHN WILEY and SONS, 1979.
6. Г.В. Горелова и А.А. Саак, “Занятость молодежи, качество жизни, когнитивное моделирование”, *Доклады XIII всероссийское совещание по проблемам управления, вступ-2019, Москва, 17–20 июня, 2019 г.*
7. Г.В. Горелова, О.В. Каурова, А.А. Саак, и В.А. Вишнякова, “Когнитивный подход к исследованию занятости молодежи на рынке труда”, *Социальная политика и социология*, т. 16, №1(120), с. 18–27, 2017.
8. Г.В. Горелова, “Рынок труда Юга России, когнитивный анализ развития”, *сб. статей «Системный анализ и информационные технологии»: материалы 14-й Международной научно-технической конференции SAIT 2012*. К.: УНК “ИПСА” НТУУ “КПИ”, 2012, с. 44–46.
9. А.Н. Дейнега, О.В. Кольцова, и Т.Н. Харламова, *Качество трудовой жизни как критерий оценки зрелости социально-трудовых отношений: монография*. Тамбов: Изд-во ИП Чеснокова, 2012, 140 с.
10. M. Erdem, “The Level of Quality of Work Life to Predict Work Alienation”, *Educational Sciences: Theory & Practice*, vol. 14, no. 2, pp. 534–544, 2014.
11. C. Eden, “Cognitive mapping”, *European Journal of Operational Research*, no. 36, pp.1–13, 1998.
12. Ю.А. Зубок, Т.К. Ростовская, и Н.Л. Смакотина, *Молодежь и молодежная политика в современном российском обществе*. М.: ИТД «ПЕРСПЕКТИВА», 2016, 166 с.
13. В.В. Загребин, “Подходы к определению категории “молодёжь”, *Концепт*, №02 (февраль), 2014. Доступно: <http://e-koncept.ru/2014/14030.htm>
14. *Инновационное развитие социо-экономических систем на основе методологий предвидения и когнитивного моделирования. Коллективная монография*; под ред. Г.В. Гореловой, Н.Д. Панкратовой. Киев: Наукова думка, 2015, 464 с.
15. И.М. Ильинский и В.А. Луков, “Государственная молодежная политика в России: философия преемственности и смены поколений”, *Знание. Понимание. Умение*, №4, с. 5–14, 2008.
16. С.В. Коврига и В.И. Максимов, “Когнитивная технология стратегического управления развитием сложных социально-экономических объектов в нестабильной внешней среде”, *Сб. 1-й трудов «Когнитивный анализ и управление ситуациями» (CASC'2001). Тр. междунар. конф.*, т.1, 2001.

17. T. Heiskanen and E. Jokinen, "Stability and Change of the Quality of Working Life in Restructuring Municipalities", *Social Indicators Research*, vol. 118, no. 2, pp. 579–599, 2014.
18. В.В. Кульба, Д.А. Кононов, С.С. Ковалевский, С.А. Косяченко, Р.М. Нижегородцев, и И.В. Чернов, *Сценарный анализ динамики поведения социально-экономических систем (Научное издание)*. М.:ИПУ РАН, 2002, 122 с.
19. V. Kul'ba, D. Kononov, I. Chernov, and P. Ponomarev, "Group Management in Social and Economic Systems: Research of Efficiency by Means of Cognitive Analysis", *18th IFAC Conference on Technology, Culture and International Stability (TECIS2018)*, IFAC Papers OnLine 51-30/2018, pp. 510–515.
20. A. Monkevicius, "Quality of Working Life Concept and Empirical Indicators", *Intelektine Ekonomika*, vol. 8, no. 1(19), pp. 8–24, 2014.
21. A. Raikov, "Cognitive Modeling Quality Rising by Applying Quantum and Optical-Semantic Approaches", *18th IFAC Conference on Technology, Culture and International Stability*.
22. F. Roberts, *Graph Theory and its Applications to Problems of Society*. Philadelphia: Society for Industrial and Applied Mathematics, 1978.
23. *Программа для когнитивного моделирования и анализа социально-экономических систем регионального уровня*. Свидетельство о государственной регистрации программ для ЭВМ № 2018661506 от 07.09.2018.
24. А.А. Саак, "Исследование взаимосвязи качества жизни молодежи с социально-экономической средой", *Проблемы экономики и юридической практики*, вып. 1, с. 57–60, 2018.
25. А.С. Тодоров, *Качество жизни: критический анализ буржуазных концепций*. М.: Прогресс, 1995, 164 с.
26. N. Flores, C. Jenaro, M.B. Orgaz, and M.V. Martin, "Understanding Quality of Working Life of Workers with Intellectual Disabilities", *Journal of Applied Research in Intellectual Disabilities*, vol. 24, no. 2, pp. 133–141, 2011.
27. И.А. Филюков, "Генезис научных идей о качестве жизни населения: отечественные и зарубежные подходы", *Известия высших учебных заведений. Поволжский регион. Общественные науки*, № 3 (11), с. 50–59, 2009.
28. И.В. Цыганкова, "Генезис формирования концепции качества трудовой жизни", *Вестник Новосибирского государственного университета*, № 3 (7), с. 183–190, 2007.

Поступила 01.03.2021

INFORMATION ON THE ARTICLE

Galina V. Gorelova, Institute of Management in Economic, Social and Ecological Systems of the Southern Federal University, Russia.

Andrey A. Saak, Russian State Social University (RSSU), Russia, e-mail: andrey614408@yandex.ru

СЦЕНАРНЕ КОГНІТИВНЕ МОДЕЛЮВАННЯ ТЕНДЕНЦІЙ РОЗВИТКУ СКЛАДНОЇ СИСТЕМИ «МОЛОДЬ, РИНОК ПРАЦІ, ЯКІСТЬ ЖИТТЯ» /

Г.В. Горелова, А.А. Саак

Анотація. Проведення досліджень якості життя молоді та впливу на неї такого істотного фактора, як молодіжний ринок праці, а також пов'язаних з цими дослідженнями проблем державної підтримки молоді, потребують використання методів імітаційного моделювання у зв'язку з неповнотою інформації та слабкоструктурованістю проблем таких складних систем, як якість життя молоді і ринок праці. Подано стислий огляд категорій «ринок праці», «якість життя», «молодь», необхідний для визначення змісту когнітивного моделювання

складної системи «молодь, ринок праці, якість життя», а також технологію когнітивного моделювання. Наведено результати застосування когнітивного імітаційного моделювання до вивчення якості життя молоді у взаємозв'язку з ринком праці. Приклад моделювання подано у вигляді відповідної когнітивної карти і ряду сценаріїв можливого розвитку ситуацій на ній, які можуть виникнути під передбачуваним впливом різних соціальних і економічних чинників та під впливом регулювальних і керувальних систем.

Ключові слова: молодь, ринок праці, імітаційне моделювання, когнітивна карта, сценарій, можливість, ситуація.

SCENARIO COGNITIVE MODELING OF DEVELOPMENT TRENDS OF THE COMPLEXITY SYSTEM “YOUTH, LABOR MARKET, QUALITY OF LIFE” /

G.V. Gorelova, A.A. Saak

Abstract. Conducting research on the quality of life of young people and the impact on it of such a significant factor as the youth labor market, as well as related research on the problems of state support for young people, require the use of simulation methods due to the incompleteness of information and poorly structured problems of such complex systems as the quality of life of young people and the labor market. The article gives a brief overview of the categories “labor market”, “quality of life”, “youth”, which is necessary to determine the content of cognitive modeling of the complex system “youth, labor market, quality of life”. The technology of cognitive modeling is briefly presented. The results of the application of cognitive simulation modeling to the study of the quality of life of young people in relation to the labor market are presented. An example of modeling is presented in the form of a corresponding cognitive map and a number of scenarios for the possible development of situations on it, which may arise under the assumed influence of various social and economic factors, as well as under the influence of regulatory and control systems.

Keywords: youth, labor market, simulation, cognitive map, scenario, opportunity, situation.

REFERENCES

1. S.A. Ayvazyan, *Cross-country analysis of integral categories of the population's quality of life (econometric approach)*. Moscow: TsEMI RAN, 2001, 61 p.
2. R. Axelrod, *The Structure of Decision: Cognitive Maps of Political Elites*. Princeton, University Press, 1976.
3. N.A. Abramova and Z.K. Avdeeva, “Cognitive analysis and management of the development of situations: problems of methodology, theory and practice”, *Problems of management*, no. 3, pp. 85–87, 2008.
4. I.V. Bestuzhev-Lada, G.S. Batygin, and N.P. Grishaeva, “The category “quality of life” in modern Western sociology”, *Modern concepts of the level, quality and way of life*. Moscow, 1978, pp. 142–158.
5. J. Casti, *Connectivity, Complexity, and Catastrophe in Large-scale Systems*. A Wiley–Interscience Publication International Institute for Applied Systems Analysis. Chichester–New York–Brisbane–Toronto: JOHN WILEY and SONS, 1979.
6. G.V. Gorelova and A.A. Sahak, “Youth employment, quality of life, cognitive modeling”, *Reports of the XIII All-Russian Meeting on Governance Problems, VSPU-2019, Moscow, June 17–20, 2019*.
7. G.V. Gorelova, O.V. Kaurova, A.A. Saak, and V.A. Vishnyakova, “A cognitive approach to the study of youth employment in the labor market”, *Social policy and sociology*, vol. 16, no. 1 (120), pp. 18–27, 2017.
8. G.V. Gorelova, “Labor market of the South of Russia, cognitive analysis of development”, *Sat. articles “System analysis and information technology”: materials of the 14th International Scientific and Technical Conference SAIT 2012*. Kiev: UNK “IASA” NTUU “KPI”, 2012, pp. 44–46.

9. A.N. Deynega, O.V. Koltsova, and T.N. Kharlamov, *The quality of labor life as a criterion for assessing the maturity of social and labor relations: monograph*. Publishing house of IP Chesnokov, 2012, 140 p.
10. M. Erdem, “The Level of Quality of Work Life to Predict Work Alienation”, *Educational Sciences: Theory & Practice*, vol. 14, no. 2, pp. 534–544, 2014.
11. C. Eden, “Cognitive mapping”, *European Journal of Operational Research*, no. 36, pp. 1–13, 1998.
12. Yu.A. Zubok, T.K. Rostovskaya, and N.L. Smakotina, *Youth and youth policy in modern Russian society*. Moscow: ITD “PERSPECTIVE”, 2016, 166 p.
13. V.V. Zagrebin, “Approaches to the definition of the category of “youth”, *Concept*, no. 02 (February), 2014. Available: <http://e-koncept.ru/2014/14030.htm>.
14. *Innovative development of socio-economic systems based on foresight and cognitive modeling methodologies. Collective monograph*; ed. G.V. Gorelova, N.D. Pankratova. Kiev: Naukova Dumka, 2015, 464 p.
15. I.M. Ilyinsky and V.A. Lukov, “State youth policy in Russia: philosophy of continuity and generational change”, *Knowledge. Understanding. Skill*, no. 4, pp. 5–14, 2008.
16. S.V. Kovriga and V.I. Maksimov, “Cognitive technology of strategic management of the development of complex socio-economic objects in an unstable external environment”, *Sb. 1st works “Cognitive analysis and situation management” (CASC’2001), Tr. Int. Conf.*, t.1, 2001.
17. T. Heiskanen and E. Jokinen, “Stability and Change of the Quality of Working Life in Restructuring Municipalities”, *Social Indicators Research*, vol. 118, no. 2, pp. 579–599, 2014.
18. V.V. Kulba, D.A. Kononov, S.S. Kovalevsky, S.A. Kosyachenko, R.M. Nizhegorodtsev, and I.V. Chernov, *Scenario analysis of the dynamics of the behavior of socio-economic systems (Scientific edition)*. Moscow: IPU RAN, 2002, 122 p.
19. V. Kul’ba, D. Kononov, I. Chernov, and P. Ponomarev, “Group Management in Social and Economic Systems: Research of Efficiency by Means of Cognitive Analysis”, *18th IFAC Conference on Technology, Culture and International Stability (TECIS 2018)*, IFAC Papers OnLine 51-30/2018, pp. 510–515.
20. A. Monkevicius, “Quality of Working Life Concept and Empirical Indicators”, *Intelektine Ekonomika*, vol. 8, no. 1(19), pp. 8–24, 2014.
21. A. Raikov, “Cognitive Modeling Quality Rising by Applying Quantum and Optical Semantic Approaches”, *18th IFAC Conference on Technology, Culture and International Stability*.
22. F. Roberts, *Graph Theory and its Applications to Problems of Society*. Philadelphia: Society for Industrial and Applied Mathematics, 1978.
23. *Program for cognitive modeling and analysis of socio-economic systems at the regional level*. Certificate of state registration of computer programs No. 2018661506 dated 09/07/2018.
24. A.A. Saak, “Research of the relationship between the quality of life of young people and the socio-economic environment”, *Problems of economics and legal practice*, no. 1, pp. 57–60, 2018.
25. A.S. Todorov, *Quality of Life: A Critical Analysis of Bourgeois Concepts*. Moscow: Progress, 1995, 164 p.
26. N. Flores, C. Jenaro, M.B. Orgaz, and M.V. Martin, “Understanding Quality of Working Life of Workers with Intellectual Disabilities”, *Journal of Applied Research in Intellectual Disabilities*, vol. 24, no. 2, pp. 133–141, 2011.
27. I.A. Filyukov, “Genesis of scientific ideas about the quality of life of the population: domestic and foreign approaches”, *News of higher educational institutions. Volga region. Social Sciences*, no. 3 (11), pp. 50–59, 2009.
28. I.V. Tsygankova, “The Genesis of the formation of the concept of the quality of working life”, *Bulletin of the Novosibirsk State University*, no. 3 (7), pp. 183–190, 2007.

