

**MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE  
SUMY STATE UNIVERSITY  
UKRAINIAN FEDERATION OF INFORMATICS**

# **PROCEEDINGS**

**OF THE VI INTERNATIONAL SCIENTIFIC  
CONFERENCE**

**ADVANCED INFORMATION  
SYSTEMS AND TECHNOLOGIES**

**AIST-2018**

**(Sumy, May 16–18, 2018)**



**SUMY**

**SUMY STATE UNIVERSITY**

**2018**

**UDC 004(063)**

**A**

**A**            **Advanced Information Systems and Technologies: proceedings of the VI international scientific conference, Sumy, May 16–18 2018 / Edited by S. I. Protsenko, V. V. Shendryk – Sumy: Sumy State University, 2018 – 145 p.**

This book comprises the proceedings of the VI International Scientific Conference “Advanced Information Systems and Technologies, AIST-2018”. The proceeding papers cover issues related to system analysis and modeling, project management, information system engineering, intelligent data processing, computer networking and telecommunications, modern methods and information technologies of sustainable development. They will be useful for students, graduate students, researchers who interested in computer science.

**UDC 004(063)**

## **International Scientific Committee:**

<b>A.N. Chornous, Sc.D (Ukraine)</b>	<b>V.I. Lytvynenko, Sc.D (Ukraine)</b>
<b>A.S. Dovbysh, Sc.D (Ukraine)</b>	<b>N. B. Shakhovska, Sc.D (Ukraine)</b>
<b>O.A. Borisenko, Sc.D (Ukraine)</b>	<b>E.A. Druzynin, Sc.D (Ukraine)</b>
<b>E.A. Lavrov, Sc.D (Ukraine)</b>	<b>S.I. Dotsenko, Sc.D (Ukraine)</b>
<b>V.O. Lyubchak, PhD (Ukraine)</b>	<b>T.V. Kovalyuk, PhD (Ukraine)</b>
<b>S.I. Protsenko, Sc.D (Ukraine)</b>	<b>A. Pakštas, PhD (United Kingdom)</b>
<b>A.M. Kulish, Sc.D (Ukraine)</b>	<b>O. Romanko, PhD (Canada)</b>
<b>M.M. Glybovets, Sc.D (Ukraine)</b>	<b>I. Polik, PhD (USA)</b>
<b>Yu. I. Grytsyuk, Sc.D (Ukraine)</b>	<b>V. Kalashnikov, Sc.D (Mexico)</b>
<b>I.V. Grebennik, Sc.D (Ukraine)</b>	<b>S. Berezyuk, PhD (Canada)</b>
<b>O.O. Yemets, Sc.D (Ukraine)</b>	<b>P. Davidsson, PhD (Sweden)</b>
<b>D.D. Peleshko, Sc.D (Ukraine)</b>	<b>M. Biagi, PhD (Italy)</b>

## **Organizing Committee:**

**Protsenko S. I., Sc.D, chairman (Ukraine);**  
**Shendryk V. V., PhD, co-chairman (Ukraine);**  
**Vaschenko S. M., PhD, co-chairman (Ukraine);**  
**Parfenenko Y. V., PhD (Ukraine), Nahornyi V. V., PhD (Ukraine),**  
**Zakharchenko V. P. (Ukraine), Shendryk S.O. (Ukraine),**  
**Boiko O. V., PhD, executive secretary (Ukraine).**

## **Contacts:**

**Address:** AIST conference, Sumy State University  
2 Rimsky-Korsakov Str., Sumy, 40000, Ukraine  
**website:** [www.aist.sumdu.edu.ua](http://www.aist.sumdu.edu.ua)  
**e-mail:** [aist@sumdu.edu.ua](mailto:aist@sumdu.edu.ua)

## CONTENTS

### SESSION 1 SYSTEM ANALYSIS AND MODELING

BRANCH AND BOUND METHOD FOR SOLVING OF LINEAR STOCHASTIC COMBINATORIAL OPTIMIZATION PROBLEM ON ARRANGEMENTS <b>Oleg Iemets, Tetiana Barbolina .....</b>	<b>8</b>
CLIENT SOLVENCY ESTIMATION USING INTELLECTUAL DATA ANALYSIS APPROACH <b>Petro Bidyuk, Vira Huskova, Oleksandr Terentiev .....</b>	<b>12</b>
FUNCTIONS OF DISPROPORTIONALITY AND THEIR APPLICATION <b>Viktor V.Avramenko .....</b>	<b>16</b>
COGNITIVE MODELING OF POTENTIAL PHARMACY DEMAND MANAGEMENT <b>Hanna Ropalo .....</b>	<b>21</b>
INFORMATION TECHNOLOGY FOR QUALITY MANAGEMENT OF ENGINEERING PRODUCTS <b>Evgeniy Lavrov, Mykola Bahmach, Galyna Mihalevska.....</b>	<b>25</b>
INTRODUCTION OF AN IRREGULAR GRID WITH RESPECT TO THE SPATIAL COORDINATE FOR THE METHOD OF LINES <b>Olga Dmytriyeva, Nadiia Huskova.....</b>	<b>30</b>
THE APPROXIMATION SURFACE REVIEW OF THE MULTIDIMENSIONAL TARGET FUNCTION FOR SURROGATE OPTIMIZATION PROBLEMS <b>Ruslana Trembovetska, Volodymyr Halchenko, Volodymyr Tychkov .....</b>	<b>34</b>
THE PROBLEMS OF AUTOMATION OF FINANCIAL TRANSACTIONS SUBJECT TO THE SIGNS OF INTERNAL FINANCIAL MONITORING <b>Serhii Mynenko .....</b>	<b>39</b>
WEB-BASED APPLICATION FOR PLANNING THE STRUCTURE OF HYBRID RENEWABLE ENERGY SYSTEM <b>Anastasia Verbytska, Yulia Parfenenko, Olha Boiko, Ozhikenov Kasymbek Adilbekovich, Assem Kabdoldina .....</b>	<b>43</b>

### SESSION 2 PROJECT MANAGEMENT

CONVERGENSE OF AGILE AND TRADITIONAL METHODOLOGIES IN IT PROJECTS <b>Juliya Kasianenko .....</b>	<b>48</b>
---	-----------

### SESSION 3 E-LEARNING TECHNOLOGIES

GAME MODEL OF BLENDED LEARNING IN A UNIFIED LEARNING ENVIRONMENT OF THE SUMY STATE UNIVERSITY <b>Inga Vozna, Serhey Shapovalov .....</b>	<b>53</b>
DEVELOPMENT OF A TECHNIQUE FOR THE CREATION OF VIRTUAL SIMULATORS OF ARTILLERY WEAPONS <b>Natalia Fedotova, Tetiana Yasinska.....</b>	<b>57</b>
ELECTRONIC MILITARY TRAINING DEVICE <b>Kateryna Savytska, Viktoriia Kudrytska, Natalia Fedotova .....</b>	<b>61</b>
INFORMATION SYSTEM FOR ACCOUNT OF STUDENTS WORK <b>Mariia Mova, Volodymyr Nahorny, Olga Zagovora, Maksym Portianoi.....</b>	<b>65</b>
INFORMATION TECHNOLOGY FOR MODELING OF HUMAN-MACHINE INTERACTIONS <b>Evgeniy Lavrov, Nadiia Pasko, Yan Voitsekhovskiy, Ruslan Plaks, Galyna Mihalevska.....</b>	<b>69</b>
THE SIMULATOR ON THE TOPIC “LAND AND DOIG METHOD” OF THE DISTANCE LEARNING COURSE “OPTIMIZATION METHODS AND OPERATIONS RESEARCH”: DEVELOPMENT AND SOFTWARE REALIZATION <b>Oleksandr Syvokin, Oleg Iemets .....</b>	<b>72</b>
PECULIARITIES OF DISTANCE LEARNING TECHNOLOGIES APPLICATION IN TRAINING FUTURE OFFICERS FOR THE ARMED FORCES OF UKRAINE <b>Olena Pavlichenko, Yurii Kalynovs’kyi, Halyna Serhieieva .....</b>	<b>76</b>

### SESSION 4 INFORMATION SYSTEMS ENGINEERING

AUTOMATED RECORDING AND PROCESSING OF LECTURE ATTENDANCE DATA USING RFID STUDENT CARDS <b>Alexander Kalashnikov, Hongwei Zhang, Misko Abramriuk .....</b>	<b>81</b>
HIERARCHICAL ALGORITHM OF THE MACHINE LEARNING FOR THE SYSTEM OF FUNCTIONAL DIAGNOSTICS OF THE ELECTRIC DRIVE <b>Anatoly Dovbysh, Victoria Zimovets .....</b>	<b>85</b>
PHYSICAL AND MATHEMATICAL MODELLING OF PITTING CORROSION <b>Nutthawut Suchato, Roger Light, Richard Smith, Alexander Kalashnikov .....</b>	<b>89</b>
SECURE MOBILE APPLICATION DEVELOPMENT <b>Roman Yatsenko, Viktor Obodiak, Valerii Yatsenko .....</b>	<b>93</b>
WEB SERVICE FOR MONITORING THE PRICES OF ONLINE STORES <b>Borys Kuzikov, Maksym Vynohradov .....</b>	<b>97</b>

## SESSION 5 INTELLIGENT DATA PROCESSIN

FEATURES OF CALCULATING RATINGS IN INFORMATION SYSTEMS <b>Olga Pronina, Elena Piatykop</b> .....	102
MATHEMATICAL MODEL OF THE MANAGING STRATEGY SELECTION IN PUBLIC CATERING ESTABLISHMENT (BASED ON MARKOV DECISION-MAKING PROBLEM) .....	
<b>Liliana Danilova</b> .....	106
OBJECT DETECTION BASED ON GROWING CONVOLUTIONAL NEURAL NETWORK FOR AUTONOMOUS SYSTEMS <b>Viacheslav Moskalenko, Alona Moskalenko, Artem Korobov, Borys Lypivets</b> .....	110
ON-BOARD GEOGRAPHIC INFORMATION SYSTEM OF IMAGES' IDENTIFICATION <b>Juliy Simonovskiy, Vladislav Piatachenko, Nikita Mironenko</b> .....	115

## SESSION 6 COMPUTER NETWORKING AND TELECOMMUNICATIONS

DETECTION AND PREVENTING LEAKS OF SENSITIVE DATA ON COMPUTER SYSTEMS <b>Mihail Babiy</b> .....	120
CROWDSOURCED MEASUREMENT: REALISATION, PRIVACY AND SECURITY IN THE FRAME OF SUMY STATE UNIVERSITY WIRELESS NETWORK QUALITY ANALYSE <b>Borys Kuzikov, Sergey Panchenko</b> .....	123

## SESSION 7 MODERN METHODS AND INFORMATION TECHNOLOGIES OF SUSTAINABLE DEVELOPMENT

INVESTIGATION OF THE INFLUENCE OF INFORMATION MANAGEMENT ON THE DEVELOPMENT OF THE COUNTRY <b>Vitaliya Koibichuk</b> .....	129
MODELING THE PROBABLE LOSSES OF BANKS FROM THEIR INVOLVEMENT IN THE PROCESS OF LEGALIZATION (LAUNDERING) OF INFLAMMABLE FUNDS <b>Anton Boiko, Tetiana Dotsenko</b> .....	133
REGIONAL SUSTAINABILITY ASSESSMENT THROUGH MULTIVARIATE STATISTICAL ANALYSIS <b>Tetiana Marynych, Stanislav Smolenko</b> .....	137
SIMULATION OF SCORING OF THE BANK'S BORROWERS CREDITWORTHINESS <b>Konstantin Gritsenko</b> .....	141

**SESSION 1**  
**SYSTEM ANALYSIS AND MODELING**

# Branch and Bound Method for Solving of Linear Stochastic Combinatorial Optimization Problem on Arrangements

Oleg Iemets<sup>1</sup>, Tetiana Barbolina<sup>2</sup>

<sup>1</sup>Poltava University of Economics and Trade, Ukraine, <sup>2</sup>V.G. Korolenko National Pedagogical University, Ukraine, tm-b@ukr.net

**Abstract** –The article deals with the solving of linear combinatorial optimization problems on a set of arrangements under probabilistic uncertainty.

We propose to branch the common set of arrangements assigning certain possible values for some of the variables. The way of bound computing is proposed and substantiated. The algorithm of branch and bound method is formulated.

**Keywords**–Euclidean combinatorial optimization problem; optimization problem on arrangements; stochastic combinatorial optimization; branch and bound method.

## I. INTRODUCTION

Actual trend of the modern theory of optimization is to study the problems of combinatorial nature under different types of uncertainty ([1–7] and others). In particular stochastic optimization problems attract the attention of researchers recently (for example [8–16]). One approach for formalization of optimization problems under uncertainty is based on the introduction of the order relation on the set of corresponding variables. Earlier the authors [11, 12] proposed two ways of ordering of random variables. The first one is based on comparison of mathematical expectation, dispersion, values and probabilities of discrete random variables. According to the second way a given set of independent random variables is partitioned by equivalence, which based on the comparison of their numerical characteristics; the linear order is defined on the quotient set. Some properties of this linear orders and corresponding optimization problems were discussed in [11–16]. In this article we propose branch and bound method for solving of stochastic combinatorial optimization problems on arrangements.

## II. STATEMENT OF H-PROBLEM

Let some of the initial data in the optimization problem be discrete random variables. The last will be denoted by Latin capital letters ( $A, B, \dots$ ). Let  $\Omega$  be a set of independent discrete random variables,  $J_n = \{1, 2, \dots, n\}$ . We say that

$$H(A) = (h_1(A), h_2(A), \dots, h_s(A))$$

(where  $h_i(A)$ ,  $i \in J_s$  is numeric characteristic of random variable  $A$ ) is an eigenvector. Suppose that eigenvector satisfies

$$h_i(\alpha A + \beta B) = \alpha^{\lambda_i} h_i(A) + \beta^{\lambda_i} h_i(B)$$

$$\forall i \in J_s \quad (1)$$

where  $A, B$  are independent random variables,  $\alpha, \beta \in \mathbb{R}^1$ ,  $\lambda_i \in \mathbb{Z}$ ,  $\lambda_i > 0$ .

We say that random variables  $X, Y \in \Omega$  are  $H$ -equivalent if and only if  $H(X) = H(Y)$  [12]. The equivalence class of an element  $X$  is called the  $H$ -class and is denoted by  $[X]_H$ .

If  $H(X) \leq_l H(Y)$  ( $\leq_l$  is a symbol of lexicographical order) for any  $X \in [X]_H$ ,  $Y \in [Y]_H$  then we say that classes  $[X]_H, [Y]_H$  are in order of non-decreasing and write  $[X]_k \preceq [Y]_k$ .

Using introduced linear order, let us order the elements of quotient set  $\omega = \Omega / H$ :  $[X_1]_H \preceq [X_2]_H \preceq \dots \preceq [X_m]_H$ .  $[X_m]_H$  is the maximum class and  $[X_1]_H$  is the minimum one. The definition of the minimum and maximum allows setting the optimization problem for finding the extreme elements in the given conditions. Such problems are called  $H$ -problems of stochastic combinatorial optimization problems.

Let  $[X]_H = ([X_1]_H, [X_2]_H, \dots, [X_k]_H)$ ,  
 $\Phi([X]_H) = \sum_{j=1}^k c_j [X_j]_H$  where  $c_j \in \mathbb{R}^1$ ,  $X_j \in \Omega$   
 $\forall j \in J_k$ . Linear  $H$ -problem of stochastic combinatorial



optimization on a sphere  $\Xi \subset \omega^k$  can be formulated as follow: find extremum and extremal of  $\Phi([X]_H)$

subject to  $[X]_H \in \Xi$ .

Further we discuss linear  $H$ -problem of stochastic combinatorial optimization where feasible set is a subset of common set of arrangements  $E_\eta^k(\Gamma)$  from elements of multiset  $\Gamma = \{[G_1]_H, [G_2]_H, \dots, [G_\eta]_H\}$  (main concepts of Euclidean combinatorial optimization are used from [17]). Suppose  $G_1, \dots, G_\eta$  are independent random variables,  $h_1(G_i) \geq 0 \quad \forall i \in J_\eta$ ; coefficients of linear function  $\Phi([X]_H)$  are positive real numbers.

Thus linear  $H$ -problem of stochastic combinatorial optimization on arrangements can be expressed as follow: find a pair  $\langle \Phi([X^*]_H), [X^*]_H \rangle$  such that

$$\Phi([X^*]_H) = \min_{[X]_H \in E_\eta^k(\Gamma)} \sum_{j=1}^k c_j [X_j]_H, \quad (2)$$

$$[X^*]_H = \arg \min_{[X]_H \in E_\eta^k(\Gamma)} \sum_{j=1}^k c_j [X_j]_H, \quad (3)$$

subject to

$$[X]_H = ([X_1]_H, [X_2]_H, \dots, [X_k]_H) \in \Xi, \quad (4)$$

where  $\Xi \subset \omega^k$ .

Further we will consider only  $H$ -classes. Therefore instead  $[X]_H$  we will write  $X$ .

#### I. PROPERTIES OF $H_D$ -PROBLEM

Consider function  $\Phi_1(x) = \sum_{j=1}^k C_j x_j$  where

coefficients  $C_j \quad \forall j \in J_k$  are  $H$ -classes ( $C_j \in \omega \quad \forall j \in J_k$ ),  $x = (x_1, \dots, x_k) \in R^k$ . Linear  $H_d$ -problem of stochastic combinatorial optimization is the problem that can be expressed as follow: find a pair  $\langle \Phi_1(x^*), x^* \rangle$  such that

$$\Phi_1(x^*) = \min_{x \in E_\eta^k(G)} \sum_{j=1}^k C_j x_j, \quad (5)$$

$$x^* = \arg \min_{x \in E_\eta^k(G)} \sum_{j=1}^k C_j x_j, \quad (6)$$

where  $E_\eta^k(\Gamma)$  is a common set of arrangements from elements of multiset  $\Gamma = \{g_1, g_2, \dots, g_\eta\}$ ,  $g_j \in R^1 \quad \forall j \in J_\eta$ .

Suppose elements of multiset  $G$  are in order of increasing:

$$0 < g_1 \leq g_2 \leq \dots \leq g_\eta. \quad (7)$$

Consider the following deterministic problem: find a pair  $\langle \bar{\Phi}_1(x^*), x^* \rangle$  such that

$$\bar{\Phi}_1(x^*) = \min_{x \in E_\eta^k(\Gamma)} \sum_{j=1}^k \bar{c}_j x_j, \quad (8)$$

$$x^* = \arg \min_{x \in E_\eta^k(\Gamma)} \sum_{j=1}^k \bar{c}_j x_j, \quad (9)$$

where  $\bar{\Phi}_1(x) = \sum_{j=1}^k \bar{c}_j x_j$ ,  $\bar{c}_j = h_1(C_j) \quad \forall j \in J_k$ .

As it follows from the lemma 1 [16], the minimal of function  $\bar{\Phi}_1(x)$  on  $E_\eta^k(\Gamma)$  is a permutation of elements of multiset  $\Gamma' = \{g_1, \dots, g_k\}$ . Consequently if point  $x$  is not an element of a common set of permutations  $E_k(\Gamma')$  then  $\bar{\Phi}_1(x^*) < \bar{\Phi}_1(x)$ . Since

$$\begin{aligned} \bar{\Phi}_1(x) &= \sum_{j=1}^k \bar{c}_j x_j = \sum_{j=1}^k h_1(C_j) x_j = \\ &= h_1\left(\sum_{j=1}^k C_j x_j\right) = h_1(\Phi_1(x)) \end{aligned}$$

we have  $\Phi_1(x^*) < \Phi_1(x) \quad \forall x \notin E_k(\Gamma')$ .

Therefore minimal in a solution of problem (5)–(6) is a permutation of elements of multiset  $\Gamma'$ , i.e. minimum

of objective function is  $L_1^* = \sum_{j=1}^k C_j g_{i_j}$  where  $g_{i_j} \in \Gamma'$ ,

$i_j \neq i_t \quad \forall i_j, i_t \in J_k$ ,  $\forall j, t \in J_k$ . Thus problem (5)–(6)

can be expressed as follow: find a pair  $\langle \varphi(Y^*), Y^* \rangle$  such that

$$\varphi(Y^*) = \min_{Y \in E_k(\Psi)} \sum_{j=1}^k g_j Y_j, \quad (10)$$

$$Y^* = \arg \min_{Y \in E_k(\Psi)} \sum_{j=1}^k g_j Y_j, \quad (11)$$

where  $F(Y) = \sum_{j=1}^k g_j Y_j$ , multiset  $\Psi = \{C_1, \dots, C_k\}$ .

Suppose elements of multiset  $\Psi$  are indexed so that satisfy the condition

$$H(C_1) \geq_l H(C_2) \geq_l \dots \geq_l H(C_k). \quad (12)$$

From theorem 4 [15] it follows that the arrangement, which satisfies the conditions  $Y_j^* = C_j \quad \forall j \in J_k$ , is minimal in a solution of the problem (10)–(11). Hence  $\sum_{j=1}^k g_j Y_j \leq \sum_{j=1}^k g_j C_j$  for any permutation of elements of

multiset  $\Psi = \{C_1, C_2, \dots, C_k\}$ . Then also  $\sum_{j=1}^k C_j x_j \preceq \sum_{j=1}^k g_j C_j$  for any permutation of elements of multiset  $\Gamma' = \{g_1, \dots, g_k\}$ , i.e.  $x^* = (g_1, \dots, g_k)$  is minimal in a solution of the problem (5)–(6). We prove the follow theorem.

*Theorem 1.* Suppose coefficients of objective function satisfy the condition (7) and elements of multiset satisfy the condition (12). Then the point  $x^*$  such that  $x_j^* = g_j \forall j \in J_k$  is minimal in a solution of problem (5)–(6).

## II. BRANCH AND BOUND METHOD FOR SOLVING OF H- AND HD-PROBLEMS

Now we consider branch and bound method for solving linear  $H$ -problem of stochastic combinatorial optimization and linear  $H_d$ -problem.

Let us first consider  $H$ -problem (2)–(4). We propose branch on common set of arrangements  $E_\eta^k(\Gamma)$  instead feasible set  $\Xi \cap E_\eta^k(\Gamma)$ . When an arrangements is obtained we check if this arrangement belong to set  $\Xi$ . Set  $\Theta' \subset E_\eta^k(\Gamma)$ , which is obtained on certain level of branching, is determined as follow

$$X_j = G_{r_j}, \quad j \in I \quad (13)$$

where  $I$  is a set of indexes  $I \subset J_k, |I| = t$ .