MODELS FOR DEVELOPMENT OF THE INFORMATION-CONTROL COMPLEX OF THE ENTERPRISE

V.N. VOLKOVA, A.E. LEONOVA, A.V. LOGINOVA

Abstract. This article provides a brief analysis of the development of concepts and terms used in the automation the processes of production and enterprise management. The conclusion is made about the preference of the term “information-control complex” (ICC) of the enterprise and the expediency of using various concepts in its creation. A model for the development of ICC is proposed, based on the definition of a system that implements the system-target approach, and the “engineering” concept.

Keywords: control system, enterprise architecture, engineering, information-control/management complex.

INTRODUCTION

During the development of automation of information support for enterprise management, and new technologies emerged, the terminology, types of information systems, and standards governing their development and functioning changed. In the initial period of control automation, the following terms were used: data processing systems (DDS); databases (DB); automated control systems and automated information systems (ACS and AIS), information-retrieval system (IRS), etc.

In connection with the political and economic transformations of 1990–1991, mainly, enterprises began to acquire foreign local or domestic information systems (IS) developed for certain types of activities of organizations (accounting, personnel, etc.), create separate pages of information on the Internet. But then it is realized that in order to manage an organization, it is necessary to combine automated information systems developed for individual functions and types of activity into a single enterprise management automation system. The concept of enterprise architecture emerged. With the intensification of the development and implementation of innovations, interest in the initial understanding of the term engineering is revived.

The variety of the notions complicates the organization of work on the automation of enterprise management. Therefore, it is advisable to conduct a comparative analysis of the usefulness of these terms for specific organizations, taking into account their emergence and modern interpretation, and to develop methodological foundations and models for the further development of automation of production processes and organizational management of the enterprise.

OVERVIEW OF APPROACHES TO CREATION OF AUTOMATION SYSTEMS FOR ENTERPRISE CONTROL

In the 1960, when automation began to be used to manage enterprises in the Soviet Union, the methodology for developing automated control systems was ini-

tially based on the fundamental principles of the control developed in the theory of automatic control. In particular, the principle control on based deviations or principle feedback

$$y(t)_{\text{exit}} = F(u(t).z(t)),$$

where $y(t)_{\text{exit}} \in R_k$; R_k — vector of output signals; $z(t) \in R_p$; R_p — disturbance vector; $F = R_m \times R_p \rightarrow R_k$ — transformation that determines the type of relationship between y_{exit} , u and z ; k — number of output actions; m — number of control actions; p — number of disturbing impact $z(t)$; $u(t) \in R_m$ — control action calculated by the formula:

$$u(t) = G[| F(x(t), O_p) - y_{\text{req}} |],$$

where $x(t) \in R_m$ — predetermined control (the law of the system functioning); $z(t) \in R_p$; R_p — disturbance vector; y_{req} — desired result (setpoint); $G: R_k \rightarrow R_m$, $G(O_k) = O_m$ — feedback block conversion; $O_k \in O_p$, $O_m \in O_p$; m — number of control actions; p — number of disturbing impact $z(t)$.

In the mid-1960s, the American Society for Production and Inventory Management (APICS) formulated a number of principles according for proposed to build models of the main production processes that were applied to solve the problems of managing the inventory of an enterprise and were called the concept of MRP (Material Requirements Planning). Later, as this approach was applied to other production processes, the concepts of MRPII, ERP, etc. appeared. On this basis, software products were created, which became quite widespread.

This approach, hereinafter called the process approach, is aimed at solving practical problems associated with the activities of enterprises, and can be successfully applied to relatively small firms. However, for scientific and production associations, large enterprises of mass production, such an approach, which was originally called functional and technological, is very difficult to implement.

Therefore, during the period began to create automated information systems (AIS) as the first stage of automated control systems in the Soviet Union, a system-targeted approach was adopted. A classification of automated systems was developed, focused on different levels of enterprise management, including their structural divisions, industries and the state as a whole. The developed branch ACS (OACS), regional or territorial ACS (RASU, TASU) and enterprise ACS were supposed to be combined into the National Automated System (OGAS) for collecting, storing and searching information.

To manage the development of automated systems, the corresponding guiding methodological materials were prepared, in which the concepts of the functional part (FP), formed on the basis of the analysis of the goals and functions of the enterprise, and the supporting part (SP), including information, technical, program and other types of ACS support. The structure of the FP was initially developed in the form of a tree-like hierarchy, then, as the number of subsystems increased, in the form of matrix, multi-level and structures.

In the 1990s, when the information market of software products began to actively develop, subject-oriented and corporate information systems (SOIS, CIS) were developing on the basis of new information technologies. The most developed

automated information systems that support the internal activities of mass production enterprises include:

$$IIS = \langle ERP, MRPII, MRP, MES, PDM \rangle,$$

where *MES* — manufacturing execution system; *MRP* and *MRPII* — Material Requirements Planning; *ERP* — Enterprise Resources Planning; *PDM* — Product Data Management.

In 1987, the concept of “enterprise architecture” emerged, which was first presented in an article by J.A. Zachman [1] in the form of a two-dimensional matrix, in columns — “the main aspects of the activity (“what”, “how”, “where”, “who”, “when”, “why”); and line by line — different descriptions of the system from the point of view of business leaders, managers and developers. In 1996–1997 E.Z. Zinder proposed a “3D model” of the enterprise [2, 3]. The model introduces a axis of time, where the intervals for the implementation of various projects and stages of development of the IS and the entire enterprise are located, and the principles of the Zachman matrix are used as other axes. This model became the basis for multidimensional structures. Then it was created multidimensional models, used in the systems of the Department of Defense and other US Federal government agencies [4–9]: TAFIM (Technical Architecture Framework for Information Management). FEAF (Federal Enterprise Architecture Framework). TOGAF (The Open Group Architectural Framework), DoDAF (Department of Defense Architecture Framework), etc. Standards ISO 14258: 1998 (as amended from 2000) and ISO 15704: 2000 have been developed.

In the 1990s to combine heterogeneous software products into a single automated information system of an enterprise, the idea of an information infrastructure was proposed [11], on the basis of a modification of which, in the next section, the concept of an information management complex is substantiated and models for its implementation are proposed.

THE CONCEPT OF A MULTI-LEVEL INFORMATION-CONTROL COMPLEX

The concept and model (Fig. 1) of a multilevel information management complex (IAC) is based on the use of a system definition that implements the system-target approach [12]

$$S_{\text{def}} \equiv \langle Z, STR, TECH, COND, N \rangle, \quad (1)$$

where *Z* — goals that can be interpreted as the structure of the functional part of the ACS, as the needs of users;

STR — the structure of the relationship between *Z* and information arrays, i.e. databases; searcher arrays of documents, repositories information of various kinds;

TECH — technologies in a broad sense, including technical means (i.e. devices or devices that are necessary for collecting, registering, storing, processing and presenting information), methods of collecting, storing, processing information, including algorithms, software procedures or packages of application programs, information technologies, etc.;

COND — conditions, i.e. external and internal factors influencing the creation and functioning of ICC; for their analysis, it is useful to use the attribute

“space of target initiation”, i.e. to identify the factors of the supersystem and the actual environment (φ_{ext}), sub-departmental and the system itself (φ_{int});

N — people interacting with the system, i.e. those who order, design, use this ICC. For a general name for everyone involved in the development and use of systems, W.R. Ashby, at the beginning of the development of cybernetics and systems theory, introduced the term “observer”; and nowadays, the term “stakeholders” is often used [13].

The definition (1) and illustrating its stratified representation (Fig. 1) help to theoretically substantiate the structure of the information-control complex, to manage its development. The definition can be interpreted taking into account the purpose and conditions of its development. It is necessary to determine the relationship between the components of the user, functional and information strat-levels, to justify the choice of the necessary technologies.

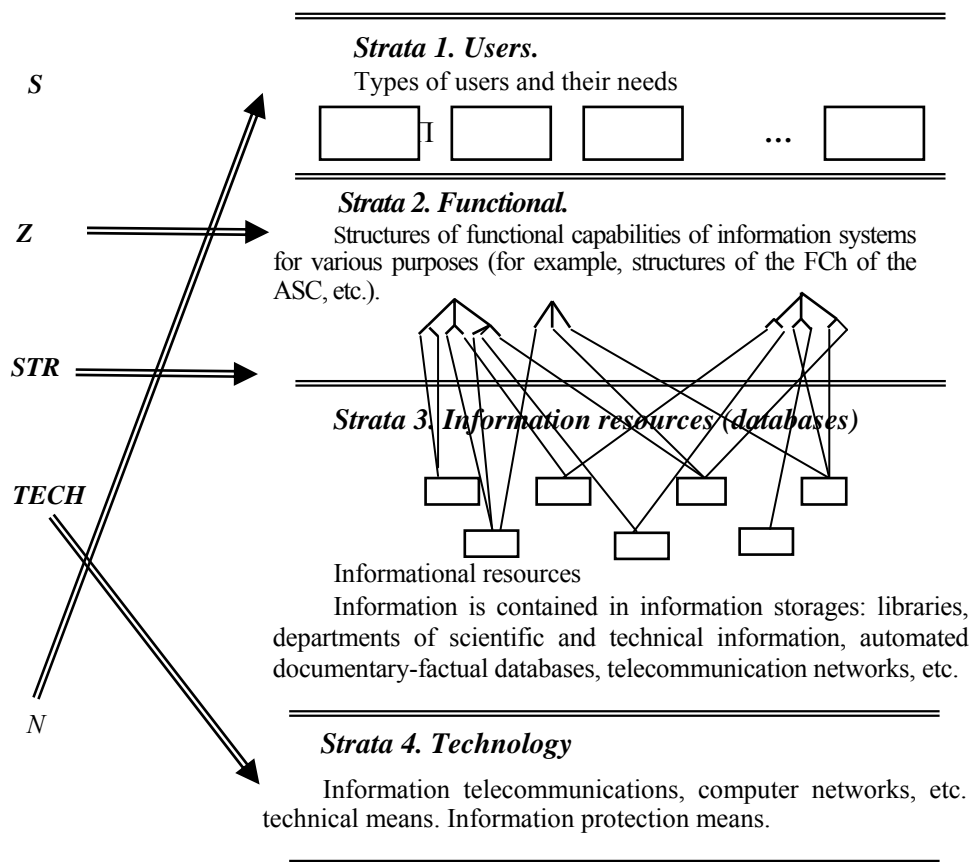


Fig. 1. Model of a multi-level information-control complex

The stratified representation helps to solve the problems of managing projects and programs for the development of ICC, allocation of financial, material and human resources based on the development and application of appropriate methods and automated procedures to determine the composition of the components of each strata, assess the impact of a project in the field of ICC on its development.

In particular, for the analysis and formation of a functional strat, it is advisable to use the methods of structuring goals and functions, methods of analyzing

information needs, developed in the theory of information retrieval. The assessment of the significance of the elements of each of the strata can be carried out using the methods of organizing complicated examinations: the method of paired comparisons in the modification of T. Saati [13], the method of deciding matrices by G.S. Pospelov [14], the informational approach of A.A. Denisov [15] and others [16, 17], take into account different criteria — technical, economic, social.

For example, you can apply the model of deciding matrices:

“*Functional strata* $\langle \alpha_j \rangle$ ” \Rightarrow “Information resources $\langle \beta_i \rangle$ ” \Rightarrow “Technology $\langle \gamma_k \rangle$ ”.

Then the algorithm for analyzing the model:

Evaluation $\alpha_j \rightarrow$ Checkness $\sum_j \alpha_j = 1 \rightarrow$ Evaluation α_{ij} for each $\alpha_j \rightarrow$
 \rightarrow Checkness $\sum_i a_{ij} = 1$ for each $\alpha_j \rightarrow$ Calculation $\beta_i = \sum_j a_{ij} \alpha_j \rightarrow \sum_i \beta_i = 1 \rightarrow$
 \rightarrow Evaluation $b_{ki} \rightarrow$ Checkness $\sum_k b_{ki} = 1$ for each $b_i \rightarrow$ Calculation
 $\gamma_k = \sum_i b_{ki} \beta_i \rightarrow \sum_k \gamma_k = 1$.

The considered concept and models make it possible to manage the development of IAI. At the same time, both the concept of enterprise architecture and the concept of a multi-level ICC help to reflect the structure of the ICC, make decisions about its development, but do not take into account the processes of the enterprise, including its interaction with the environment, receipt, execution and implementation of orders. Therefore, it is proposed to additionally use concept of the engineering in the development of ICC.

ROLE THE CONCEPT OF ENGINEERING FOR THE DEVELOPMENT OF THE INFORMATION- CONTROL COMPLEX

For the development of ICC, it is important to use the concept of “engineering” in the original meaning of this term in the 16th century (from Lat. Ingenium — ingenuity, invention, knowledge). In the XVII century engineering began to form into a separate profession. Scale drawings help to maintain the integrity of the product presentation and to see the role and place of parts in the product. Development leads to the combination of engineering with scientific progress, with technological innovation.

Engineering in the original understanding is a superstructure over design and development activities, allowing the results of engineering activities to be brought closer to their direct implementation, to obtain a new, synergistic, result, the emergence. Engineering includes all stages of the life cycle of product creation, order fulfillment — from concept to implementation and maintenance, including delivery, installation supervision, condition acceptance and field supervision, if necessary.

Engineering activities (although in the Soviet Union the foreign term “engineering” was not used) was provided by the service of the chief engineer, which coordinated not only the departments of the chief designer, chief technologist,

chief mechanic, etc. activities, but also subdivisions such as a bureau of inventions and discoveries, helping in the organization of rationalization and inventive activities, and subsequently – special design bureaus, or even research institutes, which supplemented the support of the full life cycle of engineering activities — from inventions to innovative products (new types of products or their blocks), innovative technologies that ensure the development of enterprises.

When accompanying the creation of complex products, the path from concept to implementation and support is not a unidirectional, but a complex graph, including the order or purchase of components through tenders, the distribution of work, taking into account the specialization and cooperation of not only divisions, but even enterprises. Each stage of the complete life cycle is a complex process. In particular, design and technological activities are supported by a multilevel system of drawings of various types (general block diagrams, schematic diagrams, product projections, etc.) and normative and technical documentation. To implement these works, knowledge of technical disciplines (mechanics, reliability theory, systems engineering), management disciplines (project management, quality management, etc.), economic and legal disciplines is required. This served as the basis for the implementation of engineering in the form of consulting activities in individual areas required to complete the stages and sub-stages of the project. This practice arose in the 18th century, when engineering became a professional activity and a separate discipline.

Professional associations and schools of engineering have emerged. Standards have been created [18, 19], in which engineering is interpreted as technical consulting services related to the development and preparation of the production process and the provision of the production process and the sale of products. This form of engineering is widely used in modern practice. There are specialized engineering centers, mainly engaged in automated support of the stages of the life cycle of development and implementation of innovations, professional communities and associations that unite engineers from different countries. There are different points of view both on the content of engineering activities and on the concept of “engineering”. The content of the concept of “engineering” is constantly expanding, including areas that are more and more distant from the classical engineering activity. Therefore, it is proposed to investigate the essence of this concept using the theory of an open system with active elements.

On the basis of his research, the Belgian biologist L. von Bertalanffy discovered that in open systems, as opposed to closed (isolated from the environment), thermodynamic laws appear that contradict the second law of thermodynamics. In accordance with the concept of L. von Bertalanffy “... it is quite possible to introduce negentropy”, that is, to decrease the entropy; and “... such systems can maintain their high level and even develop in the direction of increasing the order of complexity” [20].

L. von Bertalanffy explained negentropic tendencies mainly by the openness of the system. Further studies of development processes made it possible to understand that development occurs not only due to the openness of the system, but also due to the active elements that initiate innovation. The studies of the Russian scientist E. Bauer allowed a deeper understanding of the development processes. E. Bauer investigated one of the fundamentally important for understanding the development of living systems regularities of the fundamental disequilibrium of living systems, that is, the desire to maintain a stable imbalance and use energy to

maintain oneself in a disequilibrium state. E. Bauer explains this by the fact that all structures of living cells at the molecular level are pre-charged with “excess” excess energy in comparison with a non-living molecule, and the body receives external energy not for work, but to maintain itself in a nonequilibrium state [21]. On the basis of understanding the ideas of E. Bauer (eg, [22]), it can be concluded that, in fact, we are talking about the biopotential of a “living cell”, which is determined not only by physiological energy as a source of negentropy. At a certain level of development of a living thing, such a source of negentropic processes can be interpreted as a potential based on information and knowledge. Then we can conclude that, taking into account the research of E. Bauer, for the development of any organization (enterprise, territorial association, etc.), it should have a kind of “living cell” that accumulates energy/information in order to create innovations that drive the development of the organization.

Such an interpretation of the source of negentropic processes can be associated with the concept of “engineering” in the original sense of this term (from the Latin *Ingenium* — ingenuity, knowledge), that is, in accordance with the definition given in the Oxford Dictionary, in which engineering is interpreted as the activity of applying scientific knowledge for the design, construction, control production of machines and plants, etc. Based on this, it can be concluded that for the development of an organization, engineering tasks should be solved by teams of qualified professionals specializing in the relevant types of professional activity.

To manage this activity, the coordination of the relevant sub-divisions or organizations performing this work is necessary. Such coordination requires the creation of a system of information support at all stages of the life cycle of the activity, for which the concept of engineering is applied, that is, it is necessary to create a unified information-control complex, including software for engineering and design activities, accompanying regulatory methodological, regulatory-technical, organizational and administrative documentation. Therefore, it is not enough for organizations to use the advisory services of specialized organizations, but it is necessary to develop and apply the form of engineering in the original sense of the notion.

This form represents design-technological and regulatory-technical support for the entire life cycle — from the invention of an innovation or receipt of an order, the development of the structure of its research, design, technological, production implementation and delivery to the customer. To provide such support, it is necessary to create an automated information-control complex of an enterprise that implements the engineering concept. This information complex provides the necessary coordination of all stages of the life cycle of the innovation process, for which the idea of engineering is used.

Taking this into account, when creating an ICC, it is necessary to have subdivisions that, in accordance with E. Bauer’s concept, accumulate energy/information in order to invent innovations that ensure the development of the organization. To aid in decision-making, it should be possible to include in such an ICC models that help in making decisions on the analysis and selection of innovations (eg [23–31]). It is also necessary to create an environment that would ensure interconnection in the information space and access to information of persons using it at the appropriate stages of the life cycle. The basis of such an environment can be an intelligent knowledge representation system of the type proposed in [32]. The development of such a system and the coordination of engineering work at the enterprise should be part of the functions of the unit dealing with the organization of the strategic development of the enterprise (organization).

CONCLUSION

Analysis of the concepts and notions used in the creation of automated information support for production processes and enterprise management allows us to draw some conclusions.

The notion ASC, which determined the ultimate desired goals of automating the management of enterprises and organizations, outstripped the capabilities of technologies that existed to create such systems at that time. Experience in the creation and application of software products for individual arms of activity (also called automated systems — accounting, personnel, etc.), solve the problem of ensuring the business processes of small and medium-sized companies. However, for large enterprises and organizations, other forms of integrated information support are needed.

The concept of enterprise architecture has made it possible to create multidimensional information bases. However, despite the fact that, in accordance with the concept of the Gartner Group, architecture is considered not only as a static model of a complex system, but also as a process, a set of guidelines and rules that determine the construction of new subsystems, this concept still does not focus on information support processes of creation and implementation of innovations, that is, to display the dynamics of the functioning of the system. Perhaps that is why recently there has been a revival of interest in the concept of engineering, including among the author of the architectural concept of a 3D-enterprise [33].

The use of the definition of a system based on the system-target approach made it possible to substantiate the concept of a multi-level information and control complex, linking goals, sources of information and technical means to ensure enterprise management. The use of the concept of an open system by L. von Bertalanffy and the study of the features of this system in the works of E. Bauer made it possible to substantiate the role of engineering as an innovative technology for the development system of information support for control and managing the development of enterprises, territorial and other organizations.

Therefore, the idea of creating an ICC, implementing the concept of engineering, automating the collection of information at all stages of the life cycle of production and organizational management of enterprise, and the creation on this based the unified information-control environment, the basis of which can be an intelligent knowledge representation system of the type proposed, for example, in [32], seems perspective. The implementation of the ICC allows you to combine different concepts and manage the development of information support for the enterprise.

REFERENCES

1. J.A. Zachman, “The structure of information systems architecture”, *IBM Systems Journal*, vol. 26, no. 3, 1987.
2. E.Z. Zinder, “New system design: information technology and system design”, *DBMS*, no. 4, 1, 1995.
3. E.Z. Zinder, “3D-enterprise — a model of a transforming system”, *Director of Information Service*, no. 4.
4. A. Danilin and A. Slyusarenko, *Architecture and Strategy. “Yin” and “Yang” of information technologies*. M.: Internet University of Information Technologies, 2005. Available: INTUIT.ru/razym.ru/komp/59613-danilin-a-slyusarenko-a

5. R. Sessions, *Comparison of the four leading architecture building methodologies*, 2007. Available: <http://msdn.microsoft.com/ru-ru/library/ee914379.aspx>
6. S. Kotusev, “TOGAF: Just the Next Fad That Turned into a New Religion”, *TOGAF Is Not an EA Framework: The Inconvenient Pragmatic Truth*, K.L. Smith (ed.), Great Notley, UK: Pragmatic EA Ltd., 2018, pp. 27–40.
7. Cole Perks and Tony Beveridge, *Enterprise IT Architecture Guide*. New York, NY: Springer, 2003.
8. Scott R. Bittler and Greg Kreutzman, *Gartner Enterprise Architecture Process: Development*, 2005.
9. N. Dedic, “FEAMI: A Methodology to include and to integrate Enterprise Architecture Processes into Existing Organizational Processes”, *IEEE Engineering Management Review*. doi: 10.1109/EMR.2020.3031968.
10. W.J. Clinton and A. Gore, “Technology for America’s Economic Growth, a New Direction to Build”, *Executive Office of the President*. Washington, DC, 1993, 39 pp.
11. V.N. Volkova, “Razvitiye opredeleniya sistemy [System definition development]”, *Mater. Mezhdunarodnoy nauchno-prakticheskoy konferentsii «Sistemnyy analiz proyektirovaniya i upravleniya»: sb.* SPb.: Izd-vo SPbGPU, 2001, pp. 12–14.
12. R.L. Ackoff, *Re-Creating the Corporation: A Design of Organization for the 21st Century*. New York, Oxford: Oxford University Press, 1999.
13. Thomas L. Saaty and Kevin P. Kearns, *Analytical Planning: The Organization of Systems*, 1985.
14. G.S. Pospelov and V.A. Irikov, *Programmno-tselevoye planirovaniye i upravleniye* [Target planning and management]. M.: Sov. radio, 1976, 440 p.
15. A.A. Denisov, *Sovremennyye problemy sistemnogo analiza* [Modern problems of systems analysis]. SPb.: Izd-vo Politekhn. un-ta, 2008, 304 p.
16. V.N. Volkova and V.N. Kozlov (eds.), *Modelirovaniye sistem i protsessov* [Modeling systems and processes]. M.: Yurayt Publ., 2015, 449 p.
17. N.D. Pankratova, *Sistemnyy analiz. Teoriya ta zastosuvannya* [System analysis. Theory and application]. Kiiv: Naukova dumka, 2018, 347 p.
18. *GOST R 57306-2016. Engineering. Terminology and basic concepts in the field of engineering*. Available: docs.cntd.ru/document/437092138.
19. *ISO/IEC 15288: (IEEE Std 15288—2008). Systems and Software Engineering — System Life Cycle Processes*. International Organization for Standardization.
20. L. von Bertalanfy, “System Theory — a Critical Review”, *General System*, vol. 7, pp. 1–20, 1962.
21. E. Bauer, *Teoreticheskaya biologiya* [Theoretical biology]. M.L.: VIEM Padl., 1935, 206 p.
22. A.L. Shamis, *Modeli povedeniya, vospriyatiya i myshleniya* [Models of behavior, perception and thinking]. M.: Intuit, 2010, 230 p.
23. V.N. Volkova, V.I. Karpov, A.E. Leonova, A.V. Loginova, and E.V.L. Romanova, “Concept of engineering as a basis for development cyber-physical system”, *CEUR Workshop Proceedings*, 2019.
24. V.N. Volkova, Yu.A. Golub, and A.V. Loginova, “Multilevel Model for Management the Regional System of Training in Conditions of Introducing Digital Technologies”, *Proceedings of 2018 17th Russian Scientific and Practical Conference on Planning and Teaching Engineering Staff for the Industrial and Economic Complex of the Region, PTES 2018*, pp. 181–184, 8604276.
25. V.N. Volkova, A.S. Kudryavtseva, A.V. Loginova, Y.Y. Chernyy, and A.E. Leonova, “System Analysis of Innovative Technologies of the Industrial Revolutions”, *Proceedings of the 3rd International Conference Ergo-2018: Human Factors in Complex Technical Systems and Environments, Ergo 2018*, pp. 57–61.
26. V. Volkova, A. Leonova, A. Loginova, and Y. Chernyy, “System analysis of the development of information-communication technologies”, *CSIS’2019: XI International Scientific Conference Communicative Strategies of the Information Society, ACM International Conference Proceeding Series, ACM, New York, 2019*, pp. 1–6. Available: <https://doi.org/10.1145/3373722.3373765>

27. V.N. Volkova, A.E. Leonova, E.V. Romanova, and Y.Y. Chernyy, "Engineering as a Coordinating Method for the Development of the Organization and Society", *Lecture Notes in Networks and Systems*, 184, pp. 12–21, 2021.
28. V. Volkova, G. Gorelova, and N. Pankratova, "The Development of the Cyberphysical System Concept on Base of the Interdisciplinary Theories", *IEEE 2nd International Conference on System Analysis and Intelligent Computing, SAIC 2020*, 9239213.
29. V.N. Volkova, B.S. Fleishman, F.P. Tarasenko, and A.V. Loginova, "Further Development of Potential Feasibility Theory for Complicated Systems According to the Unified General-System Principle", *Lecture Notes in Networks and Systems*, 184, pp. 446–453, 2021.
30. N. Pankratova, P. Bidyuk, and I. Golinko, "Decision support system for microclimate control at large industrial enterprises", *CEUR Workshop Proceedings*, pp. 489–498, 2020.
31. N. Pankratova, P. Bidyuk, and I. Golinko, "Synthesis of a multidimensional control system for a precision complex of artificial microclimate", *System Research & Information Technologies*, no. 1, pp. 7–20, 2020. doi: <https://doi.org/10.20535/SRIT.2308—8893.2020.1.01>
32. Ye.S. Kuzin, *Predstavleniye znaniy i resheniye informatsionno-slozhnykh zadach v komp'yuternykh sistemakh* [Representation of knowledge and solution of information-complex problems in computer systems]. M.: Novyye tekhnologii, 2004, 32 p.
33. E.Z. Zinder, "Values-directed enterprise engineering", *Business Informatics*, no. 3 (45), pp. 7–19, 2018.

Received: 17.03.2020

INFORMATION ON THE ARTICLE

Violetta N. Volkova, ORCID: 0000-0001-9658-5135, Peter the Great St. Petersburg Polytechnic University, Russia, e-mail: violetta_volkova@list.ru

Alla E. Leonova, ORCID: 0000-0002-2124-0295, Scientific Research Center for Electronic Computing JSC "NITSEVT", Russia, e-mail: alla.leonova@nicevt.ru

Alexandra V. Loginova, ORCID: 0000-0002-7783-7795, Peter the Great St. Petersburg Polytechnic University, Russia, e-mail: alexandra-lo@yandex.ru

МОДЕЛІ ДЛЯ РОЗВИТКУ ІНФОРМАЦІЙНО-КЕРУВАЛЬНОГО КОМПЛЕКСУ ПІДПРИЄМСТВА / В.М. Волкова, А.Е. Леонова, О.В. Логінова

Анотація. Подано стислий аналіз розвитку понять і термінів, що використовуються в автоматизації процесів виробництва та управління підприємством. Зроблено висновок про перевагу терміна «інформаційно-керувальний комплекс» (ІКК) підприємства та доцільність використання різних концепцій під час його створення. Запропоновано модель розвитку ІКК, засновану на визначенні системи, що реалізує системно-цільовий підхід та концепцію «інжиніринг».