We say that in the set  $\Theta'$  values of variables  $X_j, j \in I$  are fixed. The rest variables are denoted  $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_\tau$  ( $\tau = k - t$ ). Let also  $\zeta_B = \{c_j | j \in I\}, \tilde{\zeta} = \{c_j | j \notin I\}$ . Suppose variables  $\tilde{X}_i$  ( $i \in J_\tau$ ) are indexed so that elements of multiset  $\tilde{\zeta} = \{\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_\tau\}$  satisfy  $\tilde{c}_1 \geq \tilde{c}_2 \geq \dots \geq \tilde{c}_\tau$ . Let also  $\Gamma_B = \{G_{r_j} | j \in I\}, \tilde{\Gamma} = \Gamma - \Gamma_B, \tilde{G}_i$  ( $i \in J_p, p = |\tilde{\Gamma}| = \eta - t$ ) are elements of multiset  $\tilde{\Gamma}$ ; we assume that  $\tilde{G}_i$  satisfy  $\tilde{G}_1 \preceq \tilde{G}_2 \preceq \dots \preceq \tilde{G}_p$ .

*Theorem 2.* Let  $\Theta' \subset E_\eta^k(\Gamma)$  satisfies (4). A bound of  $\Phi(X)$  on  $\Theta'$  may be calculated as

$$\xi(\Theta') = \sum_{j \in I} c_j G_{r_j} + \sum_{i=1}^{\tau} \tilde{c}_i \tilde{G}_i. \quad (14)$$

Proof. Clearly, for any  $X \in \Theta'$  values of variables  $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_\tau$  we have  $(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_\tau) \in E_p^\tau(\tilde{\Gamma})$ .

Suppose  $\tilde{\Phi}(\tilde{X}) = \sum_{i=1}^{\tau} \tilde{c}_i \tilde{X}_i$ . Since elements of multiset  $\tilde{\Gamma}$  are in order of non-decreasing we obtain

$$H(\tilde{G}_1) \leq_l H(\tilde{G}_2) \leq_l \dots \leq_l H(\tilde{G}_p).$$

At the same time coefficient of objective function are in order of non-increasing. From theorem 4 [13] it follows that the arrangement  $\tilde{X}^*$ , which satisfies the condition  $\tilde{X}_i^* = \tilde{G}_i \forall i \in J_\tau$ , is a minimal of function  $\tilde{\Phi}(\tilde{X})$  on the set  $E_p^\tau(\tilde{\Gamma})$ . Thus  $\tilde{\Phi}(\tilde{X}^*) \preceq \tilde{\Phi}(\tilde{X}) \forall \tilde{X} \in E_p^\tau(\tilde{\Gamma})$ .

Since eigenvector satisfies the condition (1) we see that for any  $X \in \Theta'$  the correlation

$$\begin{aligned} H(\Phi(X)) &= H\left(\sum_{j \in I} c_j X_j\right) + H\left(\sum_{i=1}^{\tau} \tilde{c}_i \tilde{X}_i\right) = \\ &= H\left(\sum_{j \in I} c_j G_{r_j}\right) + H(\tilde{\Phi}(\tilde{X})) \geq_l H\left(\sum_{j \in I} c_j G_{r_j}\right) + \\ &+ H(\tilde{\Phi}(\tilde{X}^*)) = H\left(\sum_{j \in I} c_j G_{r_j}\right) + H\left(\sum_{i=1}^{\tau} \tilde{c}_i \tilde{G}_i\right) = \\ &= \xi(Q') \end{aligned}$$

is true.

Therefore  $\xi(\Theta') \preceq \Phi(X)$  for all  $X \in \Theta'$ .

Theorem is proved.

Consider now linear  $H_d$ -problem: find a pair (5)–(6) subject to

$$x = (x_1, x_2, \dots, x_k) \in \Xi \subset R^k. \quad (15)$$

As above, set  $\Theta' \in E_\eta^k(\Gamma)$  is determined as follow

$$x_j = g_{r_j}, \quad j \in I. \quad (16)$$

Assume that

$$\xi(\Theta') = \sum_{j \in I} C_j g_{r_j} + \sum_{i=1}^{\tau} \tilde{C}_i \tilde{g}_i \quad (17)$$

(notation is similar given above at the solution of  $H$ -problem).

From theorem 1 it follows that point  $x^* = (g_1, \dots, g_k)$  is a minimal of function  $\tilde{\Phi}_1(x) = \sum_{i=1}^{\tau} \tilde{C}_i x_i$  on the set  $E_p^\tau(\tilde{\Gamma})$ . Therefore  $\tilde{\Phi}_1(x^*) \preceq \tilde{\Phi}_1(\tilde{x}) \forall \tilde{x} \in E_p^\tau(\tilde{\Gamma})$ . Further, similar to proof of theorem 2 we obtain that  $\xi(\Theta') \preceq \Phi'(x)$ . Hence we prove the follow theorem.

*Theorem 3.* Let  $\Theta' \subset E_\eta^k(\Gamma)$  satisfies (16). A bound of  $\Phi_1(x)$  on  $\Theta'$  may be calculated according to (17).

Let us formulate a branch and bound algorithm for solving problem (2)–(4) and (5), (6), (15). By  $\Theta_{r_1 r_2 \dots r_t}$  denote a set where values of  $t$  variables  $X_1, X_2, \dots, X_t$  (for  $H_d$ -problems —  $x_1, x_2, \dots, x_t$ ) are fixed ( $r_1, r_2, \dots, r_t$  are indexes of corresponding elements of multiset  $\Gamma$ ).

1. Let  $h = 0$ ,  $\Phi^0 = \infty$ ,  $t = 1$ .

2. Branch on  $E_\eta^k(\Gamma)$  to produce sets  $\Theta_{r_t}$  according to (13) or (16), where  $I = \{1\}$ ,  $G_{r_t}(g_{r_t})$  — different elements of multiset  $\Gamma$ . For all sets  $\Theta_{r_t}$  find a bound (14) (for  $H_d$ -problems — (17)).

3. Select the set  $\Theta_{s_1 \dots s_t}$  to be processed:

$\xi(\Theta_{s_1 \dots s_t}) \succeq \xi(\Theta')$  for all unbranched sets  $\Theta'$ .

4. Branch on  $\Theta_{s_1 \dots s_t}$  to produce  $\Theta_{s_1 \dots s_t r_{t+1}}$ , find a bound for all sets according to (14) (for  $H_d$ -problems — (17)).

5. Increase  $t$  by one. If  $t = k$  (set  $\Theta_{s_1 \dots s_{k-1} r_k}$  contain a single element) then go to step 6 else (if  $t < k$ ) go to step 8.

6. For all  $r_k$  compare  $\xi(\Theta_{s_1 \dots s_{k-1} r_k})$  and  $\Phi^h$ : if  $\xi(\Theta_{s_1 \dots s_{k-1} r_k}) \prec \Phi^h$  and  $(G_{s_1}, \dots, G_{s_{k-1}}, G_{r_k}) \in \Xi$  ( $(g_{s_1}, \dots, g_{s_{k-1}}, g_{r_k}) \in \Xi$ ) then increase  $h$  by one and denote  $X^h = (G_{s_1}, G_{s_2}, \dots, G_{s_{k-1}}, G_{r_k})$  (or

$x^h = (g_{s_1}, g_{s_2}, \dots, g_{s_{k-1}}, g_{r_k})$ ),  $\Phi^h = \xi(\Theta_{s_1 \dots s_{k-1} r_k})$ .

7. Decrease  $t$  by one.

8. If there exist unbranched set  $\Theta_{s_1 \dots s_{t-1} r_t}$  such that  $\xi(\Theta_{s_1 \dots s_{t-1} r_t}) \prec \Phi^h$  then return to step 3 else go to step 9.

9. If  $t > 1$  the return to step 8 else solving of problem is finished: if  $\Phi^h = \infty$  then feasible set is empty else  $\langle \Phi^h, X^h \rangle$  is a solution.

## CONCLUSIONS

Branch and bound method for solving of linear stochastic combinatorial optimization problems is considered. Minimum in the problem refers in accordance with the linear order introduced to the quotient set, which is generated by partition a given set of independent random variables based on the comparison of their numerical characteristics. The way of bound computing is proposed, algorithm is formulated.

## REFERENCES:

- [1] L.F. Hulyanytskyi, L.F. and I.I. Riasna “On formalization of combinatorial optimization problems on fuzzy sets”, *Teoriya optymalnykh rishen*, pp. 17-25, 2016. (in Ukrainian).
- [2] O.A. Yemets and A.A. Roskladka “Combinatorial optimization under uncertainty”, *Cybern. Syst. Analysis*, Vol. 44, Is.5, pp. 655-663, 2008.
- [3] V.A. Perepelitsa and F.B. Tebuevia (2007) *Discrete optimization and modeling under uncertainty of data*. Moscow : Akademiia Yestestvoznaniia, 2007, 151 p. (in Russian)
- [4] I. V. Sergienko O.O. Iemets and O.O. Yemets “Optimization problems with interval uncertainty: Branch and bound method” *Cybern. Syst. Analysis*, vol. 49, no. 5, pp. 673-683, 2013.
- [5] I.V. Grebennik, T.E. Romanova and S.B. Shekhovtsov “Interval estimation of alternatives in decision making in geometric design”, *Bionica Intellecta. Informacii. Yazyk. Intellect. scientific and technical journal*, vol. 2, pp.56-60, 2008. (in Russian)
- [6] O.O. Iemets and O.O. Yemets, *Solving combinatorial optimization problems on fuzzy sets*, Poltava : PUET, 2011, 239p. (in Ukrainian).
- [7] N.V. Semenova, L.N. Kolechikina and A. M. Nagima “Vector optimization problems with linear criteria over a fuzzy combinatorial set of alternatives”, *Cybern. Syst. Analysis*, Vol. 47, Is. 2, pp 250–259, 2011.
- [8] Yu.M. Ermol'ev and A.I. Yastremskii *Stochastic models and methods in economic planning*, Moskow : Nauka, 1979, 256 p. (in Russian).
- [9] I.V. Sergienko and M.V. Mikhalevich. “Application of stochastic optimization methods to analysis of the processes of economic transformation” *System research and information technologies*, vol. 4, pp.7-29, 2004. (in Ukrainian)
- [10] T. Yu. Ermolieva, Yu. M. Ermoliev, P. Havlik, A. Mosnier, D. Leclere, F. Kraksner, N. Khabarov, and M. Obersteiner, “Systems analysis of robust strategic decisions to plan secure food, energy, and water provision based on the stochastic GLOBIOM model,” *Cybern. Syst. Analysis*, vol 51, No. 1, pp.125, 2015.
- [11] O.O. Iemets and T.M. Barbolina, “About optimization problems with probabilistic uncertainty”, *Reports of the National Academy of Sciences of Ukraine*, No.11, pp. 40-45, 2014. (in Russian)
- [12] T.M. Barbolina, “About approach to optimization with probabilistic uncertainty using ordering of random variables” *Visnyk of Zaporizhzhya National University. Physical and Mathematical Sciences*, vol. 1, pp. 11-20, 2016. (in Russian.)
- [13] O.O. Iemets and T.M. Barbolina “Properties of the linear unconditional problem of combinatorial optimization on arrangements under probabilistic uncertainty” *Cybern. Syst. Analysis*, vol. 52, Is. 2, pp 285-295, 2016.
- [14] O.O. Iemets and T.M. Barbolina “Linear optimization problems on permutations under probabilistic uncertainty: properties and solution” *System research and information technologies*, Is. 1, pp. 107–119, 2016.
- [15] O.O. Iemets and T.M. Barbolina, “Solving Linear Unconstrained Problems of Combinatorial Optimization on Arrangements Under Stochastic Uncertainty”, *Cybern. Syst. Analysis*, Vol. 52, Is. 3, pp. 457-466, 2016.
- [16] O.A. Iemets and T.M. Barbolina, “Properties of Combinatorial Optimization Unconstrained Problems on Arrangements with Linear and Linear-Fractional Objective Functions”, *Journal of Automation and Information Sciences*, vol. 49, no. 1, pp.41–52, 2017.
- [17] Yu.G. Stoyan and O.O. Iemets, *Theory and methods of euclidian combinatorial optimization*. Kyiv : Instytut systemnykh doslidzhen osvity, 1993, 188 p. (in Ukrainian).

# Client Solvency Estimation Using Intellectual Data Analysis Approach

Petro Bidyuk, Vira Huskova<sup>1</sup>, Oleksandr Terentiev

Institute for Applied System Analysis at NTUU «I. Sikorsky KPI», Ukraine, <sup>1</sup>guskovavera2009@gmail.com

**Abstract** – This article is devoted to the estimation of client’s solvency and is an important problem for the banking system and other financial companies that provide loans to their clients. The questions devoted to the main methods of data mining for the collection of statistical information on clients, aimed at assessing the probability of default on loan repayment.

To solve the problem, modern estimation systems based on mathematical models have been developed that provide the possibility of analyzing and assessing the client's status regarding their solvency. Conducted assess the creditworthiness using approaches such as decision trees, cluster analysis (k-means), hierarchical clustering, nearest neighbor method, Binary choice regression models, nonlinear regression logit models, probit model, Bayesian networks.

**Keywords** – data mining, statistical information, mathematical models, cluster analysis, logit models, probit model, Bayesian networks.

## I. INTRODUCTION

Estimation of client’s solvency is very important problem for banking system and other financial companies that provide loans for their clients. To solve the problem it is necessary to develop modern scoring systems based on mathematical models providing a possibility for analysis and estimation of client’s state regarding their solvency. Scoring systems based on modern data analysis techniques, mathematical models and information technologies create a useful and reliable instrumentation for financial risk analysis and managerial decision making in many directions of human activities. This study is based on the application of intellectual data analysis techniques to collected clients’ statistical data aiming to estimation of default probability regarding loan return.

## II. PROBLEM STATEMENT

The problem is to develop scoring models for bank client’s solvency estimation using cluster analysis, decision trees, artificial neural networks (ANN), binary choice regression models and Bayesian belief networks (BN). Statistical data characterizing clients state and necessary for model building was taken from the Ukrainian banking system.

**Input data description.** To construct the scoring models the following dataset was hired: the training sample included 3250 items regarding clients and the test set included 100 items that had not been used for training. Each client was characterized by the following attributes:  $x_1$  is sex;  $x_2$  is age;  $x_3$  is family status;  $x_4$  is a number of children;  $x_5$  is working (not working);  $x_6$  is education;  $x_7$  is a type of work;  $x_8$  is availability of guarantying person;  $x_9$  is a sum of credit;  $y$  is a variable of result (probability of client default).

Statistical criteria characterizing quality of a scoring model:

1. Common accuracy (CA): this a relation of a number of correctly forecasted cases to the total number of cases  $N$  considered in specific problem (1):

$$CA = \frac{\text{number of correctly forecasted cases}}{N} \quad (1)$$

2. First and second type errors that are explained in table I. The first type error is direct loss of a bank; and the second type error is an income that was not realized.

TABLE I. FIRST AND SECOND TYPES OF ERRORS

Actual state of a client	Forecast: normal (client will return credit)	Forecast: default (client will not return credit)
Normal	Correct classification	Second type error
Default	First type error	Correct classification

## III. DECISION TREE

In a simple problem set decision tree is an approach to representing decision rules as a hierarchical sequential structure. As a result of application of the structure to statistical data is an answer “Yes” or “No” to respective specific questions. A typical algorithm for constructing topology (structure) of a decision tree includes two basic stages: “tree building” and “tree pruning”. There exist many known techniques of constructing decision trees. The

well known and widely applicable approaches to constructing the trees are CHAID, CART and QUEST [1].

Figure 1 shows an example of the tree constructing according to the QUEST approach using SPSS 16. The purpose of the scoring model is to give an answer to the question: “Should the client be given a loan?”

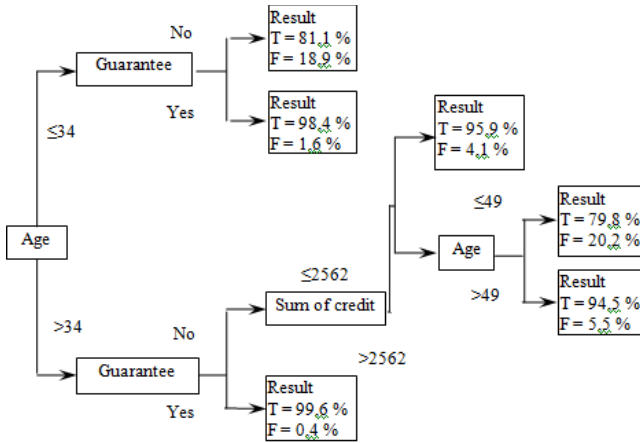


Figure 1. Credit scoring model in the form of decision tree QUEST

#### IV. CLUSTER ANALYSIS

A distinctive feature of cluster analysis from classification is that it does not require prior postulates regarding data. The clustering problem can be considered as a process of substantial order identifying in multi-dimensional data matrix what allows for clusters selecting, i.e. “condensed” groups of objects under study. The cluster analysis techniques today include about 100 various algorithms [2, 3] based on different principles. To solve the problem of constructing scoring models the following methods were used:

1) hierarchical: 1 – links between groups; 2 – links inside of groups; 3 – nearest neighbor; 4 – distant neighbors; 5 – weighted centroid; 6 – median algorithm; 7 – Varda algorithm;

2) iteration: 1 – *k*-means; 2 – *k*-median; 3 – two-step BIRCH algorithm.

The following criteria measuring the distance between clusters were used: – quadratic Euclid measure; – Pearson correlation coefficient; – Chebyshev measure, and Minkovich measure. To reduce the number of variables for model constructing the principal component approach was applied. The inverse factors were computed with varimax method (Table II).

#### V. ARTIFICIAL NEURAL NETWORKS

Another widely used intellectual data analysis method is based on ANN. In spite of the fact that there exist numerous possibilities for constructing ANN all of them have some similar features. To construct scoring models the following ANN constructing approaches were used: –

back propagation [4]; – resilient propagation [5]. The best models constructed are shown in Table III. The resilient propagation network was constructed with 2000 training epochs.

TABLE II. INVERSE FACTORS COMPUTED WITH VARIMAX METHOD

	Factor		
	1	2	3
Sex	<b>-0.538</b>	0.085	-0.223
Age	<b>0.811</b>	0.171	-0.036
Family status	<b>0.577</b>	<b>0.566</b>	-0.024
Number of children	-0.153	<b>0.769</b>	-0.033
Husband (wife) works	0.093	<b>0.697</b>	-0.014
Education	-0.158	-0.075	<b>0.537</b>
Type of work	<b>0.543</b>	-0.266	-0.369
Guarantee	0.121	-0.141	<b>0.529</b>
Sum of credit	0.085	0.181	<b>0.626</b>

TABLE III. THE BEST ANN MODELS CONSTRUCTED WITH THE DATA AVAILABLE

Model constructing algorithm	Number of neurons in a layer			Common accuracy	General classification error
	Input	Hidden	Output		
Back propagation	9	3	1	0,90	20%
Back propagation	9	5	1	0,91	15%
Resilient propagation	9	3	1	0,91	15%
Resilient propagation	9	5	1	0,90	15%

**Binary choice regression models.** In this model dependent variable accepts the value “ $y = 1$ ” if a client’s solvency is positive with probability 0.5 or higher, or “ $y = 0$ ” in a case of client’s default. Results of solvency estimation are given in Table 4.

**Linear regression.** The model constructed in the form of the linear regression is given in (2):

$$\begin{aligned}
 \text{probability} = & 0,7545 - 0,0358 \cdot x_1 + 0,0024 \cdot x_2 \\
 & + 0,01 \cdot x_3 + 0,0121 \cdot x_4 + 0,0214 \cdot x_5 + 0,0142 \cdot x_6 \\
 & + 0,0037 \cdot x_7 + 0,1112 \cdot x_8 - 1,6732 \cdot x_9
 \end{aligned} \quad (2)$$

**Nonlinear logit regression model** constructed is as follows:

$$\text{probability} = \frac{\exp(z)}{1 + \exp(z)} = \frac{1}{1 + (\exp(z))^{-1}} = \frac{1}{1 + \exp(-z)} \quad (3)$$

where

$$z = -0,9563 \cdot x_1 + 0,0695 \cdot x_2 + 0,1426 \cdot x_3 + 0,1875 \cdot x_4 + 0,2264 \cdot x_5 + 0,2727 \cdot x_6 + 0,0509 \cdot x_7 + 3,0315 \cdot x_8 - 0,0005 \cdot x_9$$

**Probit model** has the form:

$$probability = \int_{-\infty}^z \frac{1}{\sqrt{2 \cdot \pi}} \cdot \exp\left(\frac{-t^2}{2}\right) \cdot dt,$$

where

$$z = -0,2677 - 0,4469 \cdot x_1 + 0,035 \cdot x_2 + 0,0918 \cdot x_3 + 0,1097 \cdot x_4 + 0,1148 \cdot x_5 + 0,1436 \cdot x_6 + 0,1539 \cdot x_7 + 1,4139 \cdot x_8 - 0,0002 \cdot x_9$$

TABLE IV. RESULTS OF CONSTRUCTING AND APPLICATION OF REGRESSION MODELS

Type of regression	Error			Common accuracy	General classification error
	1st type	2nd type	Total		
Linear regression	175	0	175	0,947	15%
Logit model	163	6	169	0,949	13%
Probit model	168	2	170	0,948	14%

**Bayesian networks** (BN) are probabilistic and statistical model in the form of directed graph the vertices

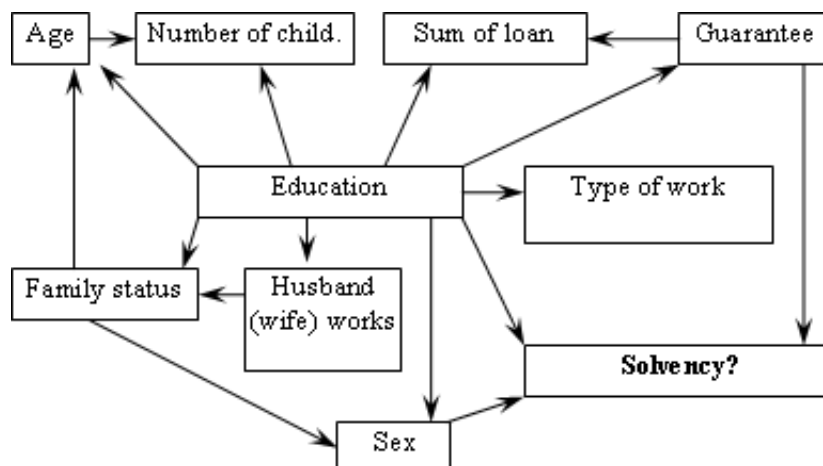


Figure 2. Scoring model in the form of Bayesian network

The BN model constructed in our case provided the following results of clients classification: (1) first type errors are equal to 3.5% – direct loss of the bank; i.e. the model classified these clients as reliable though their solvency turned out to be negative; (2) second type errors are equal to 4.8% – non-realized income of the bank; i.e. the BN model classified these clients as unreliable but they turned out to have positive solvency; (3) common accuracy of the model constructed comprises 91.8% what

of which represent model variables and directed arcs represent existing causal relations [6]. The model vertices are accompanied with conditional probability tables showing the probabilities of accepting specific value by a variable. First they were proposed by US scientist Judea Pearl in 1985 [7]. Thanks to availability of highly developed probabilistic numerical techniques today BN create very powerful approach to constructing probabilistic and statistical models in many fields of human activity (financial and economic forecasting, classification, medical and technical diagnostics, process control, financial risk analysis etc). One of the popular applications of BN is coping with possible probabilistic and amplitude uncertainties in modeling complex systems in various application fields. They may contain in the same model structure numerical and linguistic variables, and expert estimates what is very useful in many applications including solvency (and other financial risks) risk analysis using scoring models. Figure 2 illustrates the scoring model in the form of BN constructed with heuristic algorithm proposed by the authors in [5]. It is based on estimation of mutual information between vertices and minimum description length function. To test correctness of model building procedure the same kind of model was constructed with widely known GeNie system. Both results achieved turned out to be similar.

is actually one of the best scoring results we got in this study.