Ключові слова: система управління, архітектура підприємств, інжиніринг, інформаційно-керувальний комплекс.

МОДЕЛИ ДЛЯ РАЗВИТИЯ ИНФОРМАЦИОННО-УПРАВЛЯЮЩЕГО КОМПЛЕКСА ПРЕДПРИЯТИЯ / В.Н. Волкова, А.Е. Леонова, А.В. Логинава

Аннотация. Приведен краткий анализ развития понятий и терминов, используемых в автоматизации процессов производства и управления предприятием. Сделан вывод о предпочтительности термина «информационно-управляющий комплекс» (ИУК) предприятия и целесообразности использования при его создании различных концепций. Предложена модель разработки УИС, основанная на определении системы, реализующей системно-целевой подход, и концепции «инжиниринг».

Ключевые слова: система управления, архитектура предприятия, инжиниринг, информационно-управляющий комплекс.

О НЕКОТОРЫХ СТАТИСТИКАХ ФРАКТАЛЬНОГО БРОУНОВСКОГО ДВИЖЕНИЯ

В.Г. БОНДАРЕНКО

Аннотация. Предложен и обоснован метод оценивания параметров стохастического процесса — фрактального броуновского движения — дисперсии и одношаговой ковариации приращений. Доказана среднеквадратичная состоятельность построенных оценок. Полученные результаты дополняют и обобщают следствия из предельных теорем для фрактального броуновского движения, доказанных в ряде работ. Необходимость оценивания дисперсии вызвана отсутствием эталонной единицы времени, а оценка ковариации позволяет определить показатель Харста. Установленные результаты позволяют использовать известные предельные теоремы при построении критериев согласия для гипотезы «наблюдаемый временной ряд является реализацией фрактального броуновского движения» и оценить ошибку оптимального прогноза временного ряда.

Ключевые слова: фрактальное броуновское движение, оценивание параметров, проверка статистических гипотез.

ПОСТАНОВКА ЗАДАЧИ И ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

При построении стохастических моделей временных рядов в ряде случаев целесообразно использовать в качестве базового процесса фрактальное броуновское движение (fractional Brownian motion — fBm). Этот процесс обозначается $B_H(t)$, $t \geq 0$, $B_H(0) = 0$, где H — показатель Харста, $0 < H < 1$, и определяется как гауссовский случайный процесс с нулевым средним и корреляционной функцией

$$R(t, s) = E(B_H(t)B_H(s)) = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}),$$

т.е. n -мерное распределение определяется плотностью вероятностей

$$p(t_1, \dots, t_n, x_1, \dots, x_n) = c \exp \left\{ -\frac{1}{2} \sum r^{jk} x_j x_k \right\},$$

где r^{jk} — элементы матрицы, обратной к корреляционной $r^{jk} = R(t_j, t_k)$.

Заметим, что при $H = 0,5$ получаем стандартный винеровский процесс $w(t)$.

Фрактальное броуновское движение обладает свойством автомодельности (самоподобия), что означает совпадение распределений случайных величин:

$$B_H(\alpha t) \text{ и } \alpha^H B_H(t), \alpha > 0 \text{ (обозначение } B_H(\alpha t) \sim \alpha^H B_H(t)).$$

Вследствие этого свойства в выражении $\sigma B_H(t)$ можно положить коэффициент $\sigma = 1$, и отсутствие эталона для единицы времени требует оценивания временного интервала.

Приращения фрактального броуновского движения

$$\xi_k = B_H(t + ks) - B_H(t + (k - 1)s), \quad k = 1, 2, \dots, n \quad (1)$$

образуют гауссовский случайный вектор $\xi \sim \mathfrak{N}(0; B)$, где элементы b_{jk} корреляционной матрицы B имеют вид

$$b_{jk} = \frac{s^{2H}}{2} (|j - k + 1|^{2H} + |j - k - 1|^{2H} - 2|j - k|^{2H}), \quad (2)$$

т.е. последовательность $\{\xi_k\}$ стационарна.

Свойства фрактального броуновского движения рассмотрены в работах [1, 2]; в частности, обобщена формула Ито для фрактальной диффузии.

В работах [3–8] доказан ряд предельных теорем для функций от приращений

$$\eta_k = \sigma \left(B_H \left(t + \frac{k}{n} \right) - B_H \left(t + \frac{k-1}{n} \right) \right)$$

фрактального броуновского движения, соответствующих временному интервалу $s = \frac{1}{n}$. Используя автомодельность, эти результаты можно обобщить для произвольного s , что позволяет использовать их для статистической обработки данных. При этом необходимо оценить дисперсию приращения s^{2H} и показатель Харста H , который находится из равенства

$$\rho_1 = 2^{2H-1} - 1,$$

где ρ_1 — одношаговый коэффициент корреляции приращений, определенных равенством (1):

$$\rho_1 = \frac{\text{cov}(\xi_k, \xi_{k+1})}{s^{2H}}.$$

Обозначим $\theta = \text{cov}(\xi_k, \xi_{k+1})$. В данной работе доказана консистентность (состоятельность) в среднеквадратичном смысле следующих несмещенных оценок:

- оценки дисперсии $\widehat{s^{2H}} = \frac{1}{n} \sum_{k=1}^n \xi_k^2$;
- оценки одношаговой ковариации $\hat{\theta} = \frac{1}{n-1} \sum_{j=1}^{n-1} \xi_j \xi_{j+1}$.

Полученные результаты позволяют проверить ряд статистических гипотез для исследуемых временных рядов.

При доказательстве используются следующие факты:

1. Для гауссовского вектора $\xi \sim \mathfrak{N}(0; B)$ момент четвертого порядка выражается формулой

$$E(\xi, h)(\xi, k)(\xi, u)(\xi, v) = (Bu, k)(Bv, h) + (Bv, k)(Bu, h) + (Bu, v)(Bk, h)$$

или в координатной форме

$$E\xi_i\xi_j\xi_k\xi_l = b_{il}b_{jk} + b_{jl}b_{ik} + b_{ij}b_{kl}.$$

2. Для функции

$$f(x) = (1+x)^\alpha + (1-x)^\alpha - 2, \quad 0 < \alpha < 2$$

справедлива оценка

$$|f(x)| \leq 4x^2, \quad \text{если } |x| \leq \frac{1}{\sqrt{2}}, \quad (3)$$

которая следует из тейлоровского разложения $f(x)$.

3. Частная сумма ряда зета-функции Римана $\zeta_p(t) = \sum_{m=1}^p \frac{1}{m^t}$, $0 < t < 1$ удовлетворяет оценке

$$\zeta_p(t) < \frac{p^{1-t}}{1-t}; \quad \zeta_p(1) < C + \ln p. \quad (4)$$

ДОКАЗАТЕЛЬСТВО ПРЕДЕЛЬНОЙ ТЕОРЕМЫ

Теорема. Для статистик $\widehat{s^{2H}}$, $\hat{\theta}$ справедливы предельные соотношения:

$$\delta_{1n} = E(\widehat{s^{2H}} - s^{2H})^2 \rightarrow 0, \quad n \rightarrow \infty;$$

$$\delta_{2n} = E(\hat{\theta} - \theta)^2 \rightarrow 0, \quad n \rightarrow \infty.$$

Иначе, если ξ_1, \dots, ξ_n наблюдаемы, то $\frac{1}{n} \sum_{k=1}^n \xi_k^2$ является состоятельной оценкой дисперсии приращения s^{2H} , $\frac{1}{n-1} \sum_{j=1}^{n-1} \xi_j \xi_{j+1}$ — состоятельной оценкой одношаговой ковариации.

Доказательство. Исходные выражения приводятся к виду ($b_{kk} = s^{2H}$):

$$\delta_{1n} = \frac{2}{n} s^{4H} + \frac{4}{n^2} \sum_{k>j} b_{jk}^2;$$

$$\delta_{2n} = \frac{1}{n-1} s^{4H} + \frac{1}{(n-1)^2} \left(2 \sum_{k>j} b_{jk}^2 + \sum_{j=1}^{n-1} \sum_{k=1}^{n-1-j} b_{j(k+1)} b_{k(j+1)} \right).$$

Подставляя b_{jk} из формулы (2) в сумму $\sum_{k>j} b_{jk}^2$, получаем

$$\sum_{k>j} b_{jk}^2 = \frac{s^{4H}}{4} \sum_{k=2}^n \sum_{j=1}^{k-1} ((k-j+1)^{2H} + (k-j-1)^{2H} - 2(k-j)^{2H})^2,$$

и заменой $(j, k) \rightarrow (m, k)$, $m = k - j$, приведем это выражение к виду

$$\begin{aligned} & \frac{s^{4H}}{4} \sum_{k=1}^n \sum_{m=1}^{k-1} ((m+1)^{2H} + (m-1)^{2H} - 2m^{2H})^2 = \\ & = \frac{s^{4H}}{4} \sum_{k=1}^n \sum_{m=1}^{k-1} \left(\frac{1}{m^{2H}} f\left(\frac{1}{m}\right) \right)^2, \end{aligned}$$

где функция $f(x) = (1+x)^{2H} + (1-x)^{2H} - 2$, $0 < H < 1$.

Оценим слагаемые внутренней суммы.

Из неравенства (3) для слагаемых внутренней суммы

$$I_k = \sum_{m=1}^{k-1} \left(\frac{1}{m^{2H}} f\left(\frac{1}{m}\right) \right)^2$$

вытекает оценка

$$\frac{1}{m^{2H}} f\left(\frac{1}{m}\right) < \frac{4}{m^{2-2H}}, \text{ т.е. } \left(\frac{1}{m^{2H}} f\left(\frac{1}{m}\right) \right)^2 < \frac{16}{m^{4-4H}}$$

в силу неравенства (4)

$$I_k < C, H < \frac{3}{4}; I_k < C + \ln k, H = \frac{3}{4}; I_k < \frac{1}{4H-3} k^{4H-3}, \frac{3}{4} < H < 1,$$

и окончательно

$$\sum_{k=1}^n I_k = o(n^2), n \rightarrow \infty, \text{ т.е. } \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k>j} b_{jk}^2 = 0.$$

Далее оценим выражение

$$\sum_{j=1}^{n-1} \sum_{k=1}^{n-1} b_{j(k+1)} b_{k(j+1)},$$

представив его в виде суммы трех слагаемых ($j = k$, $|j - k| = 1$, $|j - k| \geq 2$):

$$\begin{aligned} & \sum_{j=1}^{n-1} \sum_{k=1}^{n-1} b_{j(k+1)} b_{k(j+1)} = \sum_{k=1}^{n-1} b_{k(k+1)}^2 + 2 \sum_{k=2}^{n-1} b_{(k-1)(k+1)} b_{kk} + \\ & + \frac{s^{4H}}{2} \sum_{k=3}^{n-1} \sum_{j=1}^{k-2} ((k-j+2)^{2H} + (k-j)^{2H} - 2(k-j+1)^{2H})(k-j)^{2H} + \\ & + (k-j-2)^{2H} - 2(k-j-1)^{2H}) = \\ & = (n-1)s^{4H}(2^{2H-1} - 1)^2 + \frac{1}{2}(n-2)s^{4H}(3^{2H} + 1 - 2^{2H+1}) + \frac{s^{4H}}{2} \sum_{k=4}^{n-1} J_k, \end{aligned}$$

где $J_k = \sum_{m=2}^{k-1} (m+1)^{2H} (m-1)^{2H} f\left(\frac{1}{m+1}\right) f\left(\frac{1}{m-1}\right)$.

Из неравенства (3) для слагаемых в J_k следует оценка

$$(m+1)^{2H} (m-1)^{2H} f\left(\frac{1}{m+1}\right) f\left(\frac{1}{m-1}\right) < 16 \frac{1}{(m-1)^{2-2H}} \frac{1}{(m+1)^{2-2H}} < \frac{C}{m^{4-4H}},$$

что, как и в предыдущем случае, приводит к соотношению

$$\sum_{k=4}^{n-1} J_k = o(n^2), \quad n \rightarrow \infty.$$

ПРИЛОЖЕНИЯ К СТАТИСТИКЕ ВРЕМЕННЫХ РЯДОВ

В процессе обработки временных данных рассматриваются, в частности, задачи оценивания параметров и проверки гипотез. В работе [3] доказана предельная теорема для абсолютных моментов приращений фрактального броуновского движения. Для приращений, определенных формулой (1), этот результат выглядит следующим образом.

Для последовательности случайных величин

$$R_{qn} = \frac{1}{n} \sum_{k=1}^n |\xi_k|^q \equiv \frac{1}{n} \sum_{k=1}^n |B_H(t+ks) - B_H(t+(k-1)s)|^q, \quad q \text{ — натуральное,}$$

справедливо утверждение: с вероятностью 1

$$R_{qn} \rightarrow E(q), \quad n \rightarrow \infty, \quad \text{где } E(q) = ER_{qn} = \frac{1}{\sqrt{\pi}} 2^{\frac{q}{2}} s^{qH} \Gamma\left(\frac{q+1}{2}\right).$$

Из доказанной предельной теоремы следует среднеквадратичная сходимость R_{2n} .

При моделировании временных рядов актуальной является задача идентификации данных. В работах [4–8] доказан ряд предельных теорем для фрактального броуновского движения. Для значений процесса $B_H(ks)$ и приращений (1) утверждения теорем принимают вид: статистики A_n , B_n , D_n , F_n , которые удовлетворяют предельным соотношениям:

$$\begin{aligned} A_n &= \frac{1}{n} \frac{1}{s^{4H}} \sum_{k=1}^n B_H((k-1)s) \xi_k^3 \rightarrow -\frac{3}{2}, \quad H \in \left(0; \frac{1}{2}\right); \\ B_n &= \frac{1}{n^{1+H}} \frac{1}{s^{sH}} \sum_{k=1}^n (B_H((k-1)s))^2 \xi_k^3 \rightarrow 3\eta, \quad \eta \sim \mathcal{N}\left(0; \frac{1}{2H+2}\right), \quad H \in \left(0; \frac{1}{2}\right); \\ D_n &= \frac{1}{n^{2H}} \frac{1}{s^{4H}} \sum_{k=1}^n B_H((k-1)s) \xi_k^3 \rightarrow \frac{3}{2} \zeta^2, \quad \zeta \sim \mathcal{N}(0; 1), \quad H \in \left(\frac{1}{2}; 1\right); \\ F_n &= \frac{1}{n^H} \frac{1}{s^{3H}} \sum_{k=1}^n \xi_k^3 \rightarrow 3\zeta, \quad H \in \left(\frac{1}{2}; 1\right). \end{aligned}$$

Здесь сходимость понимается в среднеквадратичном

$$\eta_n \rightarrow \eta, \quad \text{если } E(\eta_n - \eta)^2 \rightarrow 0.$$

При известных значениях параметра H и дисперсии s^{2H} эти соотношения можно использовать для проверки гипотезы T : «наблюдаемые данные являются реализацией фрактального движения».

Стандартный алгоритм (согласно двустороннему критерию) такой проверки при известном H состоит в следующем: предполагаем, что гипотеза T выполнена, задаемся уровнем значимости α и сравниваем значение статистики с табличными значениями β_1 , β_2 , где $F(\beta_1) = \alpha$, $F(\beta_2) = 1 - \alpha$.

В частности, для предельной функции распределения статистики D_n ($H > 0,5$)

$$F(x) = 2\Phi\left(\sqrt{x\frac{2}{3}}\right) - 1, \text{ где } \Phi \text{ — функция Лапласа.}$$

Соответствующие уровню значимости $\alpha = 0,05$ критические значения $\beta_1 \approx 0$, $\beta_2 = 6$ и гипотеза T принимается, если $0 < D_n < 6$.

Рассмотрим задачу прогноза фрактального броуновского движения. Пусть априорно известно, что наблюдаемые m значений временного ряда имеют вид

$$B_H(s), B_H(2s), \dots, B_H(ms),$$

т.е. являются реализацией фрактального броуновского движения. Вследствие гауссовости оптимальный r -шаговый прогноз (условное среднее) линеен и строится следующим образом.

Представим корреляционную матрицу Q в виде:

$$Q = \frac{s^{2H}}{2} \begin{pmatrix} A & B \\ C & D \end{pmatrix};$$

$$q_{jk} = EB_H(js)B_H(ks) = \frac{s^{2H}}{2}(j^{2H} + k^{2H} - |j-k|^{2H}), 1 \leq j, k \leq m+r,$$

где индексы матрицы A удовлетворяют неравенству $1 \leq j, k \leq m$. Оценкой оптимального прогноза являются значения

$$\hat{B}_H((m+1)s), \dots, \hat{B}_H((m+r)s),$$

определенные равенством

$$\begin{pmatrix} \hat{B}_H((m+1)s) \\ \vdots \\ \hat{B}_H((m+r)s) \end{pmatrix} = CA^{-1} \begin{pmatrix} B_H(s) \\ \vdots \\ B_H(ms) \end{pmatrix}.$$

Погрешность δ оптимального прогноза вычисляется по формуле

$$\delta = \frac{s^{2H}}{2} \text{tr}(D - CA^{-1}B).$$

Таким образом, для конструкции оптимального прогноза необходим параметр Харста H , для определения погрешности δ — дисперсия s^{2H} . Приведенные в работе результаты могут быть применены для оценивания параметров стохастических уравнений с последствием ([9–11]).

Модель, использующая фрактальное броуновское движение в качестве базового процесса, применена для анализа реальных медико-биологических временных данных $x_k = x(ks)$, $k = 1, 2, \dots, 204$. Представим наблюдаемую траекторию $x(t)$ в виде «сигнал + шум»:

$$x(t) = m(t) + y(t) \Leftrightarrow x_k = m_k + y_k,$$

где $m(t)$ — тренд, полученный полиномиальной аппроксимацией, компоненты шума $\{y_k\}$ удовлетворяют соотношению $(1/n)\sum_{k=1}^n y_k \approx 0$. Значения оценки $\hat{\rho}_1$ одношагового коэффициента корреляции для приращений $z_k = y_k - y_{k-1}$, вычисленные в шести временных окнах:

$$\hat{\rho}_1 = \{0,26 - 0,34\},$$

что позволяет считать последовательность приращений $\{z_k\}$ стационарной. Полагая $\rho_1 = 0,3$, из равенства $\rho_1 = 2^{2H-1} - 1$ находим $\hat{H} = 0,69$. Тестирование данных с помощью статистик

$$D_n = \frac{1}{n^{2H}} \frac{1}{s^{4H}} \sum_{k=1}^n y_{k-1} z_k^3, \quad F_n = \frac{1}{n^H} \frac{1}{s^{3H}} \sum_{k=1}^n z_k^3$$

подтверждает гипотезу: «последовательность $\{y_k\}$ является реализацией фрактального броуновского движения» (персистентный случай).

Анализ данных $\{x_k\}$ проводился двумя способами:

- 1) с использованием модели *ARIMA*;
- 2) в виде композиции «сигнал+шум», т.е. значения второй модели формировались в виде $\hat{x}_j = \hat{m}_j + \hat{y}_j$, где \hat{m}_j вычислялись по *ARIMA*, а для шума применялась формула

$$\begin{pmatrix} \hat{y}_{m+1} \\ \vdots \\ \hat{y}_{m+r} \end{pmatrix} = CA^{-1} \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}.$$

В обоих случаях критерием качества модели выбрано нормированное среднеквадратичное отклонение

$$\mu = \left(\frac{\sum_{j=m+1}^{m+r} (x_j - \hat{x}_j)^2}{\sum_{j=m-1}^{m+r} x_j^2} \right)^{1/2},$$

где $r = 10$, во второй модели объем обучающей выборки составил $m = \{50, 120, 180\}$. Результат вычислений: $\frac{\mu_2}{\mu_1} = 0,88$, т.е. по данному критерию модель, использующая фрактальное движение, дает выигрыш 12%.

ЛИТЕРАТУРА

1. Y. Mishura, “Stochastic Calculus for Fractional Brownian Motion and Related Processes. Lecture Notes in Mathematics 1929”, *Springer*, 393 p., 2008. doi: 10.1007/978-3-540-75873-0.
2. F. Biagini, Y. Hu, B. Øksendal, and T. Zhang, “Stochastic Calculus for Fractional Brownian Motion and Applications”, *Springer*, 329 p., 2013. doi: 10.1007/978-1-84628-797-8.
3. R.F. Peltier and J. Levy Vehel, “A new method for estimating the parameter of fractional Brownian motion”, *Rapport de recherche de l'INRIA*, no. 2396, 27 p., 1994.
4. I. Nourdin, “Asymptotic behavior of weighted quadratic and cubic variations of fractional Brownian motion”, *Ann. Probab.*, 36, no. 6, pp. 2159–2175, 2008. doi: 10.1214/07-AOP38.
5. I. Nourdin, “Noncentral convergence of multiple integrals”, *Ann. Probab.*, vol. 37, no. 4, pp. 1412–1426, 2009. doi: 10.1214/08-AOP435.
6. M. Gradinaru and I. Nourdin, “Milstein's type schemes for fractional SDEs”, *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 45, no. 4, pp. 1085–1098, 2009. doi: 10.1214/08-AHP196.

7. I. Nourdin, D. Nualart, and C. Tudor, “Central and non-central limit theorems for weighted power variations of fractional Brownian motion”, *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 46, no. 4, pp. 1055–1079, 2010. doi: 10.1214/09-AIHP342.
8. I. Nourdin, “Selected Aspects of fractional Brownian motion”, *Springer*, 124 p., 2012. doi: 10.1007/978-88-470-2823-4.
9. K. Kubilius, Yu. Mishura, and K. Ralchenko, “Parameter Estimation in Fractional Diffusion Models”, *Bocconi & Springer Series*, 380 p., 2017. doi: 10.1007/978-3-319-71030-3.
10. Y. Mishura, K. Ralchenko, and G. Shevchenko, “Existence and uniqueness of a mild solution to the stochastic heat equation with white and fractional noises”, *Theor. Probability and Math. Statist.*, 98(2019), pp. 149–170. Available: <https://doi.org/10.1090/tpms/1068>
11. O. Banna, Yu. Mishura, K. Ralchenko, and S. Shklyar, *Fractional Brownian Motion. Approximations and Projections*. Wiley-ISTE, 2019. 288 p. doi: 10.1002/9781119476771.

Поступила 29.01.2021

INFORMATION ON THE ARTICLE

Viktor G. Bondarenko, ORCID: 0000-0003-1663-4799, Educational and Scientific Complex “Institute for Applied System Analysis” of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: bondarenavg@gmail.com

ON SOME STATISTICS OF FRACTIONAL BROWNIAN MOTION / V.G. Bondarenko

Abstract. Fractional Brownian motion as a method for estimating the parameters of a stochastic process by variance and one-step increment covariance is proposed and substantiated. The root-mean-square consistency of the constructed estimates has been proven. The obtained results complement and generalize the consequences of limit theorems for fractional Brownian motion, that have been proved in the number of articles. The necessity to estimate the variance is caused by the absence of a base unit of time and the estimation of the covariance allows one to determine the Hurst exponent. The established results let the known limit theorems to be used to construct goodness-of-fit criteria for the hypothesis “the observed time series is a transformation of fractional Brownian motion” and to estimate the error of optimal forecasting for time series.

Keywords: fractional Brownian motion, parameter estimation, statistical hypothesis testing.

ПРО ДЕЯКІ СТАТИСТИКИ ФРАКТАЛЬНОГО БРОУНІВСЬКОГО РУХУ / В.Г. Бондаренко

Анотація. Запропоновано й обґрунтовано метод оцінювання параметрів стохастичного процесу — фрактального броунівського руху — дисперсії та однокрокової коваріації приростів. Доведено середньоквадратичну консистентність побудованих оцінок. Отримані результати доповнюють та узагальнюють наслідки з граничних теорем для фрактального броунівського руху, які доведено в ряді праць. Необхідність оцінювання дисперсії зумовлено відсутністю еталонної одиниці часу, а оцінка коваріації дозволяє визначити показник Харста. Установлені результати дозволяють використовувати відомі граничні теореми в побудові критеріїв згоди для гіпотези «спостережуваний часовий ряд є реалізацією фрактального броунівського руху» та оцінити похибку оптимального прогнозу часового ряду.

Ключові слова: фрактальний броунівський рух, оцінювання параметрів, перевірка статистичних гіпотез.

QUINTILE REGRESSION BASED APPROACH FOR DYNAMICAL VAR AND CVAR FORECASTING USING METALOG DISTRIBUTION

G. ZRAZHEVSKY, V. ZRAZHEVSKA

Abstract. The paper proposes a new method of dynamic VaR and CVaR (ES) risk measures forecasting. Quantile linear GARCH model is chosen as the main forecasting model for time series quantiles. To build a forecast, the values of quantiles are approximated by the metalog distribution, which makes it possible to use analytical formulas to evaluate risk measures. The method of VaR and CVaR forecasting is formulated as a step-by-step algorithm. At the first stage, an initial model is built to obtain variance estimates. The predicted variance values obtained from the constructed model are used at the second stage to find the QLGARCH model coefficients by solving the minimization problem. At the third stage, the QLGARCH models are estimated on a non uniform quantile grid. The obtained predicted values of quantiles are used to estimate the approximating metalog distribution. The investigated theory is applied to VaR and CVaR forecasting for time series of daily log return of the DJI index.

Keywords: VaR, CVaR, Expected Shortfall, dynamic risk measures, forecast, Quantile LGARCH model, metalog distribution.

INTRODUCTION

The purpose of this study is to develop the new method of dynamic VaR and CVaR risk measures estimation and forecasting. VaR and CVaR are classic measures that are used in financial risk assessment [1]. In the practice of VaR and CVaR estimating for a random variable that describes the profitability of a financial instrument, two main approaches can be distinguished. The first approach is a nonparametric estimation method that is based on an empirical distribution function. The disadvantage of this estimation method is the critical dependence of the effectiveness of the method on the presence in the initial data of values that arise with low probabilities [1]. The second approach is parametric, based on a priori estimation of the distribution function, which is the main disadvantage of this approach [1].

In time series analysis, in particular, in time series forecasting, in addition to static risks measures, in practice it becomes necessary to build more complex risk models, which take into account the changes of the series over time. In this case, to estimate risk measures, various time series models can be used, such as, for example, ARMA, GARCH. With this approach, the problem of risk measures modeling is reduced to the estimation of a model for variance and finding static risk measures for its residuals using parametric or nonparametric methods. Examples of this approach are described in [2–4].

The described above approaches of evaluation of risk measures are based on the construction of a cumulative distribution function on the full space of events.

At the same time, to estimate risk measures, it is sufficient to evaluate only the quantile of a given level (for VaR) or the distribution of values exceeding a given level (for CVaR). At the same time, from a practical point of view, the most significant is the distribution for a relatively small subset of the event space leading to extreme consequences. Accordingly, it is possible to simplify the forecasting task by using the quantile regression model proposed in [5, 6] contains a detailed description of the theory of quantile regression estimation applicable to standard time series models. Since financial time series, as a rule, are characterized by rather strong volatility, quantile GARCH models are popular for risk analysis. The problem of building quantile models of the GARCH class and their application in VaR forecasting for the series of log returns of stock market indices is considered, for example, in [7, 8].