Analysis of results of computational experiments performed. Table 5 contains basic statistical and forecasting characteristics of the scoring models constructed. When solving the classification problem the threshold was selected at 90%, i.e. if the probability of a loan return was lower than 90% then the client was classified as the one with negative solvency.

The best results of modeling in this case were achieved with CHAID decision tree and  $k$ -means classification method. The general classification errors for both of these methods were 10%. Practically this means that out of 100 loans 10 were classified incorrectly. Most of other scoring model resulted in 15-

20% general classification errors what is also acceptable results.

The best common accuracy of about 0.949, was provided by the nonlinear logit model.

TABLE V. THE BASIC STATISTICAL AND FORECASTING CHARACTERISTICS OF THE SCORING MODELS CONSTRUCTED

Type of scoring model	Error			Common accuracy	General classification error
	1 <sup>st</sup> type	2nd type	Total		
With condition of 90% probability for loan return					
CHAID	76	275	351	0.895	10%
CART	68	299	367	0.891	20%
QUEST	47	546	593	0.823	15%
The best hierarchical clustering techniques with the use of quadratic Euclid norm					
Weighted centroid	174	93	267	0.921	15%
Varda algorithm	146	296	442	0.868	15%
The best hierarchical clustering techniques based on the use of Pearson coefficient					
Links between groups	177	26	203	0.939	20%
Links inside of groups	171	370	541	0.838	20%
Distant neighbors	177	71	248	0.926	20%
Median	177	26	203	0.939	15%
Nonhierarchical clustering techniques					
$k$ means for iterations and classification	116	564	680	0.797	10%
$k$ means only for classification	172	119	291	0.914	15%
Two-step algorithm	78	1245	1323	0.605	45%
The best hierarchical scoring method with preliminary factorization of variables					
Nearest neighbor	178	1	179	0.947	15%
Bayesian network	115	157	272	0.918	15%

## CONCLUSIONS

Constructing of mathematical scoring models for estimating financial risks and results of financial activities for enterprises using intellectual data analysis techniques is one of perspective approaches to solving complex problems of managerial decision making in financial institutions. The best results of constructing scoring models and their practical application in our particular case were received with CHAID decision tree and cluster analysis technique known as  $k$ -means. The best common accuracy result was achieved with hierarchical clustering, nearest neighbor method, and nonlinear regression logit model. Also very good result (quite acceptable) was achieved with Bayesian network. This application example shows that intellectual data analysis techniques can be used successfully for solving important practical problems of financial risk analysis including client's solvency estimation. Such techniques will help to enhance stability and security of functioning for the banking system what is especially important for weak transition economies that are very sensitive to various random internal and external disturbances.

The research will be continued in directions of improvement of scoring models characteristics and the

search for new model types. A promising direction of the study is in combining classical statistical models and models based on intellectual data analysis. Development and software implementation of specialized decision support systems based on the modern system analysis principles, models and techniques mentioned above will provide a possibility for a fast knowledge and instrumentation transfer from universities to financial institutions.

## REFERENCES:

- [1] G. F. Luger, *Artificial Intelligence*. Boston, MA: Pearson Education, 2009.
- [2] T.S. Mandhulatha, "An Overview on clustering methods," *Journal of Engineering*, vol. 2, no. 4, pp. 719–725, 2012.
- [3] L. Cosmin, *Cluster analysis*. New York: ComoLab, 2011.
- [4] S. Haykin, *Neural networks and learning machines*. New York: Prentice Hall, 2009.
- [5] M. Riedmiller and H. Braun, "A direct adaptive method for faster back propagation learning: the RPROP algorithm," IEEE International conference on image processing; Chicago, Illinois, USA, 3-6 May, 1993, pp. 586-591.
- [6] A. N. Terent'yev, P. I. Bidyuk, and L. A. Korshevnyuk, "Bayesian network as instrument of intelligent data analysis," *Journal of Automation and Information Sciences*, vol. 39, pp. 28-38, 2007.
- [7] J. Pearl, *Probabilistic reasoning in intelligent systems*. San Francisco: Morgan Kaufmann Publishers, 1988.

# Functions of Disproportionality and Their Application

Viktor V. Avramenko

Sumy State University, Ukraine, vv.avramenko@cs.sumdu.edu.ua

**Abstract** – Special functions are considered for obtaining new properties of numerical functions, that have been called the disproportionalities. There are the functions of disproportionality by n-order derivative and by n-order value one. All disproportionalities can be used as a numerical measure of the discrepancy between images. They are invariant with respect to scales. The disproportionalities can be used to solve problems of technical diagnostics and in a broader context for pattern recognition.

**Keywords** – disproportionality with respect to the n-order derivative, disproportionality with respect to the value of the n-th order, disproportionality in the parametric setting of functions, recognition of the state of the object, technical diagnostics, recognition of signal distortion.

## I. INTRODUCTION

There is a class of problems for which, in addition to the known, it is necessary to determine new characteristics of numerical functions. As an example, let us consider the problem of technical diagnostics of a large class of quasistationary objects whose static characteristics taken for a fixed time have the form

$$y = k(t)x \quad (1)$$

Here:

$x, y$  are the input and output parameters, respectively;  
 $t$  is a time;

$k(t)$  in (1) changes slowly in comparison with the input process  $x(t)$ .

These objects include amplifiers, converters, telemetry channels, tracking systems, many sensors. For a fixed value of  $k(t)$  at time  $t$ , expression (1), as is known, reflects the proportional relationship between  $x$  and  $y$ . This proportionality is the main sign for assessing the technical condition of the objects of the class indicated. So long as the static characteristic obtained at a fixed time  $t$  has the form (1), the object is considered to be in good order. Deterioration of the technical state of the object leads to the fact that the proportionality between  $x$  and  $y$  is violated at least for some values of  $x$ .

The static characteristic in the general case becomes

$$y = k(x, t)x + b(t), \quad (2)$$

where  $b(t) \neq 0$ .

The system of technical diagnostic should detect this deterioration and evaluate it quantitatively. For this, it

would be easiest to compare the values of the functions (1) and (2) for the one and the same  $x$ . However, for quasistationary objects,  $k(t)$  in (1) changes in time, as a rule, randomly. So it's difficult to determine its real value for the current mode of operation.

Thus, to assess the technical state of an object of the class under consideration, it is necessary to determine for each value of  $x$  from the domain of existence of the function (2) whether its proportional connection with  $y$  remains. If the proportionality does not exist, it is needed to estimate how much the connection differs from the proportional one. In this case, the comparison of the values of the functions (1) and (2) is excluded because of the unknown value of  $k(t)$  in (1).

## II. PROBLEM STATEMENT

A set  $X$  of real numbers and a set  $T$  of real ordered numbers are given. The set  $Y$  of real numerical functions is defined

$$y = f(x, t), \quad (3)$$

where  $x \in X, t \in T$ .

It is necessary to define a functional on the set  $Y$ . It must be zero if the connection between  $x$  and  $y = f(x, t)$  is proportional for given values  $x \in X$  and  $t \in T$ .

## III. SOLUTION OF THE PROBLEM

To solve this problem, the special functions have been developed in [1]. They allow one to obtain a quantitative estimate of the deviation of the relationship between two numerical functions from the proportional one.

From the outset we solve the problem for function of one variable  $y = f(x), x \in X$ .

In mathematics, the proportion is the equality of two relations. It is known that for  $y = f(x)$  a derivative

$\frac{dy}{dx}$  is the limit of the ratio  $\frac{\Delta y}{\Delta x}$  for  $\Delta x \rightarrow 0$ . Hence, if the condition

$$\frac{y}{x} = \frac{dy}{dx}, \quad (4)$$

is performed for a given  $x \in X$ ,



we can assume that at the point under study the relationship between  $x$  and  $y$  is proportional. For  $y = cx$ ,

where  $c = \text{const}$ , the condition (4) is satisfied in the entire function definition area.

In general, for  $y = f(x)$ , it can only be performed for several values of  $x$  or doesn't exist.

**Definition 1.** The disproportionality over the first-order derivative of a function  $y = f(x)$  with respect to  $x$  is the difference between  $\frac{y}{x}$  and  $\frac{dy}{dx}$ .

We denote it as

$$@d_x^{(1)}y = \frac{y}{x} - \frac{dy}{dx} \quad (5)$$

The symbol @ is chosen to denote the operation of determining the disproportionality,  $d$ - means "derivative". The left side is read as "d one y with respect to x".

**Lemma 1.** The function  $y = cx$ , where  $c = \text{const}$ , is the only one for which the condition (4) is performed in whole its domain.

The proof reduces to obtaining a general solution of the differential equation (4) with separating variables. It is easy to see that the only general solution is  $y = cx$  where  $c = \text{const}$ , which is what we had to prove.

The transition to a function of two variables  $y = f(x, t)$ , where  $x \in X, t \in T$ , assumes setting the value of  $t$  when studying the proportionality between  $x$  and  $y$ . Hence the disproportionality with respect to the first-order derivative of the function with respect to  $x$  for a given value of  $t$  has the form

$$@d_x^{(1)}y_t = \frac{f(x, t)}{x} - f_x'(x, t) \quad (6)$$

The expression (6) meets the requirements set forth in the statement of the problem, and can be taken as the desired functional.

**Example 1.** Let  $y = k(t)x + b$ . For a fixed time  $t$   $k(t) = c$ . Here  $b, c$  are constants. Then  $@d_x^{(1)}y_t = \frac{b}{x}$ .

For  $b=0$   $@d_x^{(1)}y_t = 0$ , which should be in the presence of a proportional relationship between  $x$  and  $y$ .

**Example 2.**  $y = -kx \ln|cx|$ , where  $k, c$  are constants.

$$@d_x^{(1)}y = k$$

**Example 3.**  $y = kx^2 + cx$ , where  $k, c$  are constants.

$$@d_x^{(1)}y = kx.$$

For a multidimensional function  $y = f(x_1, x_2, x_3, \dots, x_p, t)$ , we define a particular

disproportionality with respect to the first-order derivative with respect to  $x_i$

$$@\partial_{x_i}^{(1)}y_t = \frac{f(x_1, x_2, x_3, \dots, x_p, t)}{x_i} - f_{x_i}'(x_1, x_2, x_3, \dots, x_p, t) \quad (7)$$

**Example .** Let  $y = k_1(t)x_1 + k_2(t)x_2$ . For fixed  $t$   $k_1(t) = c_1, k_2(t) = c_2$ , where  $c_1$  and  $c_2$  are constants. Then

$$@\partial_{x_1}^{(1)}y_t = \frac{y}{x_1} - \frac{\partial y}{\partial x_1} = c_2 \frac{x_2}{x_1}$$

#### IV. N-ORDER DERIVATIVE DISPROPORTIONALITY

We define the concept of disproportionality with respect to the  $n$ -order derivative, which must be equal to zero for

$$y = cx^n \quad (8)$$

the whole range of its definition.

For (8) we have the equality

$$\frac{y}{x^n} = \frac{1}{n!} \frac{d^n y}{dx^n} \quad (9)$$

**Definition 2.** The disproportionality with respect to the derivative of the  $n$ -order of a function  $y = f(x)$  with respect to  $x$  is

$$@d_x^{(n)}y = \frac{y}{x^n} - \frac{1}{n!} \frac{d^n x}{dx^n}, \quad (10)$$

The  $n$ -order derivative disproportionality of the function  $y = f(x, t)$  with respect to  $x$  for a fixed  $t$  has the form

$$@d_x^{(n)}y_t = \frac{f(x, t)}{x^n} - \frac{1}{n!} \frac{\partial^n x}{\partial x^n}, \quad (11)$$

In (10) and in (11)  $n$  is an integer greater than zero.

**Example.** Let  $y = k(t)x^3$ . For a fixed  $t, k(t)=c = \text{const}$ .

Then, in accordance with (11), for

$$n=1 \quad @d_x^{(1)}y_t = -2cx^2,$$

$$n=2 \quad @d_x^{(2)}y_t = -2cx,$$

$$n=3 \quad @d_x^{(3)}y_t = 0.$$

In the general case, for calculating the disproportionality with respect to the derivative, the origin can be transferred to an arbitrary point in accordance with the rules for parallel transfer of the coordinate axes. For fixed  $t$ , when the origin is transferred to the point  $M(x_0, y_0, t)$ , (11) is transformed to the form

$$@d_{x-x_0}^{(n)}(y - y_0) = \frac{f(x, t) - y_0}{(x - x_0)^n} - \frac{1}{n!} \frac{\partial^n f(x, t)}{\partial x^n} \quad (12)$$

It's determine the particular  $n$ -order derivative disproportionality  $y = f(x_1, x_2, x_3, \dots, x_p, t)$  with respect to  $x_i$  for a given  $t$

$$@ \partial_{x_i}^{(n)} y_t = \frac{y}{x_i^n} - \frac{1}{n!} \frac{\partial^n y}{\partial x_i^n} \quad (13)$$

#### V. DERIVATIVE DISPROPORTIONALITY IF A FUNCTION IS SPECIFIED IN A PARAMETRIC FORM

$$\text{Let } x = \varphi(t), \quad y = \psi(t), \quad (14)$$

where  $t \in [T_1, T_2]$  and the inverse function  $t = \Phi(x)$  exists.

In this case the  $n$ -order derivative disproportionality (10) for  $y = \Psi[\Phi(x)]$  with respect to  $x$  is determined

by applying the rules of calculation of  $\frac{d^n y}{dx^n}$  under parametric dependence of  $y$  on  $x$ .

In particular, for  $n=1$

$$@ d_x^{(1)} y = @_{\phi(t)}^{(1)} \psi(t) = \frac{y}{x} - \frac{y_t'}{x_t'} = \frac{\psi(t)}{\phi(t)} - \frac{\psi_t'(t)}{\phi_t'(t)} \quad (15)$$

The disproportionality (15) is a function of the parameter  $t$ . For  $\psi(t) = c\phi(t)$ , where  $c = \text{const}$ , it equals zero throughout the entire range of variation of  $t$ .

It's possible to calculate the derivative disproportionality with respect to  $x$  successively. It means the calculation of disproportionality some function with respect to  $x$ . Then the disproportionality of previous disproportionality is calculated with respect to  $x$ . It's repeated  $m$  times. Let's defined it for  $y = f(x, t)$  as

$@(m) @ d_x^{(n)} y_t$ . It called  $m$ -disproportionality. For example  $@(3) @ d_x^{(1)} y_t$ , means  $@ d_x^{(1)} [ @ d_x^{(1)} [ @ d_x^{(1)} y_t ] ]$ .

#### VI. VALUE DISPROPORTIONALITY FUNCTIONS

Sometimes it's more convenient to compare the deviations of the dependence of  $y$  on  $x$  on the proportional at some points in the domain of the function assignment, using the corresponding estimate with respect to the value  $x$  instead of estimating the disproportionality with respect to the derivative.

**Definition 3.** The  $n$ -order value disproportionality with respect to  $x^n$  is equal  $n$ -order derivative disproportionality (10) multiplied by  $x^n$ . Denote it as  $@ v_x^{(n)} y$ .

$$@ v_x^{(n)} y = y - \frac{x^n}{n!} \frac{d^n y}{dx^n} \quad (16)$$

In particular, for  $n=1$

$$@ v_x^{(1)} y = y - x \frac{dy}{dx} \quad (17)$$

The left-hand side of (17) is read as this:

" at value one  $y$  with respect to  $x$ ".

The disproportionality (17) is equal to the difference between  $y$  and its possible value in the case of a proportional relationship between  $y$  and  $x$  with a proportionality coefficient equal to the derivative at the point under study.

For  $y = f(x, t)$   $x \in X, t \in T$  the value disproportionality of  $x^n$ , where  $n$  is an integer greater than zero, is determined for a fixed  $t$ .

$$@ v_x^{(n)} y_t = y_t - \frac{x^n}{n!} \frac{d^n y_t}{dx^n} \quad (18)$$

If the functions are specified in a parametric form in accordance with (15)

$$@ v_x^{(1)} y = @ v_{\phi(t)}^{(1)} \psi(t) = \psi(t) - \phi_t(t) \frac{\psi_t'(t)}{\phi_t'(t)} \quad (19)$$

#### VII. PROPERTIES OF DISPROPORTIONALITIES FUNCTIONS

Below are several properties  $n$ -order disproportionality (10), (18).

1. Multiplying of the function by a constant factor  $C$  leads to multiplying of its  $n$ -order disproportionality with respect to  $x^n$ .

2. The disproportionality (10), (18) of the sum (difference) of functions is equal to the sum (difference) of their disproportionality (10), (18) correspondingly.

3. For  $y = k(t)x^n$ , where  $n$  is an integer greater than zero, for a fixed value of  $t$   $@ d_x^{(n)} y_t = 0$  in the entire function definition area.

4. Another property follows from the proof of the next lemma.

**Lemma 2.** Given a function  $y = f(x, t)$ ,  $x \in X, t \in T$ . Let  $@ d_x^{(n)} y_t \neq 0$ , but  $@(n) @ d_x^{(1)} y_t = 0$ . Let us prove that in this case  $y = f(x, t)$  has the form

$$y = k_n(t)x^n + k_{n-1}(t)x^{n-1} + \dots + k_1(t)x, \quad (20)$$

**Proof.** Let us verify that the disproportionality with respect to the  $n$ -order derivative is not zero. In accordance with (11)

$$@ d_x^{(n)} y_t = \frac{1}{x^n} [k_{n-1}(t)x^{n-1} + k_{n-2}(t)x^{n-2} \dots + k_1(t)x]. \quad (21)$$

By hypothesis, at least one of  $k_{n-1}(t), k_{n-2}, \dots, k_1(t)$  is not equal to zero. Consequently, the disproportionality (21) is also not equal to zero.

Now we find a sequence  $@(n) @ d_x^{(1)} y_t$  of  $n$  first-order disproportionalities (6) for (20).

For convenience, each disproportionality in this sequence will be denoted by  $Z$  with the corresponding index.

$$Z_1 = @ d_x^{(1)} y_t = -[(n-1)k_n(t)x^{n-1} + (n-2)k_{n-1}(t)x^{n-2} + \dots + 2k_3(t)x^2 + k_2(t)x]$$

$$Z_2 = @ d_x^{(1)} Z_1 = [(n-1)(n-2)k_n(t)x^{n-2} + (n-2)(n-3)k_{n-1}(t)x^{n-3} + \dots + 6k_4(t)x^2 + 2k_3(t)x]$$

⋮

$$Z_i = @ d_x^{(1)} Z_{i-1} = (-1)^i \sum_{j=i}^n k_j(t)x^{j-i} \prod_{m=1}^i (j-m)$$

⋮

$$Z_{n-1} = @ d_x^{(1)} Z_{n-1} = (-1)^{n-1} (n-1)! k_n(t)x$$

$$Z_n = @ d_x^{(1)} Z_{n-1} = 0$$

This result was required to be obtained.

## VIII. APPLICATION OF DISPROPORTIONALITY FUNCTIONS

These functions of disproportionate are a measure of the discrepancy between two numerical functions. They allow at a single point to quantify how much one function is unlike the other and this estimate does not depend on the scale of their representation. Below are some examples of their application.

1. A class of objects that, in the corrected state, have static characteristics of the form (1) is showed in introduction. Uncontrolled effects on the parameters of such an object can lead to a deterioration of its technical state and, as a consequence, to the deviation of the expression for a static characteristic from the form (1).

Calculation of the first-order value disproportionality of the static characteristic for fixed parameter (time)  $t$  for one or several values of  $x$  allows obtaining information necessary for assessing the technical state of the object and finding the reasons for its deterioration. So, for example, if for given  $x$  and  $t @ v_x^{(1)} y_t \neq 0$ , this may indicate either that for  $t=const$  the static characteristic does not pass through the origin of coordinates, or about a nonlinearity of  $y = f(x, t)$ . So nonlinearity can be the result of the appearance of uncontrolled feedbacks. Comparing values  $@ v_x^{(1)} y_t$  for several values of  $x$  can allow you to find the reason for these changes.

The disproportionality (18), computed for an integer  $n > 1$ , can be used to diagnose devices whose static characteristics are described by an expression  $y = k(t)x^n$ . One of them is a device raising a voltage to a power two.

2. Uncontrolled effects on the parameters of an object can be random processes.

Accordingly, the disproportionality (18)  $@ v_x^{(1)} y_t$  for the static characteristic  $y = f(x, t)$  will be random for a given  $x$  and  $n=1$ . By monitoring the characteristic and examining its statistical relationships with other processes, it is possible to obtain the information necessary to establish the causes of deterioration of the technical condition of the object.

3. There is a class of production process plants for which the required relationship between the consumption  $x$  of the resource and the quantity  $y$  of products released from it on the segment  $[x_0, x_k]$  is described by the expression

$$y = k(t)(x - x_0) + b_0,$$

where  $b_0 \leq 0$  is the permissible value of product losses in the production process;

$t$  is the time;

$k(t)$  - depends of randomly changing of the characteristics of the resource such as the percentage of the useful product in the processed raw materials, caloric content, ash content, humidity of the fuel being burned, etc. Usually determining the current values of these characteristics is a difficult task and is not realized. Therefore, the value of  $k(t)$  will be assumed to be unknown.