One of the possible approaches for fitting of the GARCH model residuals distribution is to use the metalog distribution proposed in [9]. This choice is based on the simplicity of quantile formulas and the availability of a sufficient set of parameters of this distribution for an adequate fitting of empirical data of various nature. Thus, in [10], the metalog distribution is used in the development of the extended FAIR-BN combined approach for cyber security risk assessment. In [11], the five-term metalog distribution is used to forecast fertility rates in Canada. SPT (symmetric-percentile triplet) metalog distribution is used in [12] to statistically compare the forecasts of annual production in the oil and gas industry in Norway. In [13], the metalog distribution is used for dynamic risk measures VaR and CVaR estimating based on a heteroscedastic time series model, taking into account the strong dependence of the data series. The method of smoothing of the autocorrelation function is used for variance modeling. A metalog distribution is proposed to use for risk measures model residuals estimating. The paper proposes two methods of metalog distribution estimating and explicit analytical formulas for VaR and CVaR modeling and forecasting with different numbers of members in the metalog distribution.

A large number of publications devoted to the risk measures estimation and forecasting testifies to the applied significance of this problem. At the same time, the task of developing the new methods and approaches for risk modeling, which more fully reflect the nature of the modeled series, remains relevant. Most of the forecasting methods are based on the estimation of the entire distribution function. On the one hand, this is an overstated requirement for the model, and on the other hand, it often leads to an incorrect description of tails of distribution. Therefore, in this paper, it is proposed to build volatile models only for the tail parts of the distribution. In this case, the obtained point values of the quantiles can be smoothed, for example, by metalog distribution.

MATERIALS AND METHODS

On the probability space (Ω, Φ_t, P) a time series $\{u_t, t \in T\}$ with a finite mean is considered (Φ_t is the information set containing all available at the time t information about the time series).

For a fixed confidence level α risk measure VaR_α^t is defined as the conditional α — quantile of the CDF of u_t : $VaR_\alpha^t = F_t^{-1}(\alpha)$. The risk measure

$CVaR_\alpha^t$ is defined as the integral: $CVaR_\alpha^t = E_t[u_\gamma | u_\gamma < -VaR_\alpha^t] = -\frac{1}{\alpha} \int_0^\alpha VaR_\gamma^t d\gamma$,

($\alpha < 0.5$), where $E_t[\cdot]$ denotes expectation with respect to Φ_t . In this paper, the continuity of the CDF is assumed.

As indicated in the introduction, most methods for dynamic VaR and CVaR risk measures forecasting are based on time series modeling. The GARCH models are among the models that describe volatility of financial time series. In this paper, we consider the Linear GARCH model LGARCH (p, q). This model is frequently used for fitting log return volatility time series and appropriate for quantile regression because of its linear structure [7].

The time series $\{u_t, t = 0, 1, 2, \dots\}$ follows LGARCH (p, q) process if:

$$u_t = \sigma_t \varepsilon_t, \quad \sigma_t = \beta_0 + \sum_{i=1}^q \gamma_i |u_{t-i}| + \sum_{j=1}^p \beta_j \sigma_{t-j}, \quad (1)$$

where $\{\varepsilon_t\}$ are independent, identically distributed random variables with zero mean and a conditional distribution function $F_\varepsilon(\cdot)$, $\beta_0 > 0$; $(\gamma_1, \gamma_2, \dots, \gamma_q)^T \in R_+^q$.

Using the heteroscedastic time series model, the dynamic risk measures can be found under the following formulas [1]:

$$VaR_\alpha^t = VaR_\alpha(\varepsilon) \sigma_t, \quad CVaR_\alpha^t = CVaR_\alpha(\varepsilon) \sigma_t, \quad (2)$$

where the model for σ_t is defined in (1), $VaR_\alpha(\varepsilon)$ and $CVaR_\alpha(\varepsilon)$ are static risk measures at time t . Then the P step forecast for dynamic risk measures can be found by model (2) extrapolation:

$$VaR_\alpha^{t+P} = VaR_\alpha(\varepsilon) \sigma_{t+P}, \quad CVaR_\alpha^{t+P} = CVaR_\alpha(\varepsilon) \sigma_{t+P}. \quad (3)$$

In this work, the following methods are used to evaluate static risk measures $VaR_\alpha(\varepsilon)$, $CVaR_\alpha(\varepsilon)$.

Historical simulation method. Let X be a random variable and its sample values are X_1, X_2, \dots, X_N . In accordance with the historical simulation method, an empirical distribution function is constructed on the sample values. Then according to [14]:

$$VaR_\alpha = -X_{([N\alpha])}, \quad CVaR_\alpha = -\left(\frac{\sum_{i=1}^{[N\alpha]} X_{(i)}}{[N\alpha]} \right), \quad (4)$$

where $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(N)}$.

Using Student's t -distribution. If the random variable has the local scale Student's t -distribution with the parameters μ , σ and the degrees of freedom $\nu > 2$, then the risk measures can be calculated as (see [14]):

$$VaR_\alpha = \mu + \sigma t_\nu^{-1}(\alpha), \quad CVaR_\alpha = \mu - \sigma \frac{g_\nu(t_\nu^{-1}(\alpha)) \cdot \nu + (t_\nu^{-1}(\alpha))^2}{\alpha \cdot (\nu - 1)}, \quad (5)$$

where $g_\nu(\cdot)$ is the standard PDF and $t_\nu^{-1}(\cdot)$ is the inverse standard CDF value at α of t -distribution.

Using metalog distribution. Suppose that X has a metalog distribution $F_X(x)$, that is defined by a quantile function $M_n(\alpha, a(X, \alpha))$ [9]:

$$M_n(\alpha, \mathbf{a}) = \begin{cases} a_1 + a_2 \ln \frac{\alpha}{1 + \alpha}, & n = 2; \\ a_1 + a_2 \ln \frac{\alpha}{1 + \alpha} + a_3(\alpha - 0,5) \ln \frac{\alpha}{1 + \alpha}, & n = 3; \\ a_1 + a_2 \ln \frac{\alpha}{1 + \alpha} + a_3(\alpha - 0,5) \ln \frac{\alpha}{1 + \alpha} + a_4(\alpha - 0,5), & n = 4; \\ M_{n-1} + a_n(\alpha - 0,5)^{\frac{n-1}{2}}, & \text{for odd } n \geq 5; \\ M_{n-1} + a_n(\alpha - 0,5)^{\frac{n-1}{2}} \ln \frac{\alpha}{1 + \alpha}, & \text{for even } n \geq 5. \end{cases}$$

The coefficients $\mathbf{a} = (a_1, a_2, \dots, a_n)^T$ can be found as a solution of the system of equations:

$$\mathbf{a} = [\mathbf{Y}_n^T \mathbf{Y}_n]^{-1} \mathbf{Y}_n^T \mathbf{X}, \tag{6}$$

where $\mathbf{X} = (X_1, X_2, \dots, X_N)^T$, the matrix \mathbf{Y}_n is defined through a set of quantiles with levels $\mathbf{a} = (\alpha_1, \alpha_2, \dots, \alpha_N)^T = (F_X(X_1), F_X(X_2), \dots, F_X(X_N))^T$ [9].

Following the definitions, risk measures can be found under the formulas:

$$VaR_{\tilde{\alpha}}(X) = -M_n(\tilde{\alpha}, \mathbf{a}(X)); \tag{7}$$

$$CVaR_{\tilde{\alpha},n}(X) = -\frac{1}{\tilde{\alpha}} \int_{\tilde{\alpha}_0}^{\tilde{\alpha}} M_n(y, \mathbf{a}(X)) dy.$$

Explicit formulas for $CVaR_{\tilde{\alpha},n}(X)$ estimating with different number of members of the metalog distribution were obtained in [13]:

$$CVaR_{\tilde{\alpha},n} = \begin{cases} a_1 + a_2 \left(\frac{\ln(1 - \tilde{\alpha})}{\tilde{\alpha}} + \ln \frac{\tilde{\alpha}}{1 - \tilde{\alpha}} \right), & n = 2; \\ CVaR_{\tilde{\alpha},2}(X) + \frac{a_3}{2} \left(1 + (\tilde{\alpha} - 1) \ln \frac{\tilde{\alpha}}{1 - \tilde{\alpha}} \right), & n = 3; \\ CVaR_{\tilde{\alpha},3}(X) + \frac{a_4(\tilde{\alpha} - 1)}{2}, & n = 4; \\ CVaR_{\tilde{\alpha},n-1}(X) + \frac{2a_n}{\tilde{\alpha}(1+n)} \left((-1)^{\frac{n-1}{2}} (0,5)^{\frac{n+1}{2}} + (\tilde{\alpha} - 0,5)^{\frac{n+1}{2}} \right), & \text{for odd } n \geq 5; \\ CVaR_{\tilde{\alpha},n-1}(X) + \frac{2-n}{2} \frac{a_n}{\tilde{\alpha}} 2(2\tilde{\alpha} - 1)^{\frac{n}{2}} \arctg(2\tilde{\alpha} - 1) + nG + \\ + \ln(1 - \tilde{\alpha}) + (-1)^{\frac{n+2}{2}} \ln \tilde{\alpha}, & \text{for even } n \geq 6. \end{cases} \tag{8}$$

where

$$G = {}_3F_2[1, 1, 1 - n/2; 2, 2; 2] + (\tilde{\alpha} - 1) {}_3F_2[1, 1, 1 - n/2; 2, 2; 2 - 2\tilde{\alpha}] + (-1)^{n/2} \tilde{\alpha} {}_3F_2[1, 1, 1 - n/2; 2, 2; 2\tilde{\alpha}],$$

${}_3F_2[c_1, c_2, c_3; d_1, d_2; z]$ is generalized hypergeometric function.

For determining the parameters of the metalog distribution the Quintile Metalog Method is proposed in [13]. This method uses the approximation of the empirical distribution function by sample quantiles $\mathbf{X} = (\tau_{\alpha_1}, \dots, \tau_{\alpha_N})^T$, where

$$\tau_{\alpha_i} = \begin{cases} X_{([n\alpha_i]+1)}, & \text{if } n\alpha_i \text{ is not integer;} \\ \frac{\bar{\alpha} - \alpha_i}{\bar{\alpha} - \alpha} X_{(\alpha)} + \frac{\alpha_i - \alpha}{\bar{\alpha} - \alpha} X_{(\bar{\alpha})}, & \text{if } n\alpha_i \text{ is integer;} \end{cases}$$

$$\alpha = \frac{[n\alpha_i]}{n}, \bar{\alpha} = \frac{[n\alpha_i + 1]}{n}, \alpha_i \in (0, 1).$$

Thus, the classical approach to risk measures forecasting (3) involves building a model for variance estimating, obtaining model residuals, and using methods for static risk measures $VaR_\alpha(\varepsilon)$, $CVaR_\alpha(\varepsilon)$ estimating.

Another approach is based on estimating of quantile time series models, which makes it possible to directly simulate the time series quantile of a given level. In this paper, in accordance with (1), the QLGARCH model is considered (see [6–8]):

$$Q_{u_t}(\tau | \Phi_{t-1}) = \theta_t(\tau)^T \mathbf{z}_t, \tag{9}$$

where $Q_{u_t}(\tau | \Phi_{t-1})$ is a conditional τ — quantile for $\{u\}$, $\mathbf{z}_t = (1, \sigma_{t-1}, \dots, \sigma_{t-p}, |u_{t-1}|, \dots, |u_{t-q}|)^T$, $\theta_t(\tau)^T = (\beta_0, \beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_q) F^{-1}(\tau) = (\beta_0(\tau), \beta_1(\tau), \dots, \beta_p(\tau), \gamma_1(\tau), \dots, \gamma_q(\tau))$.

The paper proposes the following methodology for dynamic VaR and CVaR forecasting. Model (9) is used to construct a set of quantiles. For a more detailed description of the left tail of the CDF, the quantile levels can be found on a non uniform grid. Assuming that the obtained set of quantiles can be fitted by the metalog distribution, the quantile function $M_n(\alpha, \mathbf{a}(X, \alpha))$ is estimated using (6). To obtain forecasts of risk measures, formulas (7), (8) are used. The practical implementation of the method is formulated as a step-by-step algorithm.

AN ALGORITHM FOR CONSTRUCTING THE DYNAMIC RISK MEASURES VAR AND CVAR FORECAST BASED ON THE METHOD QLGARCH - METALOG

1. Building a variance model to get estimates. The model LGARCH(p, q)

(1) is written as ARCH(∞): $\sigma_t = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j |u_{t-j}|$, where the coefficients α_j

satisfy summability conditions implied by the regularity conditions [6]. Due to the assumption of regularity, the coefficients decrease geometrically, therefore, the model can be reduced to ARCH (m). The appropriate lag for reduced ARCH model is chosen on the base of significant values of ACF and PACF for squared values of returns (values that is more than confidence bounds). Estimates of α_j can be obtained in various ways. QMLE is used in this work. The fitted model is used to obtain estimates $\hat{\sigma}_t, \dots, \hat{\sigma}_{t-p}$:

$$\hat{\sigma}_{t-i} = \hat{\alpha}_0 + \sum_{j=1}^m \hat{\alpha}_j |u_{t-j-i}|, \quad i = \overline{0, p}. \quad (10)$$

2. Building a set of quantiles predictive estimates. To obtain estimates of the τ -quantile for u_t , the QLGARCH model (9) is used. The orders of the model p, q can be estimated using Akaike Inform Criteria (AIC) and Hannan-Quinn Inform Criteria (HQIC). It is also possible to use Bayes Inform Criteria (BIC) and Shibata Inform Criteria (SIC). To estimate the vector of parameters $\theta(\tau)^T$, the minimization problem is solved using the quantile regression estimation in the form (see [8]):

$$\min_{\theta} \sum_t \rho_{\tau}(u_t - \theta((\tau)^T \hat{z}_t)), \quad (11)$$

where $\rho_{\tau}(u) = u(\tau - I(u < 0))$ is a check function, $\hat{z}_t = (1, \hat{\sigma}_{t-1}, \dots, \hat{\sigma}_{t-p}, |u_{t-1}|, \dots, |u_{t-q}|)^T$ taking into account that the estimates $\hat{\sigma}_{t-1}, \dots, \hat{\sigma}_{t-p}$ were obtained at step 1 (10). The solution of the unconstrained minimization problem (11) makes it possible to estimate the τ -quantile for u_t in the form:

$$\hat{Q}_{u_t}(\tau | \Phi_{t-1}) = \hat{\theta}(\tau)^T \hat{z}_t.$$

At this step, a grid of τ_i -quantiles, $i = \overline{1, N}$, is constructed and the problem (11) is solved N times. For every quantile regression with $\tau_i, i = \overline{1, N}$ the quantile estimates are $\hat{Q}_{u_t}(\tau_i) = \hat{\theta}_t(\tau_i)^T \hat{z}_t, i = \overline{1, N}$. In this case, the predicted values of the conditional quantiles are calculated by extrapolation $\hat{Q}_{u_{t+1}}(\tau_i) = \hat{\theta}(\tau_i)^T \hat{z}_{t+1}$, where $\hat{z}_{t+1} = (1, \hat{\sigma}_t, \dots, \hat{\sigma}_{t-p+1}, |u_t|, \dots, |u_{t-q+1}|)^T$, and $\hat{\sigma}_{t-i}, i = \overline{0, p-1}$ are obtained at the first step of algorithm (10).