If the quality of the technological process is estimated only by the value of the product losses, then the best way to control it is the measurement of proportionality between  $(y - b_0)$  and  $(x - x_0)$  regardless of the value of  $k(t)$

$$y - b_0 = k(t)(x - x_0) \quad (21)$$

Usually,  $x$  and  $y$  are random functions of time. For many technological installations,  $x(t)$ ,

$y(t)$  and  $k(t)$  are quasi-stationary random processes, which allows us to determine the current regression equation for  $x$  and  $y$  at time  $t$ . It is required that it has the form (21).

Suppose that for  $x \in [x_0, x_k]$   $k(t)$  and the magnitude of the losses do not depend on  $x$ . Then the current regression equation looks like

$$y = k(t)(x - x_0) + b(t) \quad (22)$$

where  $b(t) \in [b_0 + \Delta b(t)] \leq 0$  is the current value of product losses.

There are many technological processes for which direct determination of the current loss values  $b(t)$  is very complicated and therefore is not carried out. Thus, to assess the quality of the process by the magnitude of product losses, it is required for the value  $x \in [x_0, x_k]$  to determine whether the relationship between  $(y - b_0)$  and  $(x - x_0)$  is proportional for the current regression equation (22) and if it is not, how much it differs from the proportional equation. In this case, the comparison of the values of the functions (21) and (22) is not possible because the value of  $k(t)$  is unknown. This problem is solved by means calculation of the disproportionality (18) of the current regression equation (22) for fixed  $t$  at  $n=1$  and the transfer of the origin of the coordinates to the point

$M(x_0, b_0, t)$ . In this case the disproportionality (18) for (22) has a form

$$\Delta v_{(x-x_0)}^{(1)}(y - b_0) = \Delta b(t).$$

The increment of losses is got.

4. The disproportionality (19) with the parametric specification of the function  $y$  from  $x$  is recommended to be used to monitor the signal distortion at the output of the device being diagnosed in comparison with the input signal. The causes of distortions can be the influence of the dynamics of the object, the nonlinearity of its static characteristic, the delay of the output signal with respect to the input signal.

One of the distortion is an appearance of a constant component in the output signal. For example, it's can be happened for DC amplifiers when due to so-called "zero drift" its static characteristic does not pass through the origin of coordinates.

Obviously, distortion is absent only if the connection between the input  $\varphi(t)$  and output  $\psi(t)$  signals is proportional to the whole area of its investigation, i.e. when,  $\psi(t) = c\varphi(t)$  where  $c = \text{const}$ . In this case, as can be seen from (15),  $\Delta d_{\varphi(t)}^{(1)}\psi(t) = 0$ .

The fact that the disproportionality is different from zero indicates that there is a signal distortion  $\psi(t)$  in comparison with  $\varphi(t)$ . The magnitude of this distortion is determined by (19).

There are the apparatus and methods for measuring [2] the so-called nonlinear distortions. In this case, the input signal must be harmonic. All these methods are reduced to measuring the parameters of the higher harmonics of the output signal and, as a rule, spectral analyzers are used. This complicates the equipment used in the analysis. The use of the proposed characteristics (15), (19) makes it possible to dispense with spectral analyzers.

In addition, the restriction is removed, in which the input signal must be harmonic. As a result, the measurement of signal distortion can be performed directly in the normal operation mode of the device. It is characteristic that the magnitude of the distortion, as seen from (19), is measured in the same physical units as the output signal.

The quantitative estimation of signal distortion at the output of the device can represent not only an independent interest, but also serve as an information sign in the technical diagnostics of this device.

## CONCLUSIONS

New properties of numerical functions are introduced - a disproportionality with respect to derivatives and a disproportionality with respect to the values and also the corresponding functions for their measurement are suggested. They can be used as a measure of the discrepancy between numerical functions at individual points. Their application can be used for the recognition of images, as well as for solving problems of technical diagnostics.

## REFERENCES:

- [1] V.V. Avramenko, "Harakteristiki neproporcionalnosti chislovyh funkcyj", Dep. V GNTB Ukrainy 19.01.98, №59-Uk98. unpublished.
- [2] G. P. Shkurin, *Spravochnik po elektro- i elektronno-izmeritelnyh priboram*, Moscow, 1982.

# Cognitive Modeling of Potential Pharmacy Demand Management

Hanna Ropalo

Zaporizhzhya National Technical University, Ukraine, annropalo@gmail.com

**Abstract** – The purpose of this study is to create a decision support system for placing a new pharmacy on the basis of building a cognitive model. Based on the method of impulse processes and analysis of cognitive maps, an assessment of the attractiveness of the location of the pharmacy is determined.

**Keywords** – cognitive model, cognitive maps, impulse processes, pharmacy, demand.

## I. INTRODUCTION

Half of all retail outlets in Ukraine are concentrated in the hands of small pharmacy companies with up to 10 outlets. The number of players in the pharmacy market is quite large, which makes it highly competitive. At the same time, the process of choosing a particular pharmacy by the customer is poorly formalized. Therefore, in order to investigate this issue, we need an approach that focuses on the qualitative analysis of complex situations interpreted as weakly structured systems. The purpose of this article is to build a decision support system for managing the potential demand of a pharmacy on the basis of building a cognitive model. Using the construction and analysis of cognitive maps, an assessment of the attractiveness of the location for the location and functioning of the pharmacy is determined.).

## II. MAIN TEXT

A qualitative analysis of a complex situation involves the identification of trends in ongoing processes, a qualitative assessment of these trends and the choice of measures that promote their development in the right direction [3].

For the first time, the methodology of cognitive modeling was introduced by R. Axelrod. (Axelrod, 1976). He developed this methodology, based on the theory of graphs, the theory of decision-making, as well as on the ideas of psychology; formulated the basis of the cognitive approach to decision-making and the mathematical apparatus for the analysis of cognitive maps [6].

The cognitive model is based on the formalization of cause-effect relationships that take place between the factors characterizing the system under study. The result of formalization is the presentation of the system

in the form of a cause-effect network, called a cognitive map:

$G = \langle E, W \rangle$ , where  $E = \{e_1, e_2, \dots, e_n\}$  is a set of factors (also called concepts),  $W$  is a binary relation on the set  $E$  that defines a set of connections between its elements.

Elements  $e_i$  and  $e_j$  are considered to be related by the ratio  $W$  (denoted by  $(e_i, e_j) \in W$  or  $e_i W e_j$ ) if changing the value of the concept  $e_i$  (cause) leads to a change in the value of the concept  $e_j$  (result). In accordance with the terminology of cognitive modeling, in this case it is said that the concept  $e_i$  influences the concept of  $e_j$ .

Moreover, if the increase in the value of the concept-cause leads to an increase in the value of the concept-effect, then the influence is considered positive, if the value decreases - negative. The methodology for constructing and analyzing cognitive maps is described in more detail in [1-6]. In the process of applying this technique, the experts identified causal relationships. On their basis, a cognitive map was built.

Factors affecting the choice of location for the a pharmacy:

1. The level of potential demand. The main factor that reflects the potential profit of the pharmacy.
2. The quality of the pedestrian flow. The factor that shows the percentage of buyers who are in the pharmacy from the general flow of passers-by.
3. Purchasing power. Reflects the real possibility for people passing by the pharmacy to make a purchase. The purchasing power affects the size of the average check.
4. Type of district. This factor reflects the motivation of buyers who pass by the pharmacy (business part of the city) or purposefully go to this pharmacy (sleeping area).
5. Competitive environment. Absence or a small number of competitors will avoid price wars and other methods of luring customers.
6. Distance to medical institutions. The presence of a medical institution near the pharmacy will provide its customers with the needs for pharmaceutical services.
7. Advertising companies. They are a method of stimulating interest among clients to a pharmacy.

8. Traffic. The higher the percentage of people passing by the pharmacy, the greater the demand for the services it provides.

9. Qualification of staff. Dislike of the pharmacist will lead to reluctance to visit the pharmacy where he works.

10. Marketing strategy. It is an instrument of struggle of the pharmacy, which turned out to be in unfavorable competitive conditions.

11. Number of storeys. The factor reflects the potential number of future customers.

12. Availability of public transport stops. The best place for a pharmacy is near transport interchanges. It can be a stop for several kinds of transport, a metro station, a busy intersection.

13. Lifetime. Affects the need to purchase drugs

14. Population density. The factor shows what level of income the pharmacy will have under favorable conditions.

15. The average salary for the region. The well-being of residents directly affects the ability to make purchases in the pharmacy.

16. Ecological situation. The factor that affects people's health

17. Unemployment rate. Factor showing the financial well-being of customers.

18. The availability of an online pharmacy. Factor that simplifies the process of buying drugs. Forms a positive opinion of the client about the network.

19. Health problems. The factor that encourages people to be pharmacy clients.

20. Assortment. The factor that influences the client's satisfaction with the work of the pharmacy.

21. The presence of shopping centers, shops. Provides the pharmacy with a large number of people passing by, and, accordingly, will increase the potential demand.

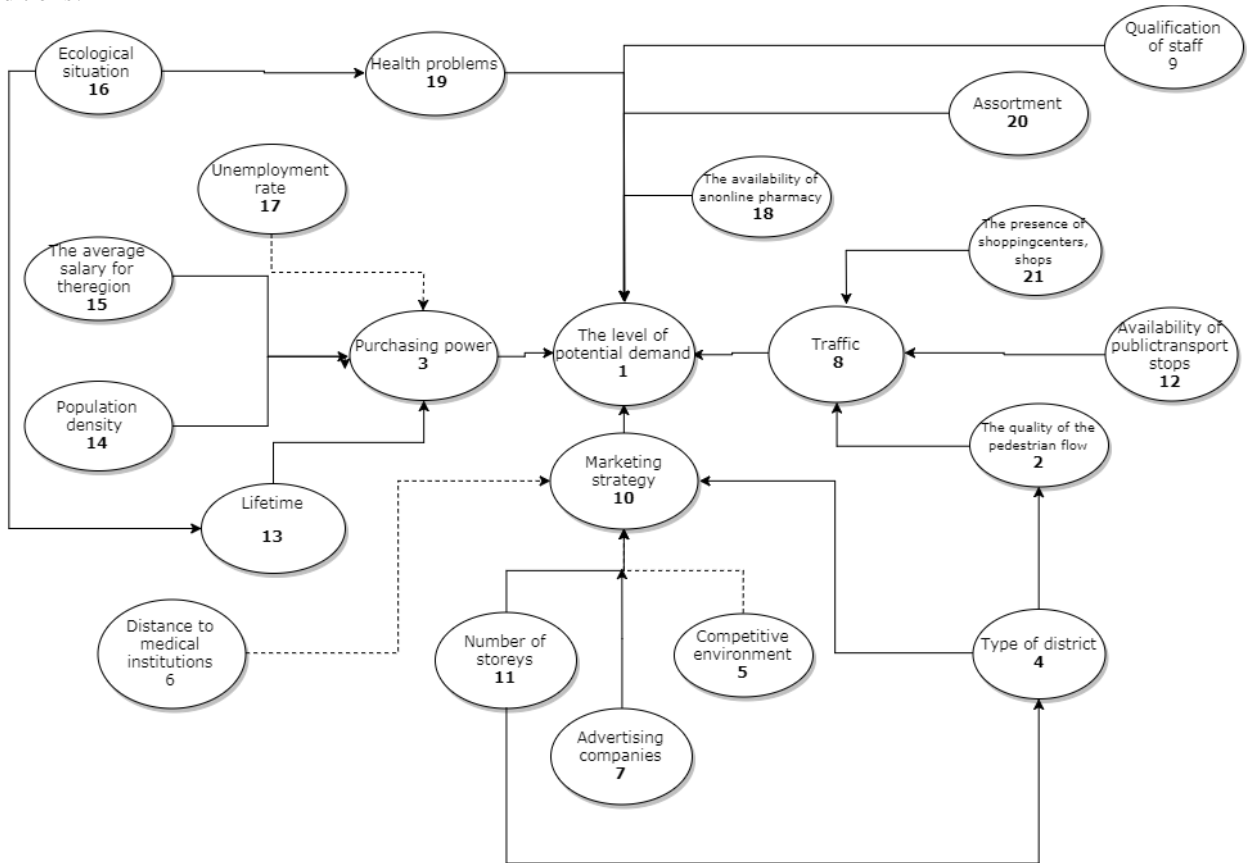


Figure 1. Cognitive site selection map for a pharmacy

It is advisable to formalize the structure of these connections using the cognitive model in the form of a signed digraph, which is shown in Fig. 1.

In the sign digraph all relations are associated with a signed weight function:

$$w(i, j) = \text{sign}(i, j) = \begin{cases} -1, \\ 0, \\ 1 \end{cases}, \quad (1)$$

where

-1 denotes the connection between such vertices, to which correspond the variables with opposite dynamics;  
 0 means no connection between vertices;

1 denotes the relationship between variables with the same direction of dynamics. The weights of the negative bonds in Fig. 1 are indicated by a dotted line, and positive - by a continuous line.

TABLE I WEIGHT OF CONNECTION OF THE COGNITIVE MODEL

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. The level of potential demand	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. The quality of the pedestrian flow	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3. Purchasing power	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4. Type of district	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5. Competitive environment	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6. Distance to medical institutions	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0
7. Advertising companies	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
8. Traffic	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9. Qualification of staff	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10. Marketing strategy	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11. Number of storeys	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
12. Availability of public transport stops	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
13. Lifetime	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14. Population density	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15. The average salary for the region	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16. Ecological situation	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
17. Unemployment rate	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18. The availability of an online pharmacy	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19. Health problems	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20. Assortment	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21. The presence of shopping centers, shops	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

To obtain a forecast of the situation on the basis of the constructed cognitive map, we use the impulse processes method was used [2, 4].

The impulse process allows to determine the state of concepts at discrete instants of time, analyzing which the expert receives a forecast of changes in the state of the system when implementing various control strategies and actions from the external environment.

The vertices are represented by the set  $u_1, u_2, \dots, u_n$ .

It is assumed that each vertex  $u_i$  takes the value  $v_i(t)$  at discrete instants of time  $t = 0, 1, 2, \dots$

The change  $p_j(t)$ , called the momentum, is given by the difference  $v_j(t) - v_j(t-1)$ , for  $t > 0$ . For  $t > 0$ , we define the impulse process:

$$v_i(t+1) = v_i(t) + \sum_{j=1}^N s(u_j, u_i) \cdot p_j(t) \quad (2)$$

Let's consider two situations.

The first is when the pharmacy is just planned. Here the factors that can be controlled when choosing a location for a pharmacy are:

2. The quality of the pedestrian flow.

3. Purchasing power.
4. Type of district.
5. Competitive environment.
6. Distance to medical institutions.
8. Traffic.
11. Number of storeys.
12. Availability of public transport stops.
21. The presence of shopping centers, shops.

In this case, the choice of a place for a new pharmacy occurs, based on the advantage of the above factors.

It is impossible to influence macroeconomic factors. These include:

13. Lifetime.
14. Population density.
15. The average salary for the region.
16. Ecological situation.
17. Unemployment rate.
19. Health problems.

The remaining factors are manageable, but at the opening stage they are the basic rules and only in the case when the scenario of the pharmacy operation is violated there is a need to influence them.

The second situation is when the pharmacy is already open and there is a need to influence the amount of its profit. In this case, the controlling factors are:

7. Advertising companies.
9. Qualification of staff.
10. Marketing strategy.
18. The availability of an online pharmacy.
20. Assortment.

By influencing them, you can change the values of other factors and, as a result, the level of potential demand. When studying the simulation of the vertex parameter dynamics in the case of interaction of the system without external pulses for  $t = 7$ , where  $t$  is the study period, the initial values of the vertex parameters  $v_i = 0$ , obtained the following dynamics.

Vectors of initial impulses for the first situation (the pharmacy is planned to open):

$$\begin{aligned}
 p_2 &= (0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0) \\
 p_3 &= (1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0) \\
 p_4 &= (0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0) \\
 p_5 &= (0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0) \\
 p_6 &= (0,0,0,0,0,0,0,0,0,-1,0,0,0,0,0,0,0,0,0,0) \\
 p_8 &= (1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0) \\
 p_{11} &= (0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0) \\
 p_{12} &= (0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0) \\
 p_{21} &= (0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0)
 \end{aligned}$$

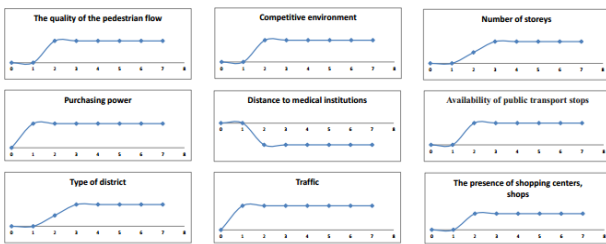


Figure 2. Dynamics of the target vertex (1) under the condition of the vertex pulse (2), (3), (4), (5), (6), (8), (11), (12), (21)

Figure 2 shows a positive effect of the factors (2), (3), (4), (5), (8), (11), (12), (21) and the negative factor (6) on the target vertex (1) - "level of potential demand" - during the period under study.

The stabilization of the target indicator occurs from the first period when the vertices (3) and (8) are affected, from the second period when the vertices (2), (5), (6), (21) are acted on, and (3) (4) ).

Vectors of initial impulses for the second situation (the pharmacy is already open):

$$\begin{aligned}
 p_7 &= (0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0) \\
 p_9 &= (1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0) \\
 p_{10} &= (1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0) \\
 p_{18} &= (1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0) \\
 p_{20} &= (1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
 \end{aligned}$$



Figure 3. Dynamics of the target vertex (1) under the condition of the vertex pulse (7), (9), (10), (18), (20)

The factors (7), (9), (10), (18), (20) have a positive effect on the target vertex (1) - during the period under study. The stabilization of the target indicator comes from the first period when it affects the vertices (9), (10), (18), (20). When influencing the top (7) from the first period, there is a negative trend, but already from the second period - positive, and stabilized in the third period.

On the basis of the results of two cases, we can say that the impulse processes in the weighted digraphs are stable and the model is suitable for analyzing the processes and predicting the results of these processes.

Based on the results of the simulation experiment based on the cognitive model, it is possible to develop control strategies for the relevant factors to achieve a positive result.

## CONCLUSIONS

Studying the behavior of the model by means of impulse processes, the expert receives a forecast of the development of the situation under the conditions of his chosen strategy, which allows to evaluate the consequences of the decisions made and to choose the most optimal strategy.

In the future, the model can be represented as a fuzzy map. And a decision support system based on fuzzy logic is built.

## REFERENCES:

- [1] A.V. Bakurova, Self-organization of socio-economic systems: model and methodology: monograph. Zaporizhzhya: KPU, 2010.
- [2] D.V. Yerokhin, "Modeling an innovative enterprise mechanism using fuzzy cognitive maps", *Management in Russia and abroad*, no. 3, pp. 95-111, 2006.
- [3] V.I. Maksimov, "Structurally-objective analysis of the development of socio-economic situations", *Management issues*, no. 3, pp. 30-38, 2005.
- [4] F.S. Robertc, *Discrete mathematical models with applications to social, biological and ecological problems*. M.: Nauka, 1986.
- [5] V.B. Silov, *Adopting strategic decisions in a fuzzy environment: in politics, in macroeconomics, sociology, management, medicine, economics*. M.: INPRO-RES, 1995.
- [6] R.M. Axelrod. *The Structure of Decision: The Cognitive Maps of Political Elites*. Princeton: University Press, 1976.



# Information Technology for Quality Management of Engineering Products

Evgeniy Lavrov<sup>1,2</sup>, Mykola Bahmach<sup>1</sup>, Galyna Mihalevska<sup>3</sup>

<sup>1</sup>Sumy State University, Ukraine, <sup>3</sup>Khmelnitsky National University, Ukraine

<sup>2</sup>prof\_lavrov@hotmail.com, <http://www.famous-scientists.ru/13333>

**Abstract** – Development of tools for improvement the functioning of the overall production process and the technical control department, which consists of building a model of the production process, calculation the necessary performance and reliability of its operation, developing software for automation of statistical quality control and stability of the production process.

**Keywords**—modeling, quality control, probability, automatical control, economic efficiency, damage minimization

## I. INTRODUCTION

One of the subsystems of the automated control system of machine building enterprise is the subsystem of quality management. As a rule, in most factories the quality problem is particularly acute in connection with the occurrence of significant percentage of defective products, as well as the complexity of the choice of economically justified measures aimed at improving quality. In most enterprises there is a Department of technical control, but the efficient organization of its activities requires precise planning with the use of qualimetry methods, mathematical analysis and statistics [1,2,3].

## II. PROGRAM COMPLEX FOR TECHNICAL DEPARTMENT OF QUALITY CONTROL

By the order of Lebedinsky plant of piston rings (Sumy region, Ukraine) it was started the work for creating informational technology for support of decision-making in quality management. Some aspects of the work described in [4, 5, 6]. The aim of this work is to complement [6] using the recent developments of the system and analyze the effectiveness of development.

### A. General characteristics

It was developed the software system for the Department of technical quality control, which is based on an existing mathematical tools of statistical analysis of the production process, for example [6,7], which allows calculations needed to support decision making on quality.

General characteristics of the software system and their basic functions (figure 1):

- determining the likelihood of defective products;
- construction of number allocation likelihood of defective products based on the average number per unit time of marriage;
- calculation of probabilities yield products for a given statistical parameters;
- calculation of statistical unknown parameters of the known data;
- implementation of statistical analysis of a sample of product parameters, calculate the mean and variance limits of access and control, histogram graphs superimposed with probability distribution function of the density of the normal law;
- construction of control charts (X-card and S-Cards);
- calculation of risk-producer and the customer at the receiving party quality control of finished products using various combinations of the control samples and risk graph depending on percentage of defective products.



Figure 1. The main window

**B. Subsystem «Acceptance quality control»**

Purpose: used to calculate the risk of the manufacturer and the customer during the final control, depending on the different parameters of the sample particles and permissible percentage of defective products.

Risk is a manufacturer in the sense of failure probabilities of receiving party products lack percentage of which is permissible. The risk of the customer is the meaning of the likelihood of unwanted party take percentage of marriage.

Figure 2 shows the main window of the subsystem.

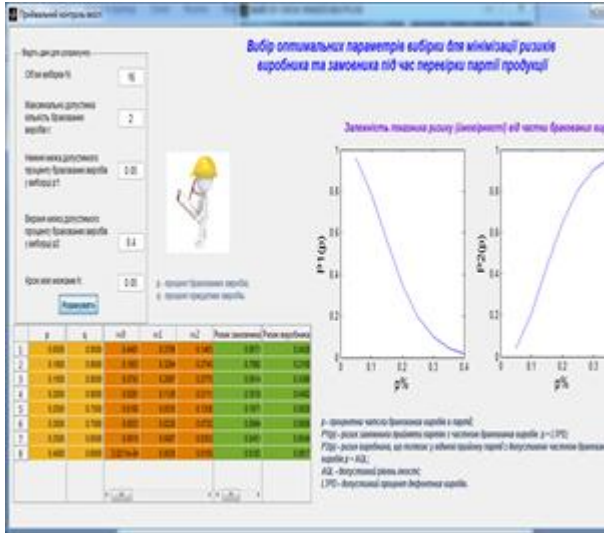


Figure 2. Main window of subsystem «Receiving quality control»

**C. Subsystem «Control for qualitative indicators».**

Purpose: construction of a number of the distribution of defective products for a certain period. We use statistics: the average number of defective products per unit of time and the maximum number of defective products for a certain period, calculated the probability that the number of defective products does not exceed a given. The main window of subsystem presented in fig.3.

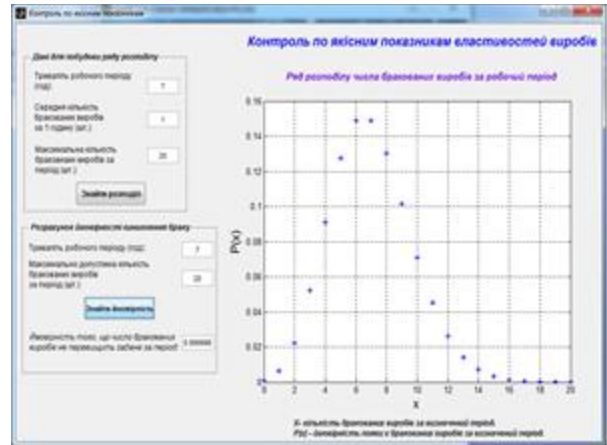


Figure 3. Main window of subsystem «Control of the Quality indicators»

**D. Subsystem «Control for quantitative indicators».**

Purpose: construction schedule probability of being the value set in parameter within a given standard deviation.

The main window of subsystem presented in figure 4.

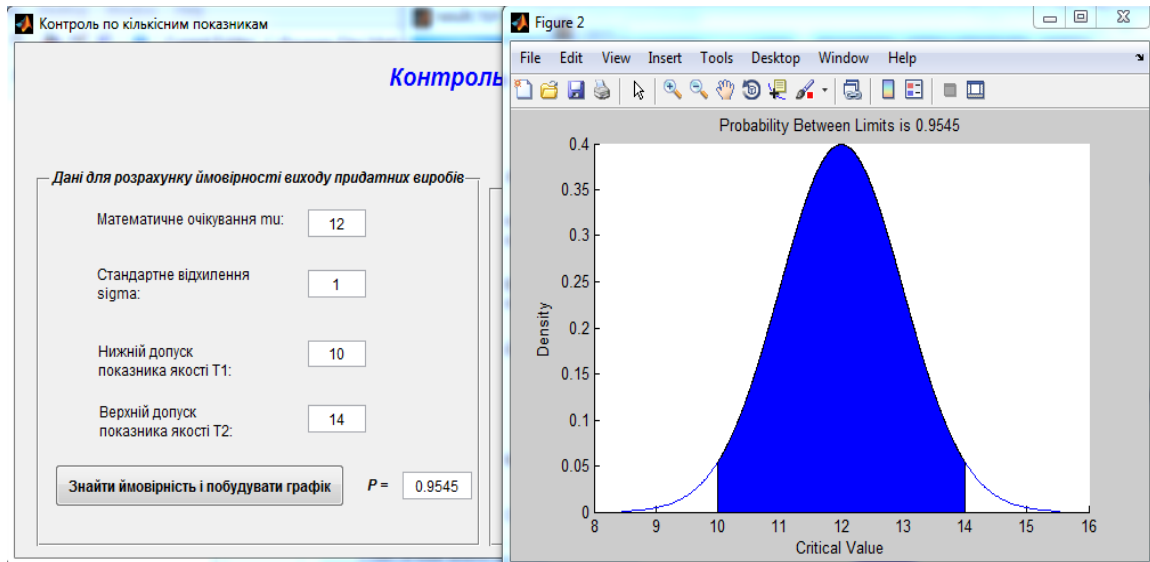


Figure 4. Main window of subsystem «Control for Quantitative indicators»

**E. Subsystem «Statistical analysis of sample».**

Purpose: statistical analysis of sample parameters, histogram superimposed on the graph of the probability

distribution density of the normal law. The main window of subsystem presented in figure 5.

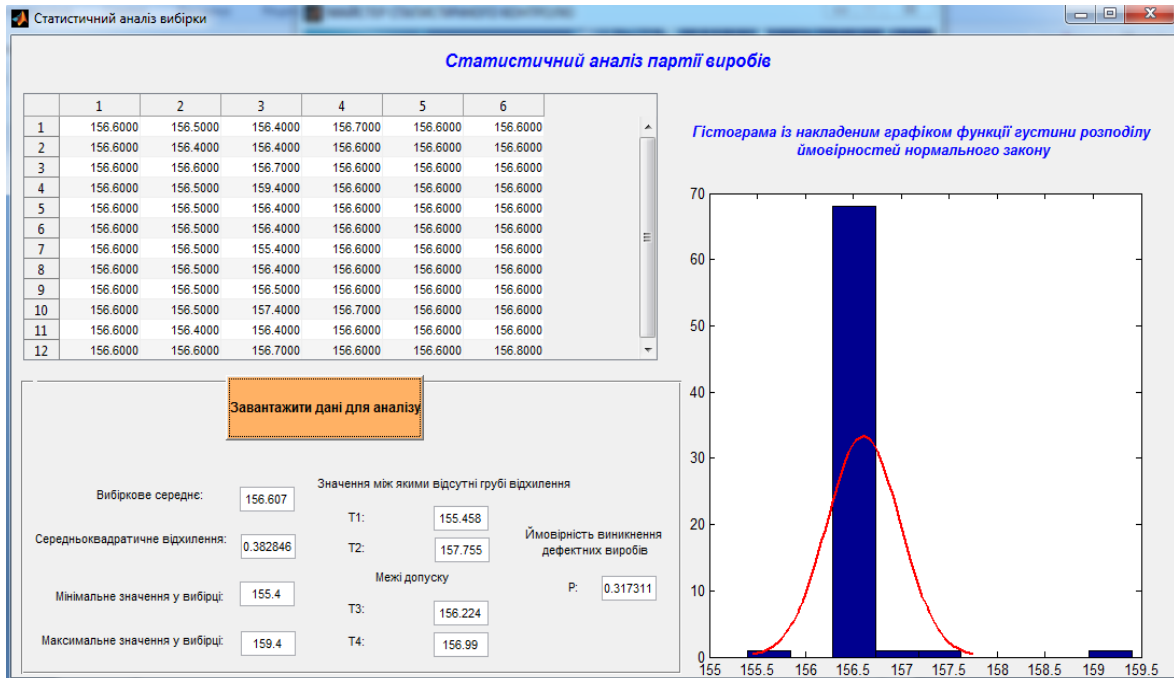


Figure 5. Main window of subsystem «Statistical analysis of the sample»

#### F. Subsystem “Control cards”

Purpose: building control arithmetic cards (X-Card), which is applied selectively arithmetic mean of the controlled parameter, building S-Card, that contains the value of sample standard deviation..

The purpose of control charts - find abnormal changes in the data for processes that are repeated and give criteria to identify lack of statistical control. The process is statistically controlled state, if only due to random variability causes. After determining that the received variability of any deviation from it considered the result of special reasons, which should identify, eliminate or reduce their impact.

The using of the control cards and careful analysis will lead to better understanding and improvement process, minimizing losses marriage.

The problem of statistical process control - maintenance and support processes at a reasonable and stable level that guarantees the products and services specified requirements.

The main window subsystem presented in fig. 6.

#### G. The simulation subsystem of production processes

For building of functional models was used the software package «Computer simulation technology of discrete human-machine interaction» [7,8], which is included as a subsystem in the software package.

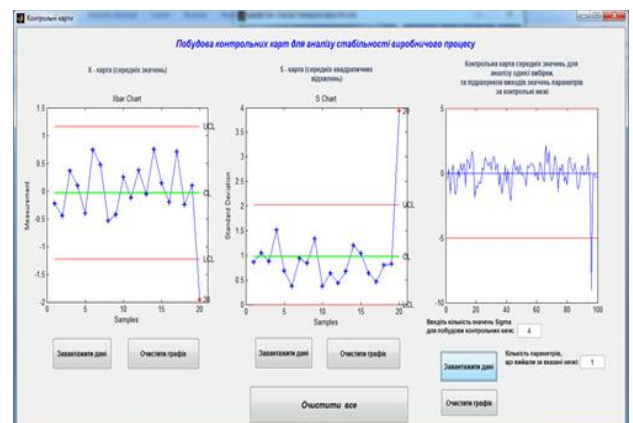


Figure 6. Main window of subsystem «Control cards»

The main functions are automated implementation of complex tasks of assessment, calculation of the probability of timely and error-free performance algorithms describing the production process.

The input data of the subsystem may be the results produced by the other subsystems (see 1.2.-1.6.).

Thus it is possible to compare the probability of faults for possible production technologies.

Figure 7 presents diagrams of alternative processes of production (existing and modified).

Figures 8-9 present indicators of the quality of work execution and inspection operations (for the modified process).

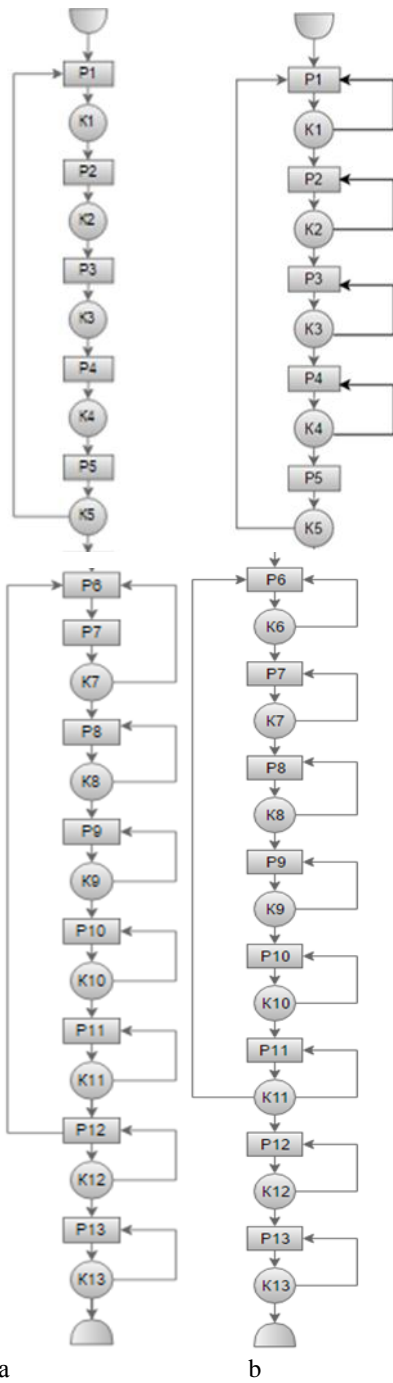


Figure 7. Structure of production determination [9], (a – modified, b – existing)

Operation	Settings		
	Probability $B^1$	Time	
		M, m	D, m <sup>2</sup>
P1	0,998	5	1
P2	0,988	4	0,3
P3	0,989	3	1
P4	0,960	6	0,7
P5	0,990	13	0,9
P6	0,991	8	1
P7	0,973	9	0,89
P8	0,950	15	2
P9	0,950	25	2
P10	0,931	20	3
P11	0,934	10	1
P12	0,951	20	0,5
P13	0,990	15	0,7

Figure 8. Quality indicators of operations performance

Operation	Settings		
	Probability $B^1$	Time	
		M, m	D, m <sup>2</sup>
P1	0,998	5	1
P2	0,988	4	0,3
P3	0,989	3	1
P4	0,960	6	0,7
P5	0,990	13	0,9
P6	0,991	8	1
P7	0,973	9	0,89
P8	0,950	15	2
P9	0,950	25	2
P10	0,931	20	3
P11	0,934	10	1
P12	0,951	20	0,5
P13	0,990	15	0,7

Figure 9. The quality indicators of control operations

Options:

- $B^1$  – the probability of error-free performance;
- $K00$  – his probability of an error in its presence;
- $K11$  – probability of detecting the absence of errors in its absence;
- M – expectation runtime;
- D – dispersion runtime.

After a complete description of the functional model network, it is necessary perform the reduction (convolution) for identifying typical structures in the general algorithm and for calculating of reliability parameters of algorithm functioning.

Figure 10 shows the results of the upgraded reduction algorithm.

### III. ANALYSIS OF THE EFFECTIVENESS OF THE PROPOSED TECHNOLOGY

Figure 11 shows the result of the one of numerical experiments, characterizing the influence of the structure of the production process in production losses.

Calculation results						
Model	The probability of error-free execution of the algorithm	Expectation runtime of the algorithm	The dispersion time of the algorithm	The probability of timely execution of the algorithm	The probability of timely and error-free execution of the algorithm	The specified time directive, $T_0$ :
existing technology	0,92064543	239,600625	5276,781	0,9995325	0,920215	480
upgraded technology	0,98854995	308,70947	4719,321	0,9936738	0,980891	480

Figure 10. The results of the evaluation of alternative production methods

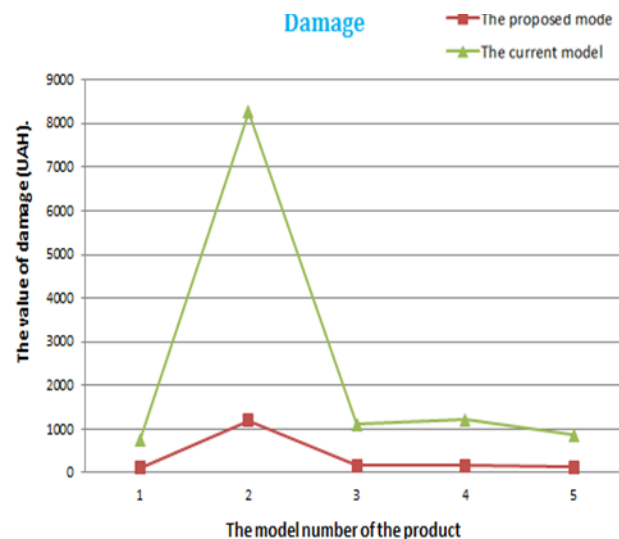


Figure 11. Comparison of possible damage when using the proposed models

### CONCLUSION

The system passes test operation at Lebedinsky plant of piston rings.

Full-scale implementation of the system will solve the problem of assessment of the likelihood of violations of the process, optimize production processes, choice points, methods and techniques of control measures to ensure high quality products.

The results may be useful for engineering enterprises with discrete production.

The use of the developed information technology of quality management contributes to sound decision-making in quality management systems based on objective quantitative and qualitative indicators

### REFERENCES:

[1] A. I. Gubinsky, M. G. Grif, and E. B. Tsoy, "About some of the optimization algorithms of the system "man-equipment", in *Application of computers for optimal planning and design*. Novosibirsk: NSU, 1981, pp. 148-154.

[2] E. S. Ventzel and D. Ovcharov, *A. probability Theory and its engineering applications*. M.: Nauka, 1988.

[3] H. Kume, eds., *Statistical methods for quality: trans. from English*. M.: Finance and statistics, 1990.

[4] E.A. Lavrov, N.V. Bahmach, A.S. Krivodub, "The approach to modeling production and quality control of the production process production plant at Lebedinsky piston rings", in *Proceedings of VI International Scientific Conference of Young Scientists "Information technology: economy, technology, education materials"*, 2015, pp.106.

[5] E.A. Lavrov and M.V. Bahmach, "Formalized description of the process plant at Lebedinsky piston rings for quality control problems", in *Proceedings of scientific and technical conference "Computer science, mathematics, engineering, IMA"*, 2016, pp.90.

[6] M.V. Bahmach and E.A. Lavrov, "Program complex of statistical calculations for the quality control of products at Lebedinsky plant of piston rings", in *Proceedings of VI International Scientific Conference "Modern information technology systems and materials, AIST"*, 2016, pp.82.

[7] E.A. Lavrov and N.B. Pasko, "Computer simulation technology of discrete human-machine interaction", *Certificate of registration of copyright "computer program"*, № 45262, 21.08.2012.

[8] E. Lavrov and N. Pasko, "Ergonomics of the of flexible systems "man-computer", in *Proceedings of International Scientific Conference «UNITECH 10»*, 2010, pp. 354-359.

[9] Gubinsky A.I. *Reliability and quality of functioning ergatic systems*. Leningrad: Nauka, 1982.

# Introduction of an irregular grid with respect to the spatial coordinate for the method of lines

Olga Dmytriieva<sup>1,2</sup>, Nadiia Huskova<sup>1,3</sup>

<sup>1</sup>Donetsk National Technical University, Ukraine

<sup>2</sup>Research Centre for Simulation Technology (Simtech), Germany

<sup>3</sup>Higher Technical School of the University of Applied Sciences, Germany

<sup>2</sup>olha.dmytriieva@donntu.edu.ua, <http://fknt.donntu.edu.ua/pmi>

<sup>3</sup>huskovanadiia@gmail.com, <https://www.th-bingen.de/person/nadiia-huskova/>

**Abstract** – The paper deals with the problem of reducing evolutionary partial differential equations to systems of ordinary differential equations with discretization over space. It is assumed that the obtained systems will be implemented in parallel using the method of lines. The questions devoted to the parallel control of the step of time integration on the basis of collocation block methods are considered. For the spatial coordinate, an irregular grid with a Chebyshev arrangement of nodes is introduced, which makes it possible to improve the accuracy of the results without significantly increasing the computational complexity. The obtained results are confirmed by computer experiments for partial parabolic partial differential equations with different types of boundary conditions and stiffness parameters.

**Keywords** – irregular grid, direct method, Cauchy problem, Chebyshev nodes, parallel block methods

## I. INTRODUCTION

One of the main questions of developing numerical methods for modeling dynamic objects is the possibility of the implementation in parallel computing systems. The widespread use of parallel computers has created an urgent need for effective methods aimed at solving problems with special properties: rigid, ill-conditioned, rapidly oscillating. If earlier mathematical models describing the dynamics of such systems took into account only the most significant factors, displaying individual aspects of the system, then with the development of information technology, the emergence and distribution of parallel computer systems, it was possible to significantly accelerate the process of obtaining results, and therefore, to create more complex models that allow increase the accuracy of the solution by taking into account additional factors. At the same time, any dynamical system that is modeled by systems of ordinary differential equations (SODEs) or partial differential equations and has physical components with strongly different time constants leads to a rigid problem, it becomes necessary to provide control of the integration step.

In this paper we consider issues of increasing the efficiency of parallel computing by developing new and restructuring well-known methods of parallelizing the computational processes of modeling complex dynamic systems. The proposed methods can be used for parallel modeling of dynamic objects with lumped parameters whose behavior is described by the Cauchy problem for SODEs, as well as objects with distributed parameters for which descriptive partial differential equations can be discretized and reduced to SODEs using the method of lines.

This paper is a continuation of the research [1-2], connected with the numerical solution of evolutionary partial differential equations by the method of lines, which is a semi-discrete method of discretization with respect to spatial variables. After reducing the original partial differential equation to the Cauchy problem described by a system of ordinary differential equations, the solution of the latter can be obtained in parallel using collocation block methods that support the parallel control of the step of integration ( $\tau$ -refinement).

It is known [3] that the formation of a system of grid nodes has a significant effect on the quality of the solution of a partial differential equation by difference methods. A high density of grid points is desirable in areas with complex geometry or with a high gradient. The distribution of grid nodes in physical space is established initially and usually does not change in the course of calculations [4]. However, sometimes it makes sense to restructure the grid after performing a certain number of steps [5]. In particular, it is initially difficult to establish a suitable grid in a situation where the location of particular sites is not known in advance. In this case, the best results are obtained by using adaptive grids, which can be transformed as the solution changes.

## II. SELECTION OF THE NODES FOR AN IRREGULAR GRID

The need for adaptive grids also arises when using the semi-discrete method of straight lines with discretization over spatial variables. As the initial evolution equation, we consider a one-dimensional parabolic partial differential equation

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} + f(x, t), x \in [x_0, L], t \in [t_0, T], \quad (1)$$

with the initial and the boundary conditions. When reducing partial differential equations to the method of lines, as a rule, a step in space is chosen to be regular. In the case of boundary conditions of the first kind, such a choice is justified. However, if we observe the behavior of the error in solving equation (1) for the case of boundary conditions of the second or third kind, we can see that the maximum error value is observed when the boundary conditions are approximated. This fact can be explained by objective reasons connected with different order of approximation with respect to the spatial variable,  $O(h^2)$  by the basic scheme and  $O(h)$  by boundary conditions. In this case, a simple step-by-step reduction in space leads to a significant increase in the nodes of the computational grid and, as a consequence, to a significant increase in the dimension of the SODEs when the partial differential equation is reduced for solving by the method of lines.

In this paper, it is suggested that when constructing an ODE system instead of a grid with a constant step in space

$$x_i = x_0 + ih, i = 0, 1, 2, \dots, n, h = \frac{L}{n-1}, \quad (2)$$

use an irregular grid (3) whose interior nodes are introduced as the roots of the Chebyshev polynomial, the so-called Chebyshev interpolation nodes [6]. For simplicity of exposition, the computational domain is chosen to be rectangular, although the considered region in physical space can have an arbitrary shape. An auxiliary variable  $\xi \in [-1, 1]$  is introduced and a one-to-one correspondence is established between the newly entered variable and  $x \in [x_0, L]$  by correlating

$$x = \frac{(x_0 + L)}{2} + \frac{(L - x_0)}{2} \xi. \quad (3)$$

Then the original uniform arrangement of nodes  $x_i \in [x_0, L]$  can be associated with points

$$x_i = \frac{(x_0 + L)}{2} + \frac{(L - x_0)}{2} \xi_i$$

where

$$\xi_i = -\cos\left(\frac{2i+1}{2n+2}\pi\right), i = 1, 2, \dots, n-1.$$

Those for the case under consideration on the solution interval with respect to the spatial variable  $x \in [x_0, L]$ , internal nodes can be introduced by (3).