3. Risk measures forecasting. The predictive quantiles $\hat{Q}_{u_{t+1}}(\tau_1), \dots, \hat{Q}_{u_{t+1}}(\tau_N)$ (from previous step) are fitted using the metalog distribution. Estimates of the metalog distribution parameters $\hat{a} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n)^T$ are found in accordance with (6), where $X = (\hat{Q}_{u_{t+1}}(\tau_1), \dots, \hat{Q}_{u_{t+1}}(\tau_N))^T, \alpha = (\tau_1, \tau_2, \dots, \tau_N)^T$. Specifying the quantile function $M_n(\tilde{\alpha}, \hat{a})$ for a given level $\tilde{\alpha}$ allows the use of analytical expressions (7), (8) to find the predicted values $VaR_{\tilde{\alpha}}^{t+1}$ and $CVaR_{\tilde{\alpha}}^{t+1}$.

EXPERIMENT, RESULTS AND DISCUSSIONS

Proposed algorithm was applied for dynamic risk measures $VaR_{0,1}$, $CVaR_{0,1}$ forecasting for the time series of daily log returns of the Dow Jones Industrial Average index (the DJI time series). The sample length was 3500 values from 2007/02/16 to 2021/01/11. The forecast model was based on 1500 historical values and was extrapolated one value forward. After that, the modeling window was one step shifted, and the model was rebuilt. The procedure was repeated 2000 times (Rolling Forecast Method). The obtained one-step forecasts were compared with real values for the corresponding period of time. To obtain variance estimates (10), the model ARCH(30) was built. Based on the historical values of the time series and the estimated variance values, quantile LGARCH(p, q) models were built for different levels of quantiles. The orders of the models were found using the AIC and HQIC criteria: $p = 3$, $q = 3$. For more detailed description of the left tail of the distribution, the non uniform grid of quantiles was used: $\tau_i = ih$, $h = 0,01$ for $i = \overline{1,20}$, and $h = 0,05$ for $i = \overline{5,19}$. Estimates of the parameters were obtained using QMLE. The estimated models were used to generate one-step quantile predictions. The predicted quantile values were used to estimate the metalog distribution. Using (6), the estimates for the coefficients of the metalog distributions for $n = 4, 5, 6, 7$ were obtained. Risk measures estimates were calculated using formulas (7) and (8). The results were obtained using *R* packages *rugarch* [15] and *quantreg* [16].

The results of dynamic VaR and $CVaR$ risk measures forecasting using the QLGARCH – Metalog (rq_met) method are shown in Fig. 1 along with the historical values (the first 1500) and the real values (1501–3500) of the DJI time series. As can be seen from the graph, the obtained forecast estimates describe the dynamic behavior of the time series quite well.

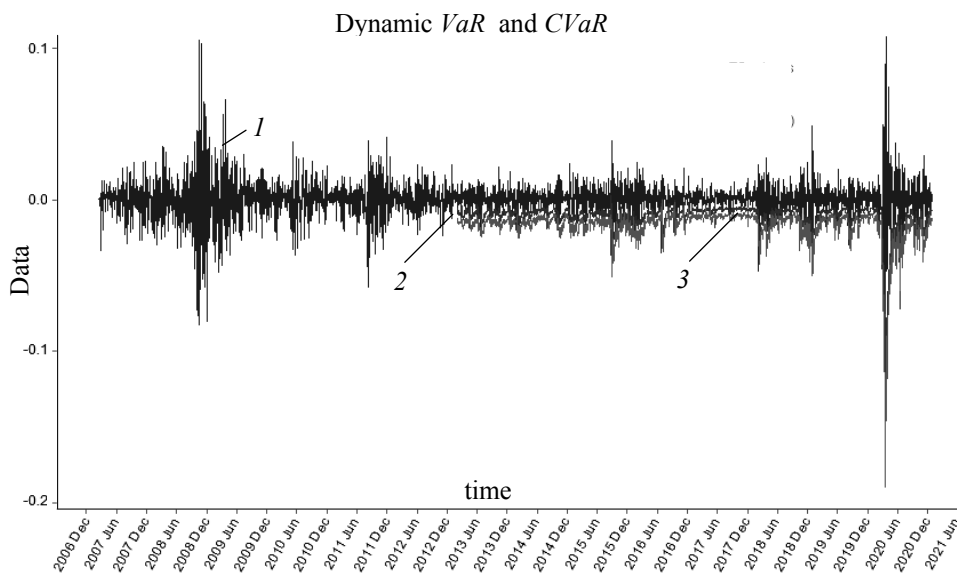


Fig. 1. Historical data of the time series of daily log return of the DJI index (1 — TS returns) from 2007/02/16 to 2021/01/11 and the forecast values for $VaR_{0,1}^{t+1}$ (2 — $VaR(0,1)$) and $CVaR_{0,1}^{t+1}$ (3 — $CVaR(0,1)$) obtained by the QLGARCH – Metalog method ($n = 5$) for the period 2013/02/01 – 2021/01/11

Historical data and the forecast values of dynamic risk measures over a short period of time is shown in Fig. 2 for more convenient visual analysis.

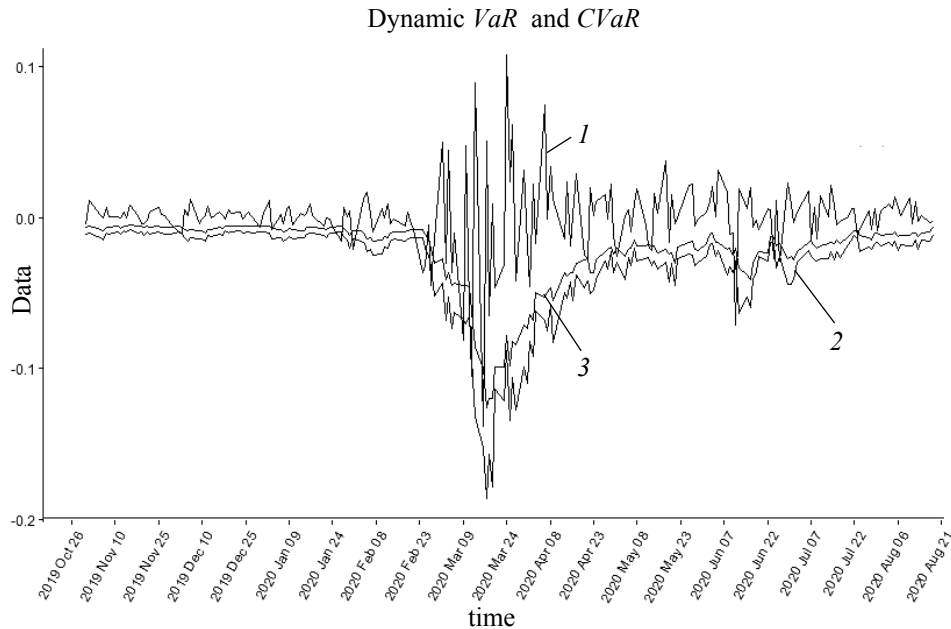


Fig. 2. Historical data of the time series of daily log return of the DJI index (I — TS returns) and the forecast values for for $VaR_{0,1}^{t+1}$ (2 — $VaR(0,1)$) and $CVaR_{0,1}^{t+1}$ (3 — $CVaR(0,1)$) obtained by the QLGARCH – Metalog method ($n = 5$) for the period 2019/10/31 – 2020/08/18

For a comparative analysis of the effectiveness of the proposed method, a forecast of dynamic risk measures for a given time series was built with standard approach on the basis of a heteroscedastic model (3). To estimate the variance, the LGARCH(3,3) model was considered. The AIC and HQIC criteria were used to determine the orders of the model. The QMLE was used to estimate the model coefficients. To determine the risk measures for the residuals of the LGARCH(3,3) model, the following methods were used: the Historical simulation method (4) (hist method), explicit formulas (5) under the assumption that the model residuals have the local scale t -distribution (tLS method), explicit formulas (7), (8) based on Quantile Metalog Method for $n = 4, 5, 6, 7$ (metal method).

The analysis of the constructed forecast estimates was carried out using the backtesting procedure. The following tests were used in the work:

- for VaR estimates: the Kupiec test (LRuc), Christoffersen's independence test (LRind), PoE statistics [17];
- for CVaR estimates: two tests proposed in [18]: one-sided simple conditional calibration test (scc_1) and two-sided simple conditional calibration test (scc_2); three regression based backtests proposed in [19]: the auxiliary ESR backtest (Aux), the strict ESR backtest (Str), the intercept ESR backtest (Int).

The p -values of these tests are shown in Table 1. Table 2 shows the PoE statistic values for the forecast estimates of dynamic risk measure VaR .

Table 1. The results of the qualitative analysis of the forecast estimates of dynamic risk measures *VaR* and *CVaR* for DJI time series

Method	Risk measure						
	<i>VaR</i>		<i>CVaR</i>				
	LRuc	LRind	scc_1	scc_2	Aux	Str	Int
hist	0,0677	0,0091	0,0866	0,0076	0,6453	0,6191	0,5824
tLS	0,4582	0,3359	0,0657	0,1136	0,5364	0,5286	0,4710
metal, $n = 4$	0,1497	0,0802	0,2301	0,0625	0,6980	0,7289	0,6267
metal, $n = 5$	0,1726	0,1497	0,2411	0,0630	0,6883	0,7338	0,6461
metal, $n = 6$	0,0945	0,0393	0,1251	0,0115	0,5707	0,6047	0,5355
metal, $n = 7$	0,0945	0,0474	0,1245	0,0098	0,5261	0,5521	0,5180
rq_met, $n = 4$	0,4983	0,7643	0,1572	0,0869	0,8454	0,8914	0,8072
rq_met, $n = 5$	0,8810	0,5796	0,2741	0,1808	0,8842	0,9164	0,8475
rq_met, $n = 6$	0,4272	0,4072	0,1863	0,1082	0,8566	0,8931	0,8121
rq_met, $n = 7$	0,5476	0,3655	0,2688	0,1757	0,8671	0,8940	0,8450

Table 2. PoE statistic values for the forecast estimates of dynamic risk measure *VaR* for DJI time series

Method	hist	tLS	metal				rq_met			
			$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 4$	$n = 5$	$n = 6$	$n = 7$
PoE	0,0879	0,1051	0,09045	0,09095	0,0889	0,0889	0,0954	0,0989	0,0944	0,0959

As follows from Table 1, the worst estimates were obtained using the historical simulation method (hist). In particular, as a result of applying the Christoffersen's independence test (LRind) for *VaR* and the two-sided simple conditional calibration test (scc_2) for *CVaR*, hypotheses with a significance level of 0,05 were rejected. This indicates the inapplicability of the historical simulation method for predicting the values of the DJI time series. At the same time, all tests showed consistently good quality of forecasts obtained by the QLGARCH — Metalog method (rq_met) and maximum *p*-value statistics compared to other methods (tLS and metal).

In the article the metalog distribution with the different number of parameters ($n = 4, 5, 6, 7$) was considered (see Table 1). An increase in the number of parameters potentially increases the accuracy of the estimates, but can lead to the problem of overfitting. As a result of the backtesting for forecasted dynamic risk measures *VaR* and *CVaR* obtained with the Quantile Metalog Method (metal), the choose of large *n* probably leads to overfitting. The estimates of *VaR* and *CVaR* obtained using the metalog distribution for $n = 4$ and $n = 5$ are consistently better than the estimates obtained using the same sample for $n = 6$ and $n = 7$. At the same time, the QLGARCH – Metalog (rq_met) method shows less dependence on the number of parameters of the metalog distribution. The results of the qualitative analysis for *VaR* and *CVaR* forecasts obtained by this method are relatively uniform for all *n*. Although it should be noted that according to the results of all tests, the highest quality forecasts for risk measures were obtained at $n = 5$. The best results of *VaR* forecasting (see Table 2) according to PoE statistics is obtained also by rq_met method for $n = 5$ (has the least deviation from the target value of 0,1).

The paper proposes a method that is a natural continuation of existing research and methods devoted to dynamic risk models developing. It combines parametric and nonparametric statistical approaches to time series modeling. Practical application of the method shows its effectiveness in the case of risk modeling for highly volatile financial time series. The simplicity of the method and its background make it possible to recommend it for using in various fields. However, the determination of the restrictions on applicability of this method, as well as the automation of the procedure for estimation of its parameters, requires further mathematical research, in particular, the construction of asymptotic estimates of the convergence of the model.

CONCLUSIONS

The paper considers the problem of dynamic VaR and $CVaR$ risk measures modeling and forecasting for financial time series. Since the VaR measure is a conditional quantile of the distribution function of a given level, and $CVaR$ for continuous distributions can be specified as the average of the quantile function it is proposed to use QLGARCH as a model for risk measures forecasting. The advantage of using of this model is the ability to estimate and predict not the full distribution, but the values of the quantiles of the required levels. Since the risk measures are determined for the tail part of the distribution, a non uniform grid is used in the work, which makes it possible to detail the quantiles with a low level. To smooth point values, it is proposed to fit a set of quantiles with metalog distribution. This approach is also convenient due to the presence of explicit analytical expressions VaR and $CVaR$ for the metalog distribution. The proposed method for dynamic VaR and $CVaR$ risk measures forecasting is formulated in the form of a step-by-step algorithm.