### III. FORMATION OF A SYSTEM OF ORDINARY DIFFERENTIAL EQUATIONS WITH NON-EQUIDISTANT NODES

It was shown in [6] that the choice of such nodes minimizes the maximum error in solving the interpolation problem and proves the best possible error. If we use this

idea to solve evolutionary partial differential equations using the method of lines, we can try to reduce the approximation error in the spatial variable without increasing the dimension of SODEs. The main issue here will be the formation of a system of ordinary differential equations with non-equally spaced nodes (fig. 1) over a spatial variable.

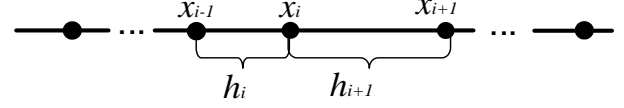


Figure 1. Scheme for introducing nonuniform knots with respect to a spatial variable

For this purpose, one can use the approach proposed in [7] and is used in the case of the curvilinear boundary of the solution search area. The proposed interpolation options can be used not only on the boundaries, but also on the entire uneven grid. If we are guided by the arrangement of nodes (fig. 1), then the difference representation of the second-order derivative takes the form

$$\frac{\partial^2 u}{\partial x_i^2} \approx \frac{1}{\tilde{h}_i} \left( \frac{u_{i+1} - u_i}{h_{i+1}} - \frac{u_i - u_{i-1}}{h_i} \right), i = 1, 2, \dots, n-1$$

where

$$\tilde{h}_i = \frac{h_{i+1} + h_i}{2}.$$

Then the original matrix of the right-hand sides of the system with a regular step (2)

$$\frac{a^2}{h^2} = \begin{pmatrix} -2u_1 & u_2 & 0 & 0 & 0 \\ u_1 & -2u_2 & u_3 & 0 & 0 \\ 0 & u_2 & -2u_3 & u_4 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & u_{n-2} & -2u_{n-1} & u_n \\ 0 & \dots & \dots & u_{n-1} & -2u_n \end{pmatrix}$$

with an irregular arrangement (3) of nodes must be transformed. In this case, the total number of nodes, and, consequently, the dimension of the system of differential equations remains unchanged.

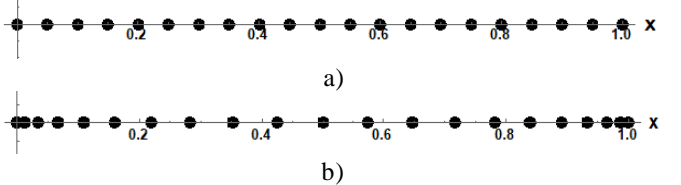


Figure 2. Formation of a system of nodes for the method of lines with respect to the spatial coordinate, a) a regular grid, b) Chebyshev location

In fig. 2, a) the variants of the uniform arrangement of nodes are given for the choice of a constant step in space based on (2), in fig. 2 b) an irregular mesh with Chebyshev arrangement of nodes (3) is given.

### IV. TEST IMPLEMENTATION ON AN IRREGULAR GRID

*Test problem 1.* The choice of the Chebyshev arrangement of nodes can be demonstrated by the example of the following homogeneous problem [8]

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2}, x \in [x_0, L], t \in [t_0, T] \quad (4)$$

with the values of the parameters  $L=1, T=1, a=1$  with the initial condition

$$u(x, 0) = \cos\left(\frac{\pi x}{L}\right), \quad (5)$$

and the boundary conditions of the second kind

$$\frac{\partial u(0, t)}{\partial x} = \frac{\partial u(L, t)}{\partial x} = 0, \quad (6)$$

the exact solution of which is known

$$u(x, t) = e^{-(\pi^2/L^2)t} \cos\left(\frac{\pi x}{L}\right).$$

and will be for determine the global error.

Let's introduce a discretization with respect to a spatial variable and fix the dimension of SODEs ( $n=21$ ). Since the boundary conditions in the original problem (4) are given in the form of the first derivatives with respect to the space (6). The order of approximation of which is  $O(h)$ , and for internal nodes this order is  $O(h^2)$ , it makes sense to proceed to the uneven distribution of nodes i.e. Reduce the step along the spatial variable as you approach the boundaries of the region and increase to the center (fig. 2, b). A numerical solution of problem (4) with the initial and boundary conditions (5) - (6) is shown in fig. 3.

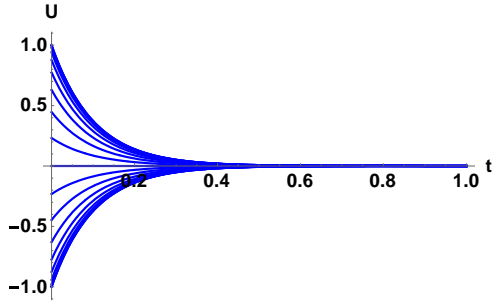


Figure 3. Numerical solution of the test problem (4) with conditions (5) - (6) by the method of lines on an irregular grid

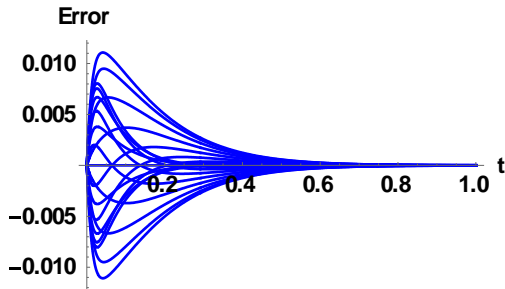


Figure 4. Graphs of the global error in the numerical solution of the problem (4-6) by the method of lines on an irregular grid

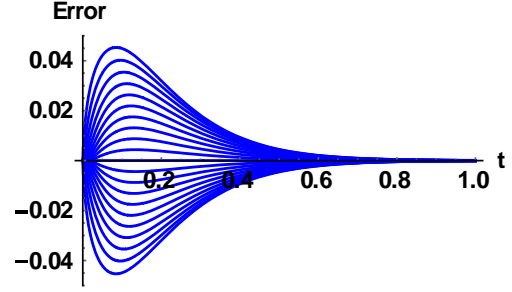


Figure 5. Graphs of the global error in the numerical solution of the problem (4-6) by the method of lines on a regular grid

It should be noted that in this experiment the time control procedure for the variable time was not used, since the questions of correlating errors on uniform grids and meshes with Chebyshev arrangement of nodes were of practical interest. The resulting global error in the numerical solution on the irregular grid of problem (4) with the conditions (5-6) is shown in fig. 4, the global error obtained on a regular grid is shown in fig. 5.

When examining the error graphs in detail, it can be seen that the largest values of the distance of the numerical solutions obtained by the direct method from the exact ones are observed just at those grid points that correspond to the boundary ones and close to them. As far as distance from the boundary nodes, the errors in the numerical solution become damped. The introduction of an irregular Chebyshev grid made it possible to reduce errors precisely at nodes close to the boundaries.

*Test problem 2.* The use of an irregular grid with respect to a spatial coordinate can significantly improve the solution also in the case of using the time control procedure for the variable time. Moreover, such improvements are also observed in the case of boundary conditions of the first kind. Consider the problem (4) with the initial condition

$$u(x, 0) = \sin\left(\frac{\pi x}{L}\right) \quad (7)$$

boundary conditions of the first kind

$$u(0, t) = u(L, t) = 0 \quad (8)$$

An exact solution of this problem is known [8]

$$u(x, t) = e^{-(\pi^2/L^2)t} \sin\left(\frac{\pi x}{L}\right)$$

A numerical solution of problem (4) with the initial and boundary conditions (7) - (8) is shown in fig. 6.

We leave the same dimensionality of the problem with the location of the nodes along the spatial coordinate shown in fig. 2, b. We form a system of ordinary differential equations and solve it with the help of the method of lines. The obtained graphs of the global error in the numerical solution of the problem (4) with the initial and boundary conditions (7) - (8) with an irregular arrangement of nodes are shown in fig. 7. The global error graphs obtained on a grid with a uniform step over the spatial variable are shown in fig. 8. Here, unlike the



previous problem with boundary conditions of the second kind, the discrepancy between the values of the global error is not so significant, but nevertheless, it is present. Thus, when conducting a comparative analysis, one can speak of the advantages of the approach that forms the solution on irregular grids.

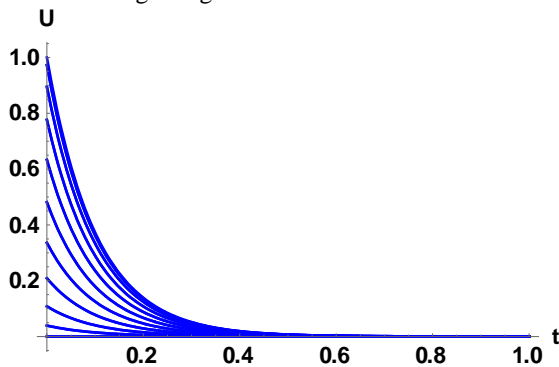


Figure 6. Numerical solution of the test problem (4) with conditions (7) - (8) by the method of lines on an irregular grid

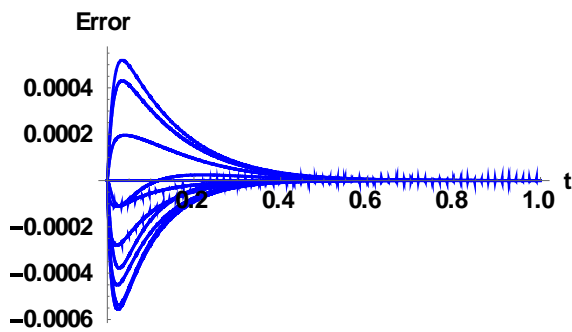


Figure 7. Graphs of the global error in the numerical solution of the problem (4), (7) - (8) by the method of lines on an irregular grid

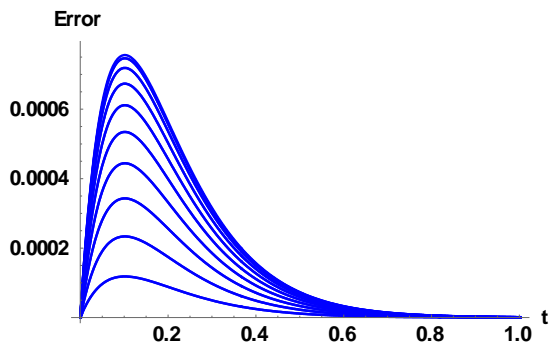


Figure 8. Graphs of the global error in the numerical solution of the problem (4), (7) - (8) by the method of lines on a regular grid

### CONCLUSIONS

The paper is devoted to the study of the problem of reduction of evolutionary partial differential equations to systems of ordinary differential equations with discretization in terms of a spatial variable. In contrast to the classical information schemes with a uniform step in operation, an irregular grid with Chebyshev arrangement of nodes is introduced for the spatial coordinate, which

makes it possible to improve the accuracy of the results without significantly increasing the computational complexity. In this paper, we consider the choice of nodes for an irregular grid, the formation of a system of ordinary differential equations with non-equally spaced nodes, the test implementation of SODEs on the introduced system of nodes for the spatial coordinate.

As the basic numerical method for solving the resulting system of equations, we propose the use of collocated block multistep methods. The main feature of these methods is the possibility of effective parallel implementation due to the block layout of the calculated and reference points, as well as the automatic adjustment of the integration step. Carried out computer experiments for known test equations have shown a significant advantage of the proposed approach for equations with boundary conditions containing derivatives and a slight improvement in the solution for boundary conditions of the first kind.

The theoretical assumptions given in the paper are supported by experimental studies that were carried out on partial parabolic test equations with different types of boundary conditions and stiffness parameters. The presence of exact solutions of the test equations allowed us to estimate the real values of global errors.

### ACKNOWLEDGMENT

The material presented in this report is the result of research conducted at the Research Centre for Simulation Technology (Simtech) of the University of Stuttgart, Germany, and is focused on developing and validating methods for modeling large-dimensional dynamical systems with extended the area of stability possessing high indexes of parallelism.

The authors are extremely grateful to the director of the SimTech Institute, the president of the International Association of Applied Mathematics and Mechanics, Professor Wolfgang Ehlers, for many years of support and the opportunity to conduct research.

### REFERENCES:

- [1] O. Dmytriyeva, *Parallel simulating of dynamic objects with lumped parameters*, Kharkiv: Noulidg, 2014. (in Ukrainian)
- [2] O. Dmytriyeva, *Realization of the method of lines for partial differential equations by collocation block difference schemes*, in press.
- [3] S.Franco, F.Gaspar, M.Pinto and C.Rodrigo, "Multigrid method based on a space-time approach with standard coarsening for parabolic problems", *Applied Mathematics and Computation*, 2018, vol.1317, pp. 25 - 34.
- [4] W.Schiesser, *Method of Lines Analysis of Turing Models*, World Scientific Publishing Co., 2017.
- [5] V. Chandrakera, A. Awasthib, and S. Jayaraja, "Numerical treatment of Fisher Equation", *Procedia Engineering*, vol. 127, pp. 1256 – 1262, 2015.
- [6] V.Verzhbitskii, *Fundamentals of numerical methods*, Moscow: Direct-Medi", 2013. (in Russian)
- [7] A.Tikhonov and A.Samarskii, *Equations of mathematical physics*. Moscow: Scienc", 1977. (in Russian)
- [8] W.Schiesser, *Differential Equation Analysis in Biomedical Science and Engineering*, John Wiley & Sons, 2014.

# The Approximation Surface Review of the Multidimensional Target Function for Surrogate Optimization Problems

Ruslana Trembovetska, Volodymyr Halchenko, Volodymyr Tychkov<sup>1</sup>  
Cherkasy State Technological University, Ukraine, <sup>1</sup>tvvpanda@ukr.net

**Abstract** – The computational technology of metamodels construction is developed, using modern achievements in the experimental planning theory, intellectual data analysis and artificial intelligence, and are determined experimentally by regularities that allow to effectively carry out the metamodels construction. The metamodels construction is performed on numerical examples with a goal function, which depends on two variables. As an experiment plan, a point generator is used which fills the search space and in the process of implementation of which the Sobol LP $\tau$ -sequence is used. The numerical experiments results indicate the possibility of using the computational technology proposed for constructing RBF-metamodels for approximation, multidimensional objective functions with a rather complicated response surface.

**Keywords** – metamodel, computer experiment plan, response surface, neural network, surrogate optimization.

## I. INTRODUCTION

Optimum synthesis refers to progressive ways of solving the problems of new technical devices efficient designing. When using it, the source data for designing are the necessary design object characteristics, and the result is a variant of the design, which ideally has the optimal structure and technical parameters. Implementation of design such approach in most cases is not trivial. At the same time, special attention should be paid to cases that are characterized by the solution increased complexity. In such tasks a priori given technical devices initial characteristics are not individual indicators, but their totality. Examples of such characteristics can be pre-determined temperature distributions, electric, magnetic fields, mechanical deformations and voltages, which are specified in a set of control points located in the device working volume. The definition of such characteristics is most often the result of solving the boundary field problems in a differential or integral formulation. Numerical such problems solving by known methods is rather costly in

terms of computing and time resources. With their multiple solution in the optimal synthesis problems framework, resource requirements increase in catastrophic proportions, which leads to the practical implementation impossibility of the acceptable result search.

One of the solving this tasks class ways is surrogate optimization, which involves replacing the resource-intensive, complex for the computational process implementation, the objective function, formulated on the physical laws basis, into a less costly, so-called metamodel. Moreover, the metamodel is no longer obtained from physical considerations, but as a result of the data approximation that is the result of resource-intensive model calculations and performs the function of the reference constructing points an approximation model on the physical model. Thus, we solve the problem of unlimited computational resources increase in optimal synthesis problems solving.

The **purpose** of these studies is to effective computing technology create based on current advances in information technology, artificial intelligence, experiment planning theory, which allows metamodels constructing for surrogate optimization tasks in a multidimensional search space.

## II. FORMULATION OF THE PROBLEM

The metamodel construction solving three interrelated tasks involves: the computational experiment plan definition, the approximation model construction and the adequacy and informative verification of the metamodel resulting.

Consistently, accompanied by numerical examples, consider the each of these separate tasks solution. Because of the possible complexity of the hypersurface response topology in this study, it is expedient to use experiment planning non-classical methods, and computer methods to fill the multidimensional search space, which provide with a high probability homogeneous filling it with reference points, in which the resource-intensive target function values are then calculated.

## III. SELECTION OF COMPUTATIONAL METHODS

When experiment plan choosing among the possible options variety, the point generators that search space fill

and in the implementation process of which the Sobol LP $\tau$ -sequences are used should prevail. The following sequences properties, noted in [1-3]: the high falling probability into the search probes in the search area in the extremes points vicinity and overlaps of the target function response surface; weakly correlated main effects and effects of the factors interaction. In addition, the researchers found that the best number of uniform filling points of the multidimensional search space can be determined from the expression  $N = 2^z - 1$ , for which  $z=2,3,4,\dots$ , for example

$N=1,\dots,7,15,31,63,127,255,\dots$ . There is also a weak correlation between the main effects and the factors interaction effects [2, 3].

That is, the Sobol LP $\tau$ -sequences have the best of uniform point's distribution present properties in a unit hypercube than any other sequences science known. Therefore, the use of LP $\tau$ -sequences in planning a multivariate experiment to obtain regression models is also promising for solving problems surrogate optimization.

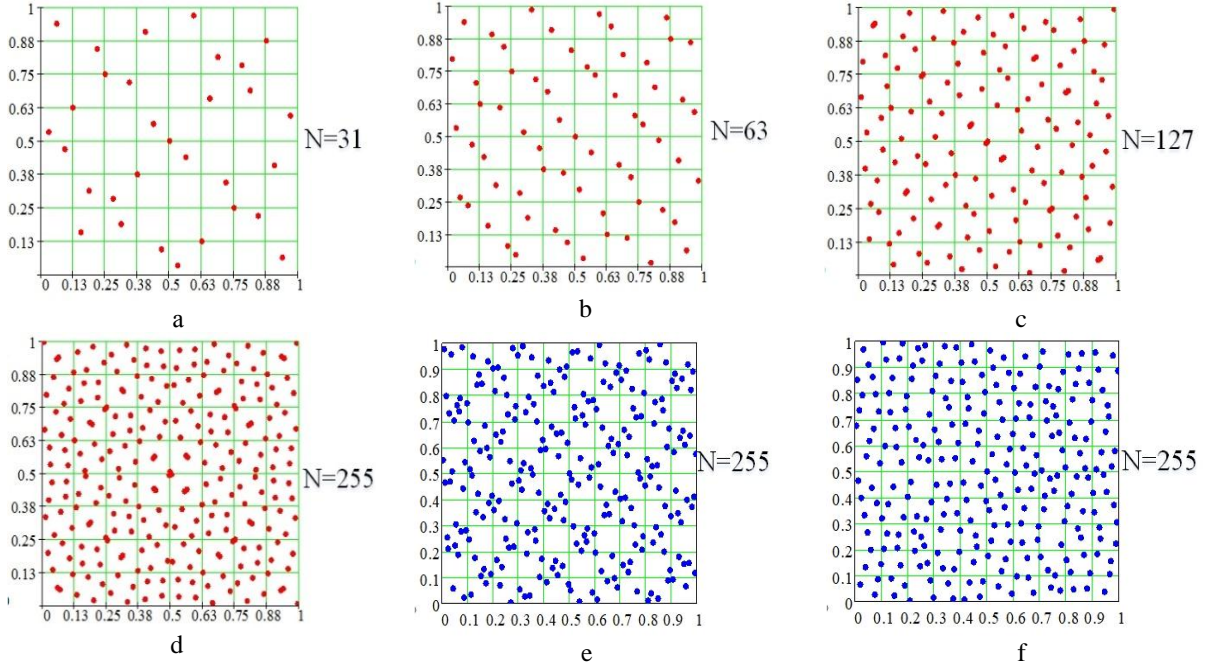


Figure 1. Generation of LP $\tau$ -sequences for search area filling with sensing points: a-d) sequence  $(\xi_1, \xi_2)$  with different point's number; e) sequence  $(\xi_1, \xi_5)$ ; f) sequence  $(\xi_1, \xi_7)$ .

“Fig. 1 a-d” shows the results of the Sobol LP $\tau$ -sequences use for different points number plans. With the Sobol LP $\tau$ -sequences help, we can get a variety of plans by combining variants sequence, for example,  $(\xi_1, \xi_5)$ ,  $(\xi_1, \xi_7)$  and td “Fig. 1d,e”. It is desirable to check the generated sequences correlation, but this procedure is not obligatory, since the weak generated LP $\tau$ -sequences correlation fact [1-3] is proved and sequences with avoided correlations moderate are revealed.

According to the authors, for approximation model constructing, the mathematical artificial neural networks device based on nuclear functions, namely radial-basic [4], is the most promising. This choice is due to the RBF-network fact has only one hidden neurons layer, which greatly simplifies the problem of the hidden layers characteristic number selecting of neural networks and makes this choice a definite and fast learning due to the using well-studied methods linear optimization possibility.

For further research, test goal functions were used:

$$\begin{aligned}
 f_1(x, y) &= x^2 + y^2 & x &\in [0;1], y \in [0;1] \\
 f_2(x, y) &= (0.5 + 0.5 \cdot x)^4 \cdot y^4 \cdot e^{[2 - (0.5 + 0.5 \cdot x)^4 - y^4]} & x &\in [-4;2], y \in [0.5;1.5] \\
 f_3(x, y) &= \left[ y - \left( \frac{5.1}{4 \cdot \pi^2} \right) \cdot x^2 + \left( \frac{5}{\pi} \right) \cdot x - 6 \right]^2 + 10 \cdot \left( 1 - \frac{1}{8 \cdot \pi} \right) \cdot \cos(x) + 10 & x &\in [-4;10], y \in [0;15]
 \end{aligned}$$

for which the function value was calculated at the experiment plan points.