The proposed methodology was tested on the time series of daily log return of the Dow Jones Industrial Average (DJI) index for the period from 2007-02-16 to 2021-01-11. Using the formulated algorithm, a set of one-step forecasts of risk measures was obtained. An analysis of the quality of the forecasts was carried out using various standard backtesting techniques on real data. The results were compared with the forecasts obtained by standard methods that are based on the LGARCH model and various assumptions about distribution of the residuals. The carried out qualitative analysis of the obtained predicted values showed the effectiveness of using the method proposed in this work and its advantage in comparison with standard methods.

The results of the work can be directly applied in dynamic risk modeling for highly volatile time series, in particular, financial time series, and also can serve as the basis for the development of new methods and algorithms for random processes prediction.

REFERENCES

1. McNeil R. Frey and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools*. New Jersey, USA: Princeton University Press, 2005.
2. N. Pankratova and N. Zrazhevskaya, "Method of dynamic var and cvar risk measures forecasting for long range dependent time series on the base of the heteroscedastic model", *Intelligent Control and Automation*, vol.8, pp. 126–138, Jan. 2017. Available: <https://doi.org/10.4236/ica.2017.82010>

3. H. Badaye and J. Narsoo, “Forecasting multivariate var and es using mc-garch-copula model”, *The Journal of Risk Finance*, vol. 21, no. 5, pp. 493–516, 2020. Available: <https://doi.org/10.1108/JRF-06-2019-0114>
4. S. Paul and P. Sharma, “Quantile forecasts using the realized garch-evt approach”, *Studies in Economics and Finance*, vol. 35, no. 4, pp. 481–504, Oct. 2018. Available: <https://doi.org/10.1108/SEF-09-2016-0236>
5. Z. Xiao and R. Koenker, “Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models”, *Journal of the American Statistical Association*, vol. 104, pp. 1696–1712, Dec. 2009. Available: <https://doi.org/10.1198/jasa.2009.tm09170>
6. R. Koenker, V. Chernozhukov, X. He, and L. Peng, *Handbook of Quantile Regression* [1st ed.]. Boca Raton, USA: CRC Press, 2017. Available: <https://doi.org/10.1201/9781315120256>
7. Y. Zheng, Z. Qianqian, G. Li, and Z. Xiao, “Hybrid quantile regression estimation for time series models with conditional heteroscedasticity”, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 80, issue 5, pp. 975–993, Nov. 2018. Available: <https://doi.org/10.1111/rssb.12277>
8. G. Wang, K. Zhu, G. Li, and W.K. Li, “Hybrid quantile estimation for asymmetric power garch models”, *Journal of Econometrics*, Aug. 2020. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0304407620302220>, Accessed on: 6 August 2020. Available: <https://doi.org/10.1016/j.jeconom.2020.05.005>
9. T. Keelin, “The metalog distributions”, *Decision Analysis*, vol. 13, issue 4, pp. 243–277, Dec. 2016. Available: <https://doi.org/10.1287/deca.2016.0338>
10. J. Wang, M. Neil, and N. Fenton, “A bayesian network approach for cybersecurity risk assessment implementing and extending the fair model”, *Computers & Security*, vol. 89, 2020. Available: <https://doi.org/10.1016/j.cose.2019.101659>
11. P. Dion, N. Galbraith, and E. Sirag, “Using Expert Elicitation to Build Long-Term Projection Assumptions”, in *Developments in Demographic Forecasting. The Springer Series on Demographic Methods and Population Analysis*, vol. 49, Springer, Cham, 2020, pp. 43–62. Available: https://doi.org/10.1007/978-3-030-42472-5_3
12. R. Bratvold, E. Mohus, D. Petutschnig, and J. Bickel, “Optimistic and Overconfident—Over and Over Again”, *SPE Reservoir Evaluation and Engineering*, vol. 23, May 2020. Available: <https://doi.org/10.2118/195914-PA>
13. V. Zrazhevskaja and G. Zrazhevskij, “Generalized approach for estimating and forecasting of dynamical var and cvar based on metalog distribution”, *Lecture Notes in Computational Intelligence and Decision Making*, in Babichev S., Lytvynenko V., Wójcik W., Vyshemyrskaya S. (eds), ISDMCI 2020. *Advances in Intelligent Systems and Computing*, vol. 1246. Springer, Cham. Available: https://doi.org/10.1007/978-3-030-54215-3_15
14. N.G. Zrazhevskaja and A.G. Zrazhevskij, “Classification of methods for risk measures var and cvar calculation and estimation”, *System research and information technologies*, vol. 3, pp. 118–125, Sep. 2016. Available: <https://doi.org/10.20535/SRIT.2308-8893.2016.3.11>
15. A. Ghalanos, *Introduction to the rugarch package (Version 1.4-3)*. Available: https://cran.r-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf [July 15, 2020]
16. R. Koenker, *Quantile regression in R: a vignette*. Available: <https://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf> [June 28, 2019].
17. G. Zrazhevskij, A. Golodnikov, S. Uryasev, and A. Zrazhevskij, “Application of buffered probability of exceedance in reliability optimization problems”, *Cybernetics and Systems Analysis*, vol. 56, pp. 1–9, May 2020. Available: <https://doi.org/10.1007/s10559-020-00263-4>
18. N. Nolde and J.F. Ziegel, “Elicibility and backtesting: Perspectives for banking regulation”, *The Annals of Applied Statistics*, vol. 11, no. 4, pp. 1833–1874, 2017. Available: <https://doi.org/10.1214/17-AOAS1041>

19. S. Bayer and T. Dimitriadis, "Regression-Based Expected Shortfall Backtesting", *Journal of Financial Econometrics*, September, 2020. Available: <https://academic.oup.com/jfec/advance-articlepdf/doi/10.1093/jjfinec/nbaa013/33796824/nbaa013.pdf>, Accessed on: 27 April 2020. Available: <https://doi.org/10.1093/jjfinec/nbaa013>

Received 20.03.2021

INFORMATION ON THE ARTICLE

Grigoriy M. Zrazhevsky, ORCID: 0000-0001-8475-2469, National Taras Shevchenko University of Kyiv, Ukraine, e-mail: zgrig@univ.kiev.ua

Vira F. Zrazhevskaya, ORCID: 0000-0001-5117-8093, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: vera.zrazhevskaya@gmail.com

ПРОГНОЗУВАННЯ ДИНАМІЧНИХ VAR І CVAR НА ОСНОВІ КВІНТИЛЬНОЇ РЕГРЕСІЇ З ВИКОРИСТАННЯМ МЕТАЛОГ РОЗПОДІЛУ / Г.М. Зражевський, В.Ф. Зражевська

Анотація. Запропоновано новий метод динамічного прогнозування мір ризику VaR і CVaR (ES). Як основна модель прогнозування для квантилів часових рядів обрано квантильну лінійну модель GARCH. Для побудови прогнозу значення квантилів апроксимуються розподілом металог, що дозволяє використовувати аналітичні формули для оцінювання мір ризику. Методику прогнозування VaR і CVaR сформульовано у вигляді покрокового алгоритму. На першому етапі будується вихідна модель для отримання оцінок дисперсії. Отримані за моделлю значення дисперсії використовуються на другому етапі для знаходження коефіцієнтів моделі QLGARCH шляхом розв'язання задачі мінімізації. На третьому етапі моделі QLGARCH оцінюються на неоднорідній квантильній сітці. Отримані прогнозні значення квантилів використовуються для оцінки параметрів розподілу металог. Розроблений метод застосовується до прогнозування VaR і CVaR для часового ряду логарифмічної доходності індексу DJI.

Ключові слова: VaR, CVaR, Expected Shortfall, динамічні міри ризику, прогноз, квантильна модель LGARCH, розподіл металог.

ПРОГНОЗИРОВАНИЕ ДИНАМИЧЕСКИХ VAR И CVAR НА ОСНОВЕ КВАНТИЛЬНОЙ РЕГРЕССИИ С ИСПОЛЬЗОВАНИЕМ МЕТАЛОГ РАСПРЕДЕЛЕНИЯ / Г.М. Зражевский, В.Ф. Зражевская

Аннотация. Предложен новый метод динамического прогнозирования мер риска VaR и CVaR (ES). Квантильная линейная модель GARCH выбрана в качестве основной модели прогнозирования для квантилей временных рядов. Для построения прогноза значения квантилей аппроксимируются распределением металог, что позволяет использовать аналитические формулы для оценки мер риска. Методика прогнозирования VaR и CVaR сформулирована в виде пошагового алгоритма. На первом этапе строится исходная модель для получения оценок дисперсии. Полученные по модели значения дисперсии используются на втором этапе для нахождения коэффициентов модели QLGARCH путем решения задачи минимизации. На третьем этапе модели QLGARCH оцениваются на неоднородной квантильной сетке. Полученные прогнозные значения квантилей используются для оценки параметров распределения металог. Разработанный метод применен к прогнозированию VaR и CVaR для временного ряда логарифмической доходности индекса DJI.

Ключевые слова: VaR, CVaR, Expected Shortfall, динамические меры риска, прогноз, квантильная модель LGARCH, распределение металог.

ВІДОМОСТІ ПРО АВТОРІВ

Белас Андрій Олегович,

аспірант кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Белас Олег Миколайович,

професор, доктор технічних наук, професор кафедри спеціальних телекомунікаційних систем Інституту спеціального зв'язку та захисту інформації КПІ ім. Ігоря Сікорського, Україна, Київ

Бідюк Петро Іванович,

професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Бондаренко Віктор Григорович,

професор, доктор фізико-математичних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Васильченко Іван Петрович,

професор, доктор технічних наук, професор кафедри інформаційних технологій Уманського національного університету садівництва, Україна, Умань

Волкова Віолетта Миколаївна,

професор, доктор економічних наук, професор Санкт-Петербурзького політехнічного університету Петра Великого, Росія, Санкт-Петербург

Гвоздецька Наталія Андріївна,

аспірантка КПІ ім. Ігоря Сікорського, Україна, Київ

Глоба Лариса Сергіївна,

професор, доктор технічних наук, в.о. завідувача кафедри інформаційно-телекомунікаційних систем КПІ ім. Ігоря Сікорського, Україна, Київ

Горелова Галина Вікторівна,

професор, доктор технічних наук, науковий керівник Інституту керування в економічних, соціальних й екологічних системах Південного федерального університету, Росія, Таганрог

Додонов Олександр Георгійович,

професор, доктор технічних наук, заступник директора з наукової роботи Інституту проблем реєстрації інформації НАН України, Київ

Завгородній Валерій Вікторович,

доцент, кандидат технічних наук, в.о. завідувача кафедри інформаційних технологій та дизайну Державного університету інфраструктури та технологій, Україна, Київ

Завгородня Ганна Анатоліївна,

кандидат технічних наук, старший викладач кафедри інформаційних технологій та дизайну Державного університету інфраструктури та технологій, Україна, Київ

Зилевіч Максим Олегович,

аспірант кафедри конструювання електронно-обчислювальної апаратури КПІ ім. Ігоря Сікорського, Україна, Київ

Зражевський Григорій Михайлович,

доцент, кандидат фізико-математичних наук, доцент кафедри теоретичної та прикладної механіки механіко-математичного факультету Київського національного університету імені Тараса Шевченка, Україна, Київ

Зражевська Віра Федорівна,

доцент, кандидат фізико-математичних наук, доцент кафедри диференціальних рівнянь фізико-математичного факультету КПІ ім. Ігоря Сікорського, Україна, Київ

Концеба Сергій Михайлович,

доцент, кандидат економічних наук, доцент кафедри інформаційних технологій Уманського національного університету садівництва, Україна, Умань

Корнага Ярослав Ігорович,

доцент, доктор технічних наук, доцент кафедри технічної кібернетики КПІ ім. Ігоря Сікорського, Україна, Київ

Леонова Алла Євгенівна,

начальник Управління стратегічного аналізу та інформаційної взаємодії Науково-дослідного центру електронно-обчислювальної техніки АТ «НІЦЕВТ», Росія, Москва

Ліщук Роман Ігорович,

доцент, кандидат технічних наук, завідувач кафедри інформаційних технологій Уманського національного університету садівництва, Україна, Умань

Логінова Олександра Вікторівна,

доцент, кандидат економічних наук, доцент Санкт-Петербурзького політехнічного університету Петра Великого, Росія, Санкт-Петербург

Масляк Павло Павлович,

старший науковий співробітник, кандидат технічних наук, доцент кафедри прикладної математики факультету прикладної математики КПІ ім. Ігоря Сікорського, Україна, Київ

Мацукі Йошіо,

професор, доктор технічних наук, професор Національного університету «Києво-Могилянська академія», Україна, Київ

Мухін Вадим Євгенійович,

професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Надеран Марьям,

аспірантка кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Новогрудська Ріна Леонідівна,

доцент, кандидат технічних наук, доцент кафедри інформаційно-телекомунікаційних систем КПІ ім. Ігоря Сікорського, Україна, Київ

Редько Ігор Володимирович,

професор, доктор фізико-математичних наук, професор кафедри конструювання електронно-обчислювальної апаратури КПІ ім. Ігоря Сікорського, Україна, Київ

Родащук Галина Юрївна,

доцент, кандидат економічних наук, старший викладач кафедри інформаційних технологій Уманського національного університету садівництва, Україна, Умань

Саак Андрій Андрійович,

аспірант Російського державного соціального університету, Росія, Москва

Сельський Євгеній Петрович,

студент КПІ ім. Ігоря Сікорського, Україна, Київ

Скуртол Світлана Дмитрівна,

доцент, кандидат економічних наук, доцент кафедри інформаційних технологій Уманського національного університету садівництва, Україна, Умань

Яганов Петро Олексійович,

доцент, кандидат технічних наук, доцент кафедри конструювання електронно-обчислювальної апаратури КПІ ім. Ігоря Сікорського, Україна, Київ