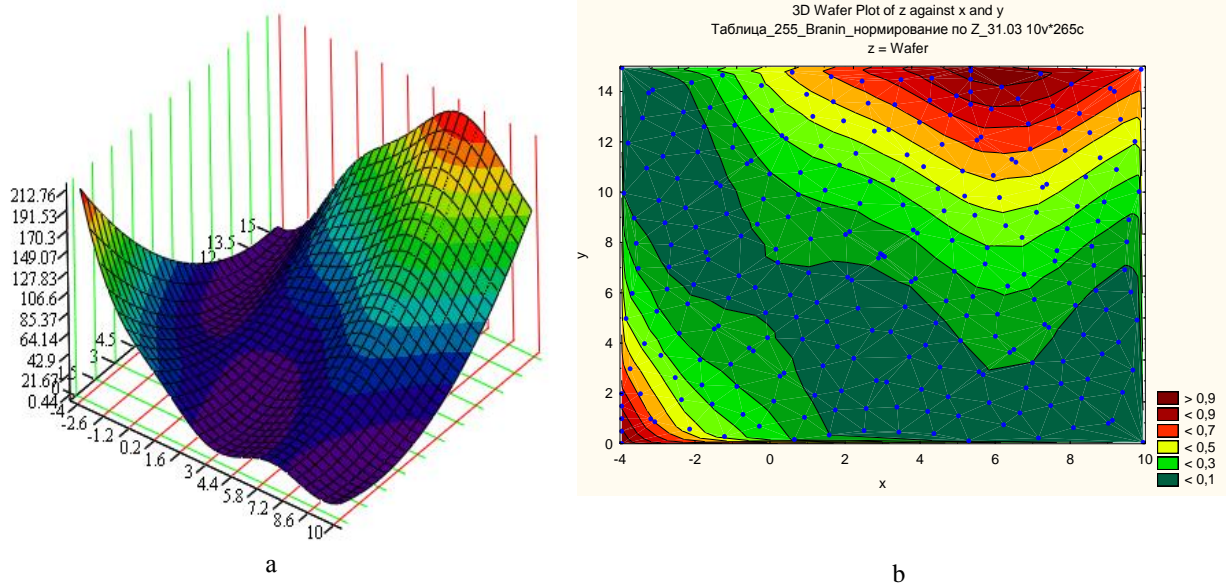


Figure 2. a) The function target  $f_3(x, y)$ ; b) The experiment plan  $N = 255$ , applied on the target function level line.

The obtained probing points coordinates and the calculated target function at these points values "Fig. 2b" constitute a output data table for the second stage implementation - the metamodel construction. For construction of RBF-metamodels, an automatic and user-defined construction strategy with random sampling is used in the following ratio: 70% - educational, 15% - control, 15% - test. If necessary to improve the received metamodels parameters, these ratios were changed by 80%, 10%, 10%, respectively. At the stage of neural networks training, best selection was carried out according to the indicators: coefficient

determination  $R^2$ , the standard forecast error deviations ratio and training data S.D.ratio, average relative model error magnitude MAPE,%, residual average error square MSR, histogram residual, diagram scattering. Neural networks are created for the plan  $N = 255$  with the hidden neurons number from 110 to 185. The following results are summarized in Table 1. The metamodel RBF-2-185-1 best under the number 6 of Table 1, for which the developed computing technology effectiveness on the residues histogram and the diagram scatter in Fig. 3 a, were evaluated.

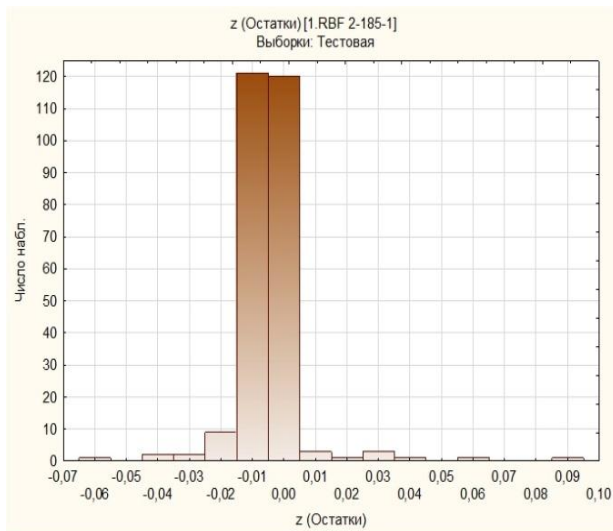
TABLE I. BEST RBF-METAMODELS.

Type	Metamodel	$R^2$ for educational, control, test sample	$S.D.ratio$	MAPE, %	$MS_R$
1	RBF-2-110-1	0.996; 0.993; 0.994	0,0975	0,47	0,00041
2	RBF-2-185-1	0.992; 0.989; 0.991	0,0405	6,6	0,000103
3	RBF-2-185-1	0.995; 0.993; 0.995	0,0411	3,4	0,000105
4	RBF-2-185-1	0.991; 0.99; 0.993	0,0413	10,94	0,000107
5	RBF-2-185-1	0.996; 0.989; 0.992	0,0441	7,1	0,000103
6	RBF-2-185-1	0.996; 0.992; 0.991	0,0401	3,6	0,000102

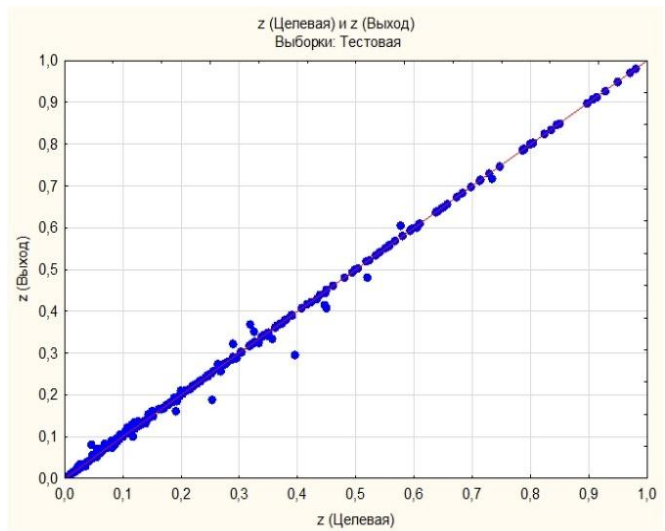
An important constructing a meta-model stage is to its adequacy check. In the process of its multi-step validation creation is carried out, the purpose of which is to many numerical indicators control obtained during the meta-model construction, including the neural network quality and evaluation of recovery with its response surface use. Adequacy is usually established by checking F-criterion of the hypothesis about the statistical insignificance of the adequacy dispersion

$\sigma_R^2$  and the experiments reproducibility results dispersion  $\sigma_D^2$ , according to which the mathematical model coefficients were obtained [3].

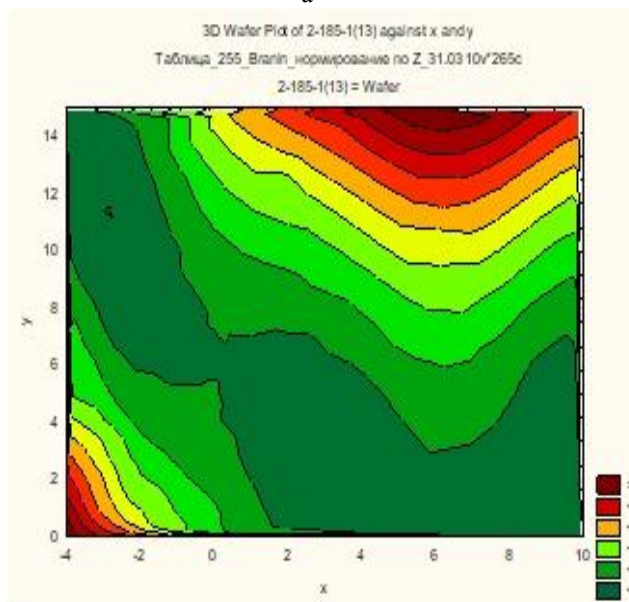
The informatively model verification is performed by correlation coefficient R multiplicity calculating and checking its significance statistical. The model is considered to be informative and meaningful in the level of significance for F-criterion  $p \leq 0,05$  (reliability  $\geq 0,95$ ).



a



b



c

Веса (Таблица_255_Branin_нормирование по Z_31.03)		
Веса ID	Соединения 1.RBF 2-185-1	Значения весов 1.RBF 2-185-1
1	x -> скрытый нейрон 1	0,367957
19	x -> скрытый нейрон 10	0,269142
199	x -> скрытый нейрон 100	0,865646
201	x -> скрытый нейрон 101	0,644467
203	x -> скрытый нейрон 102	0,383704
205	x -> скрытый нейрон 103	0,024270
207	x -> скрытый нейрон 104	0,944669
209	x -> скрытый нейрон 105	0,573317
211	x -> скрытый нейрон 106	0,458755
213	x -> скрытый нейрон 107	0,924949
215	x -> скрытый нейрон 108	0,597082
217	x -> скрытый нейрон 109	0,284975
21	x -> скрытый нейрон 11	0,205938
219	x -> скрытый нейрон 110	0,356111
221	x -> скрытый нейрон 111	0,770875
223	x -> скрытый нейрон 112	0,632621
225	x -> скрытый нейрон 113	0,518058
227	x -> скрытый нейрон 114	0,755056
229	x -> скрытый нейрон 115	0,083502
231	x -> скрытый нейрон 116	0,514086

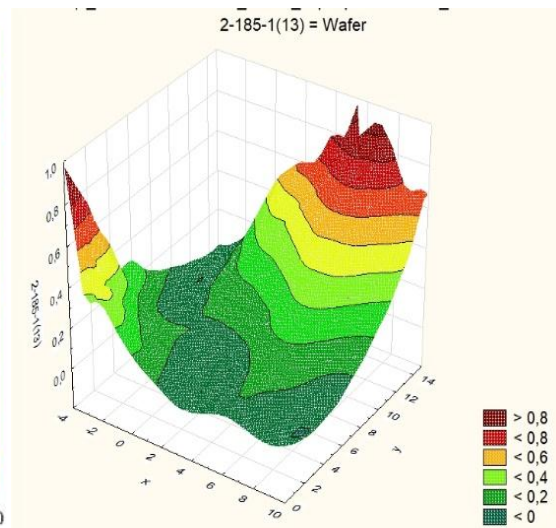
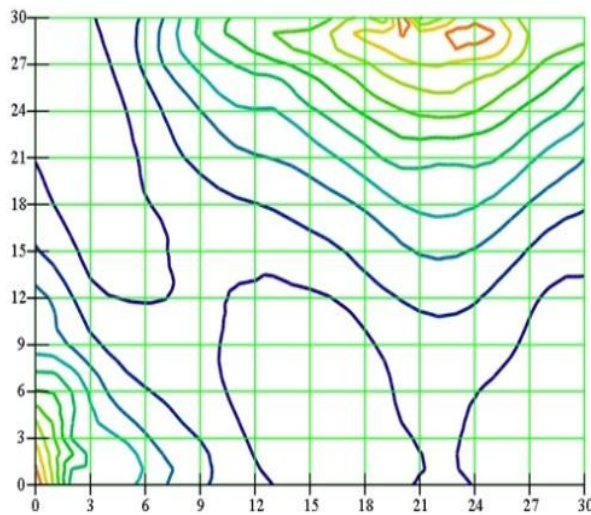
d

Figure 3. Neuronal network RBF-2-185-1 (sample N = 255): a) residues histogram; b) the diagram target values scattering and the function approximated; c) the level line of the response training sample points surface reproduced; d) the value parameters and weight created RBF-metamodel coefficients.

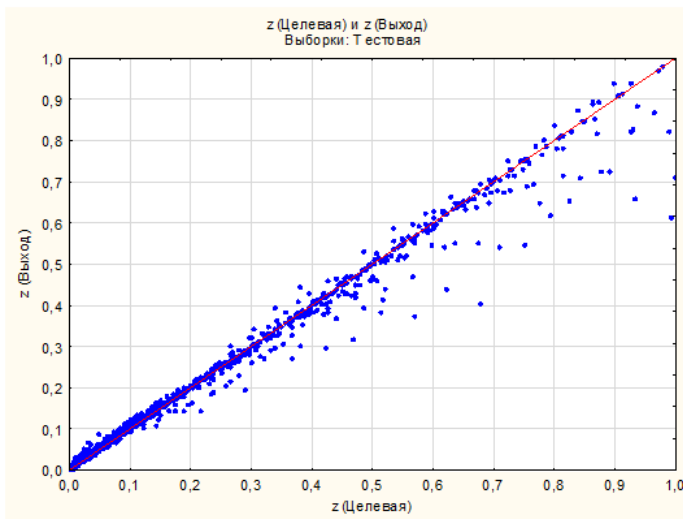
The response surface recovery evaluation is made using the formula that describes the neural network output and is formed as a linear outputs combination of the hidden layer neurons with the received source neuron coefficients with the k-th neuron of the layer hidden wk, the center k-th neuron coordinates  $C_{x_1k}, C_{x_2k}, \dots, C_{x_ik}$ , width k-th neuron  $a_k$  [4]. Some weight coefficients values are shown in "Fig. 3d". "Fig. 4a" shows the result of the response obtained surface restoration using the RBF-2-185-1 metamodel, executed throughout the range  $x \in [-4; 10]$ ,  $y \in [0; 15]$  with a step of 0,033, that is, at 961 points.

At the response surface reproduction stage, the received metamodel adequacy was evaluated according to the indicators: the average regression square MSD, the residual average error square MSR; reproducibility variability  $\sigma_D^2$ , adequacy variance  $\sigma_R^2$ ; reproducibility evaluation standard error  $S_D$ , adequacy estimation standard error  $S_R$ .

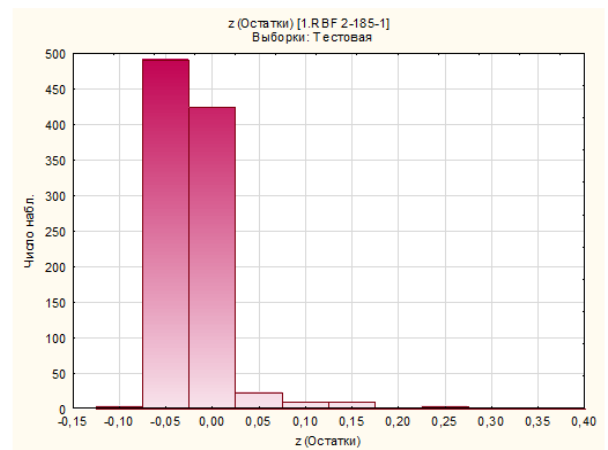
The reconstructed RBF-2-185-1 based on the metamodel has the following parameters:  $R^2=0.977$ ,  $MAPE=12,6\%$ ,  $MSR=0,00095$ ,  $S.D.ratio=0,131$ .



a



b



c

Figure 4. Response surface reproduction: a) level line and restored response surface 3D - graph; b) the diagram target and restored function values dispersion; c) residues histogram

### CONCLUSIONS

The numerical experiments results indicate the possibility of proposed computing technology using for RBF-metamodels constructing for the multidimensional target functions approximation with a rather response surface complicated.

### REFERENCES:

[1] I.M. Sobol and R.B. Statnikov, Eds., *The choice of optimal parameters in problems with many criteria*. Moscow: Drofa, 2006.  
 [2] S.G. Radchenko and O.V. Kozyr, "Experiment Plans for High Precision Models," *Mathematical Machines and Systems*, no. 2, pp. 117-127, 2014.  
 [3] S.G. Radchenko, *Methodology of regression analysis*, Kyiv: Korniychuk, 2011.

[4] S. Khaikin, Eds., *Neural networks: full course*, Moscow: Williams Publishing House, 2006.  
 [5] V.Ya. Halchenko and R.V. Trembovetska, "Construction of RBF-metamodels in the tasks of surrogate optimization", presented at 2<sup>nd</sup> All-Ukrainian Conference Theoretical and Practical Problems of Using Mathematical Methods and Computer-Oriented Technologies in Education and Science, Kiev. Un-t them. Boris Grinchenko, 2018, pp. 179 – 184.  
 [6] V.Ya. Halchenko, R.V. Trembovetska and V.V. Tychkov, "Approximation of the response surface by means of artificial intelligence", In Materials of the 2<sup>nd</sup> International Scientific and Practical Conference Modern Trends in the Development of Science'18, 2018, pp. 54-57.

# The Problems of Automation of Financial Transactions Subject to the Signs of Internal Financial Monitoring

Serhii Mynenko

Sumy State University, Ukraine, e-mail: minenserger@gmail.com

**Abstract** – In recent years, the risk of using a banking institution to legalize (launder) illegally-obtained income, terrorism financing or financing the spread of weapons of mass destruction has significantly increased in its volume. According to the State Financial Monitoring Service of Ukraine, only in the first quarter of 2018, 2,120,689 transactions were accounted for, which should be directed to the legitimization of income. Among them, 96.36% of transactions were presented on the grounds of obligatory financial monitoring, 3.05% on the grounds of internal financial monitoring and 0.59% of transactions combined the signs of mandatory and internal financial monitoring. These figures indicate the scale of use of banks for legalization of illegally-obtained income and insufficient attention to internal financial monitoring. Hence the problem of imperfection of methodological and software of internal financial monitoring occurs. Therefore, there is a significant number of risky transactions that remain beyond the attention of the banking systems of financial monitoring and, as a consequence, outside the attention of the state.

*Keywords* – automation, financial monitoring, counteraction to legalization of incomes, internal financial monitoring, bank business processes, business process modeling.

## I. INTRODUCTION

The negative effect of the legitimization of illegally-obtained income directly or indirectly is experienced by all subjects of economic activity [1]. The negative effect for the state first of all comes in the form of destabilization of the banking system and the national economy, for business this is reflected in the the complication of the conditions of functioning in an unfair competitive environment, for the population it is the reduction of social standards and the quality of life [2].

Since banks due to the specificity of their activities are at the very center of the process of the legitimization of illegally-obtained income, it is they who are responsible for identifying financial transactions that have signs of risk of legitimization of illegally-obtained income [3]. Therefore, banks need to have an effective

system of financial monitoring [4]. Due to the low level of detection of financial transactions that come under the signs of internal financial monitoring, special attention should be paid to the development of methodology, algorithms and software to implement a rapid and qualitative process for identifying financial transactions that come under the signs of internal financial monitoring [5].

To ensure the effective functioning of the system for identifying financial transactions subject to the signs of internal financial monitoring that would meet the requirements and challenges of the present, it is necessary to develop a reliable, efficient, easy-to-use automated information system.

## II. SYSTEM REQUIREMENTS

The system should:

- consider the criteria for the riskiness of operations in accordance with the risk-oriented approach and separately store them, indicating the reason for attributing such an operation to a risky one;
- take into account the risk criteria of clients on the risk-oriented approach and separately store them, indicating the reason for classifying a client as risky;
- display financial transactions assigned to risky for further analysis by the person responsible for financial monitoring by the bank's employee;
- provide an opportunity to exclude from the list of financial transactions classified as risky after the establishment of their non-transparency by the responsible employee of the bank [6].

## III. DECISION

Such a system should assess on a real-time basis the probability of belonging to a banking transaction as a subject to the risk of legitimization (laundering) of illegally-obtained income, terrorism financing or financing the spread of weapons of mass destruction [8].

Figure 1 shows the model of the process of identifying risk transactions that fall under the signs of internal financial monitoring.

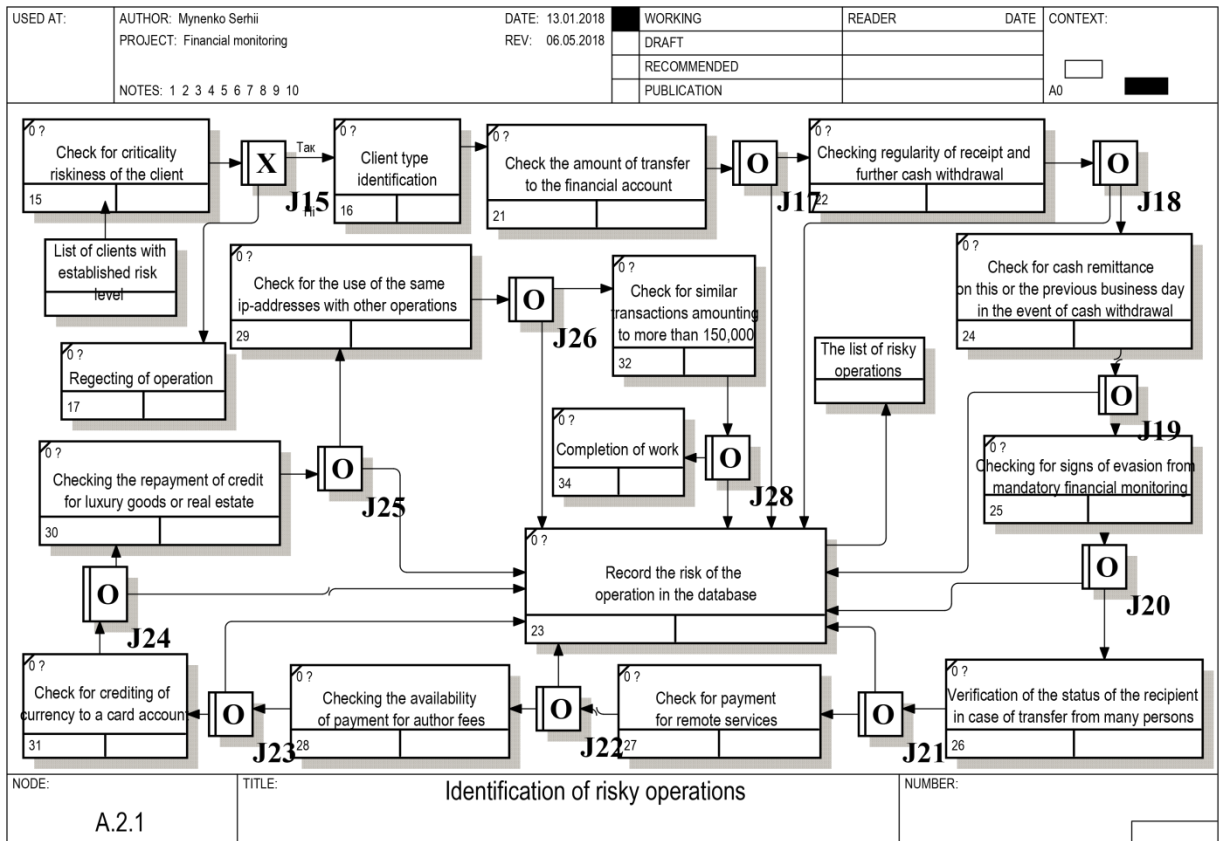


Figure 1. The model of the process of identification of risky operations

Table 1 summarizes the work performed by each functional block.

TABLE I. TABLE I. DESCRIPTION OF WORK PERFORMED BY EACH FUNCTIONAL BLOCK

Function Block	Description
Checking the criticality of the client's riskiness	There is a check, the level of risk of the client is critical
Identifying the type of client	There is a definition of the type of client
Abnormal operation	There is a blocking of the operation
Checking the amount of transfer to the financial account	There is a check, the amount of the operation corresponds to the financial condition of the client
Checking the regularity of receipt and further cash withdrawal	After checking the regularity of the client's transactions receipt and further cash withdrawals
Checking the transfer of cash on this or the previous business day in the event of cash withdrawal	If the cash withdrawal operation, then there is a check, this or the previous operational day there was a transfer of cash
Checking for signs of evasion from mandatory financial monitoring	After checking for signs of evasion from mandatory financial monitoring
Verification if the client is registered in an offshore country	After checking the client's status in an offshore country
Verification of the status of the recipient in case of enrollment from many persons	After checking the status of the client-recipient in the case of transfer from many individuals or legal entities
Check for payment for remote services	After checking the appointment of an operation for the payment of remote services
Check for royalty calculation	After checking the appointment of an operation for royalty
Check for currency crediting to the card account	After checking the currency crediting to the card account
Check for repayment of a loan for luxury goods or real estate	After checking the appointment of an operation to repay a loan for luxury goods or real estate
Checking for a joint ip-addresses with other operations	After checking the use of joint ip-addresses by various legal entities
Check for similar transactions amounting to more than 150,000	After checking for similar operations, the total amount of more than 150000 UAH
Record the risk of the operation in the database	Entering data on the riskiness of an operation in a database
Completion of work	Completion of work proceeding to the next step
List of risky operations	Database with a list of risky operations
List of clients with established risk level	Client database describing the degree of riskiness



This model allows you to assess the riskiness of operations considering the characteristics of economic activities of various types of customers in accordance with the classification of economic entities (individuals, individual entrepreneurs, legal entities) [7]. This system has the ability to:

- comparison of the amount of the transaction by volume or financial condition of the client;
- checking the regularity of receipt and further cash withdrawals;
- identification of cash withdrawal in the event of the transfer of an equivalent amount on this or the previous business day [8];
- checking for signs of evasion from mandatory financial monitoring;
- verification of the existence of a special status of the recipient in case of receipt of funds from many persons;
- checking the availability of payment for remote services and author fees;
- check for crediting of currency to a card account, especially from abroad;
- checking the direction of the operation to repayment of credit for luxury goods or real estate;
- check for the use of the same ip-addresses by different clients [9].

The proposed algorithm will work with the following information. For individuals, this information is about sources of funds and their amount: whether the scholarship, the amount of the aggregate average monthly income, the amount of financial assistance, the amount of income from securities, the amount of income from the sale of claim rights, the amount of loans or loans received, the amount of proceeds from the conclusion of term contracts or the use of other derivative financial instruments and derivatives, the amount of cash receipts and the amount of funds from other sources of income.

For clients of individual entrepreneurs it is information about the code assigned from the Unified State Register of Enterprises and Organizations of Ukraine (EDRPOU) at the time of registration, on the number of employees employed by the entrepreneur, the date of state registration of entrepreneurial activity, location of business activity, e-mail address, telephone number, website, type of economic activity for the Classifier of economic activities, information on the availability of accounts in other banks, the amount of profit for the year, led the amount of loss for the year, the size of long-term and short-term payables, information on loans received and placed deposits, on balances in accounts.

Information about clients of legal entities having information about the head of the organization, representative, co-owner, account manager, the amount of authorized capital, the amount of net income, and the proper name and form of ownership of the enterprise.

A more detailed process of the program is more appropriate to consider on the example of the algorithm. At first, there is a check of the critical level of risk of the client (in case of obvious fictitious activity of the enterprise). Then branching into cash and cashless transactions. In the next step, there is a check: if the transaction amount is more than 50,000 UAH then the algorithm continues to run, if not the completion of the algorithm. In the case of continuation of the algorithm, an analysis of the regularity of receipt and subsequent withdrawal of cash. If there is regularity - then the corresponding entry in the database is entered, and if not the next check of the size of the transaction amount. According to this check, if the transaction amount is more than 150,000 UAH - the direction of the transaction is checked; otherwise, the output from the algorithm. If in the direction of the operation of withdrawal of funds from an account - it is necessary to analyze whether this or the previous operating day there was a cash deposit for a significant amount. After that, the decision is made or the information in the database (in the case of positive analysis) is written down or to the end of the algorithm. In the case of crediting the same cash on the account is checking the financial condition of the client [10]. If the calculated financial condition of the client exceeds the amount of the transaction - then the algorithm needs to go to the end, and if not - to make a corresponding entry in the database.

Next, the algorithm's work is divided into two directions according to the amount of transaction. If the transaction amount is less than 150,000 UAH but more than 140000, then it is necessary to enter information into the database and mark this transaction as intended to evade obligatory financial monitoring. After that, there is a search for transactions for this operational day that have similarities: counterparty coincidence, direct, type of account and currency. When all similar operations are found - their amount is calculated and compared in the next branching decision with the sum of 150000. If the transaction amount is more than 150000 then the corresponding entry is entered into the database [11].

In the case where the transaction amount is more than 150,000, there is a check on entering the foreign currency into the card account, as a result of which it is necessary to make the corresponding entry in the database. If the account is not a card, the algorithm ends. If the currency of the operation is UAH, then the postal code checks the area of the counterparty and analyzes the purpose of payment for the presence of signs of payment for distance services. If the client and counterparty have different areas and on the basis of the appointment there are sufficient conditions to consider the transaction as

payment of distance services, then the corresponding entry is entered in the database.

The next step is to check the assignment of an operation in the event of payment of royalties. If there are sufficient reasons to consider an operation as such, it should be marked in the database. The calculation of the financial condition of the client is then carried out. If the financial position is less than the amount of the transaction, the transaction is recognized as risky [12].

The direction of operation is then determined. If the direction of the transaction is enrollment, then the type of the account on which the transfer is carried out is checked. If an account is a credit, then the purpose of a financial transaction as a loan repayment for elite goods is analyzed and, in the case of a positive analysis, is recorded in the database.

If the card account, then the algorithm needs to calculate the number of unique counterparties that carry out transactions to transfer funds to this client, and in case of exceeding this amount for 50, the social status of the client is determined. If a client receives a scholarship he is considered a student and he is assigned a complex social status. Also, persons with complex social status are minors and persons over 70 years old. If the client is recognized as having a complex social status, the record of the riskiness of the transaction is entered in the database [10].

After conducting all of these checks, checking the matching of the ip-address of this transaction with the ip-addresses of other operations is performed. If there is such a coincidence - it is recorded in the database.

At the output of the algorithm we get a list of financial transactions that are subject to the features of internal financial monitoring.

#### CONCLUSIONS

The proposed system for identifying financial transactions subject to the signs of internal financial monitoring solves the problem presented only in terms of identifying the risk of operations. The proposed approaches to identification of risk operations are effective when there is complete information about the bank's clients and about the operation being performed. In further research it is necessary to pay attention to the riskiness assessment of the client.

#### REFERENCES:

- [1] O. Hlushchenko, I. Semehen, and R. Slav'uk, *Antylehalizatsiyni finansovyi monitorynh: Ryzkyk-oriyentovanyi pidkhid*, 2014 [E-book] Available: [http://finmonitoring.in.ua/wp-content/uploads/2016/06/afm\\_risk-oriented.pdf](http://finmonitoring.in.ua/wp-content/uploads/2016/06/afm_risk-oriented.pdf) [Accessed Mar, 28, 2018]
- [2] State Financial Monitoring Service of Ukraine, "Statystychni dani shchodo otrymanykh Derzhfinmonitorynhom povidomlen' pro finansovi operatsiyi protyahom I kvartalu 2018 roku" [Statistical data on reports received by State Financial Monitoring Service about financial operations during the first quarter of 2018], [Online]. Available: [http://www.sdfm.gov.ua/content/file/Site\\_docs/2018/20180411/Ikv\\_2018ukr.pdf](http://www.sdfm.gov.ua/content/file/Site_docs/2018/20180411/Ikv_2018ukr.pdf) [Accessed Mar, 28, 2018]
- [3] J. Dean, O. Syniavska, and S. Mynenko, "Using economic-mathematical modeling in the study of the economic component of terrorism", *SocioEconomic Challenges*, 1(2), 103-109, 2017. [Online serial]. Available: [http://amgpublishing.sumdu.edu.ua/wp-content/uploads/2016/12/files/sec/issue2/Dean\\_SEC\\_2\\_2017.pdf](http://amgpublishing.sumdu.edu.ua/wp-content/uploads/2016/12/files/sec/issue2/Dean_SEC_2_2017.pdf) [Accessed Mar, 28, 2018]
- [4] The Financial Action Task Force Group on Money Laundering, "Guidance on the risk-based approach to combating money laundering and terrorist financing: high level principles and procedures", The Financial Action Task Force Group on Money Laundering, 2007. [Online] Available: <http://www.fatf-gafi.org/media/fatf/documents/reports/High%20Level%20Principles%20and%20Procedures.pdf> [Accessed Mar, 28, 2018]
- [5] S. Yehorycheva, "Ryzkyk-oriyentovanyi nahlyad u systemi finansovoho monitorynhu", *Naukovyy visnyk Poltavskoho universytetu ekonomiky i torhivli*. Ser.: Ekonomichni nauky – Poltava University of Economic and Trade. Ser. Economical Researches, 3, 126–132, 2016 [Online] Available: <http://dspace.puet.edu.ua/handle/123456789/5390> [Accessed Apr, 4, 2018]
- [6] V. Grushko, "Osoblyvosti vnurtrishn'oho finansovoho monitorynhu v Ukraini" [Features of Internal Financial Monitoring in Ukraine]. *Zbirnyk naukovykh prats' "Problemy i perspektyvy rozvytku bankivskoi systemy Ukrainy"*, #36, 2012. [Online] Available: <http://www.essuir.sumdu.edu.ua/bitstream/123456789/58263/2/Grushko.pdf>. [Accessed Apr, 29, 2018]
- [7] L. Kliusko, "Finansovi monitorynh ta komplyens kontrol' v bankakh" [Financial Monitoring and Compliance Control in Banks]. *Finance, Accounting, Banks*, # 1 (20), 2014. [Online] Available: <http://jpub.donnu.edu.ua/article/download/1014/1031>. [Accessed Apr, 28, 2018]
- [8] State Financial Monitoring Service, "Typolohichne doslidzhennia "Ryzkyky vykorystannia hotivky" [Typological study "Risks of using cash", *State Financial Monitoring Service*. [Online] Available: [http://www.sdfm.gov.ua/content/file/Site\\_docs/2018/20180103/2017%20Nalichka.pdf](http://www.sdfm.gov.ua/content/file/Site_docs/2018/20180103/2017%20Nalichka.pdf) [Accessed Mar, 28, 2018]
- [9] State Financial Monitoring Service, Typolohichne doslidzhennia "Typovi instrumenty, sposoby ta mekhanizmy rozmishchennia i vidmyvannia kryminal'nykh dokhodiv" [Typical tools, ways and mechanisms of placement and laundering of criminal proceeds], *State Financial Monitoring Service*. [Online], Available: [http://www.sdfm.gov.ua/content/file/Site\\_docs/2015/20151230/ty\\_p2015+.pdf](http://www.sdfm.gov.ua/content/file/Site_docs/2015/20151230/ty_p2015+.pdf) [Accessed Apr, 3, 2018]
- [10] State Financial Monitoring Service, Typolohichne doslidzhennia "Aktualni metody, sposoby ta finansovi instrumenty finansuvannya terorizmu ta separatizmu" [Topical methods, methods and financial instruments for financing terrorism and separatism], *State Financial Monitoring Service*. [Online], Available: [http://www.sdfm.gov.ua/content/file/Site\\_docs/2014/20141229/tipolog2014.pdf](http://www.sdfm.gov.ua/content/file/Site_docs/2014/20141229/tipolog2014.pdf) [Accessed Apr, 3, 2018]
- [11] State Financial Monitoring Service, Typolohichne doslidzhennia "Aktualni metody, sposoby ta finansovi instrumenty finansuvannya terorizmu ta separatizmu" [Topical methods, methods and financial instruments for financing terrorism and separatism], *State Financial Monitoring Service*. [Online], Available: [http://www.sdfm.gov.ua/content/file/Site\\_docs/2012/2012121228/2012\\_27\\_12\\_2012.pdf](http://www.sdfm.gov.ua/content/file/Site_docs/2012/2012121228/2012_27_12_2012.pdf) [Accessed Apr, 12, 2018]
- [12] National bank of Ukraine, Lyst "Shchodo finansovykh operatsiy z oznakamy fiktyvnosti" [Letter about financial transactions with signs of fictitiousness] National bank of Ukraine. [Online]. Available: <http://zakon3.rada.gov.ua/laws/show/v7888500-17> [Accessed May, 12, 2018]  
State Financial Monitoring Service, Typolohichne doslidzhennia "Aktualni metody i sposoby lehalizatsiyi (vidmyvannya) dokhodiv, oderzhanykh zlochynnym shlyakhom, ta finansuvannya terorizmu" [Current methods and methods of legalization (laundering) of proceeds from crime and terrorist financing], *State Financial Monitoring Service*. [Online], Available: [http://www.sdfm.gov.ua/content/file/Site\\_docs/2012/2012121228/2012\\_27\\_12\\_2012.pdf](http://www.sdfm.gov.ua/content/file/Site_docs/2012/2012121228/2012_27_12_2012.pdf) [Accessed May, 12, 2018]

# Web-based Application for Planning the Structure of Hybrid Renewable Energy System

Anastasia Verbytska<sup>1</sup>, Yulia Parfenenko<sup>2</sup>, Olha Boiko<sup>3</sup>, Ozhikenov Kasymbek Adilbekovich<sup>4</sup>,  
Assem Kabdoldina<sup>4</sup>

<sup>1,2,3</sup>Sumy State University, Ukraine, <sup>2</sup>yuliya.parfenenko@gmail.com, <sup>3</sup>o.shulyma@ssu.edu.ua

<sup>4</sup>The Kazakh National Research Technical University after K.I. Satpaev, the Republic of Kazakhstan

**Abstract** – The work is devoted to the information system design supporting the selection of the hybrid renewable energy system components. The proposed system allows to plan the optimal structure of the power system based on an expert estimation. It is performed as web-based application with user-friendly interface.

**Keywords** – information system, web-based application, hybrid renewable energy system, planning, structure.

## I. INTRODUCTION

Nowadays, one of the global problems of humanity is the constant increasing in the cost of electricity and the limitation of fossil fuels. Power stations are located at long distances from the end user. As a result, significant losses of useful energy are obtained. The problem can be solved by using renewable energy sources (RES) for electricity generation. RES can be implemented in already existing energy systems (ES). New ES may be created by combining different types of RES. For example, the ES can consist of solar panels (SP), wind power units and storage batteries (B). Such systems are called hybrid renewable energy systems (HRES).

Despite the significant benefits of using such systems, it is very difficult to assess which HRES structure is suitable for the certain household. When making decisions at the planning and design stages of a hybrid energy system, there is a need to take into account all the peculiarities of the operation of HRES before the construction of the power grid. It is also necessary to consider the location of the household and the amount of energy consumed by it.

Today, information technologies are widely used in various fields of human activity. Planning the structure of HRES is no exception. It is convenient to use a specialized decision support system in planning a HRES structure which helps to solve the tasks of HRES components selection.

## II. PRELIMINARY ANALYSIS

The use of information systems when choosing the HRES components allows to achieve the required level of interactivity, get consolidated analytical information on the energy and economic indicators of the power system. An information system is a tool that allows the user to

consider possible variants of combinations of elements of a hybrid power system when planning the structure of HRES.

To date, there are several commercial software products on the market that can be used in planning the structure of the energy system. Their short description and comparison are given below.

1. HOMER [1] – software intended for calculation of power systems of low power and comparison of different methods of energy production.
2. RET-Screen [2] – software that allows to comprehensively identify, assess and optimize the technical and financial viability of potential renewable energy and energy efficiency projects.
3. PVSyst [3] is a tool that allows the user to accurately analyze various configurations from solar panels, diesel generators and storage batteries, evaluate the results and determine the best possible solution.
4. Solarius PV [4] – software for the photovoltaic systems development.
5. Hybrid2 [5] – flexible and easy-to-use software for long-term use in predicting the performance of a hybrid power system.

In [6,7] the information technology and specialized software for solving the problem of HRES components selecting is proposed. However, this software needs to be improved. In particular, it requires the development of a convenient and understandable user interface of the system. Also, in this information system, planning is based on expert estimations, but there is no an expert panel where it could carry out estimations according to defined criteria.

Comparative characteristic of described above software, is presented in table 1.

According to the analysis of systems [1-7] functions, it can be concluded that these software can not be used as the sole tool for supporting decision-making when planning the HRES structure, as they do not cover the whole complex of tasks that arise in the decision-making process.

TABLE 1 - COMPARATIVE CHARACTERISTICS OF SOFTWARE FOR PLANNING THE HYBRID ENERGY SYSTEM STRUCTURE

Software	Solaris PV	RET-Screen	PVSyst	Hybrid2	HOMER	Decision support system, presented in the works [6,7]
Cost	99 USD	free	400 Euro	free	free	free
User-friendly interface	+	-	-	-	+	-
Required level of decision maker knowledge	not high	not high	not high	very high	high	not high
Selection of the generating capacity type	SP + diesel generator + B	SP + diesel generator + B	SP + diesel generator + B	SP + diesel generator + wind power installation + B	SP + wind power installation + B	SP + wind power installation + B
Access to the database with weather data	To values of solar radiation	To values of solar radiation	To values of solar radiation	To values of solar radiation	To values of solar radiation	+
Ability to enter data of consumption user	+	-	+	+	+	+
Ability to enter data on specific RES characteristics	+	-	-	-	-	+
Reporting	+	+	+	-	+	+
Availability of expert cabinet	-	-	-	-	-	-

In table 1 the following abbreviations are used: SP – solar panel; B-storage battery.

On the analysis of the HRES characteristics presented in Table 1, the software HOMER satisfies most of the criteria, that the system for ES planning should correspond, but it still can not perform all the required tasks. After all, this software does not have the ability to enter data on the specific characteristics of RES and the experts panel, which is necessary for the expert evaluation of the power system installations, according to

certain criteria. This is its significant disadvantage. The information system presented in works [6,7] is free, allows to plan energy systems with different types of components, and its shortcomings can be eliminated when upgrading the system. Therefore, there is a necessity to develop the web-application which will be the improved version of the software [6, 7].

### III. PROBLEM STATEMENT

The purpose of the work is to develop a web-based application supporting the selection of HRES components when HRES structure planning.

The developed application will help the user to determine the optimal HRES configuration, consisting of solar batteries, storage batteries and wind power installations.

### IV. RESULTS

The HRES components selection support information system is developed as web-based application with client-server architecture.

The web application consists of three modules:

1. The module of experts;

2. The module of the decision maker;
3. The Operator module.

Its functions are listed as the follows:

1. Provide the interface for several types of users: experts, operator, administrator, decision maker (DM);
2. Input information about energy consumption in the economy (DM);
3. Possibility of weighting the criteria and assessing the range of changes in linguistic variables according to scenarios (Experts);
4. Input information about the installation of RES (Operator).

The main types of users and the functions they perform are presented in the form of a use case diagram in Fig.1.

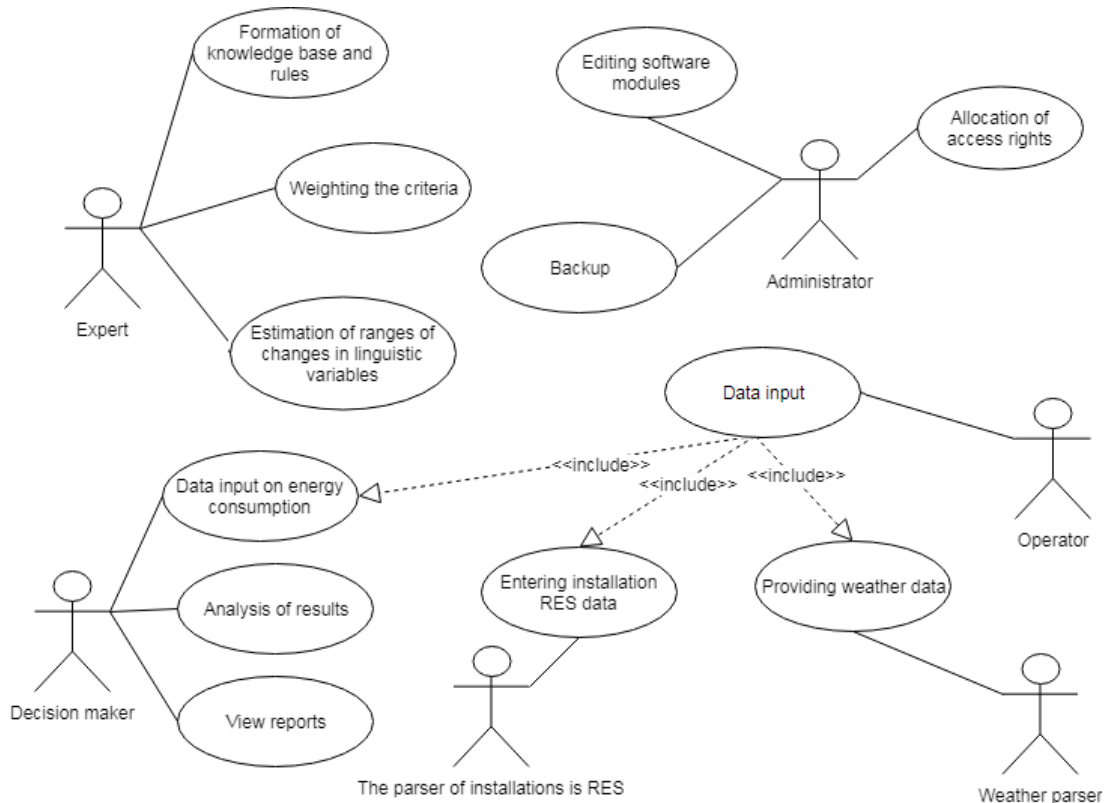


Figure 1 – Use case diagram

Let's consider more detail each of the modules of the proposed web-based application.

The module of the decision maker. In the role of the person making the decision, as a rule, is the user who plans the energy system at the pre-project stage. After authorization, the decision maker fills in the data on electricity consumption in the household. On the next stage, DM inputs information about the geographic location of the power system. Then it selects a point on the map, indicating the location of the future power

system. According to this information, the weather conditions are determined.

After the input of all information, the optimal configuration of the power system is determined.

After clicking on the corresponding button, the information about the selected possible configurations is given to DM.

All actions of the decision maker are presented in the form of a sequence diagram in Fig.2.

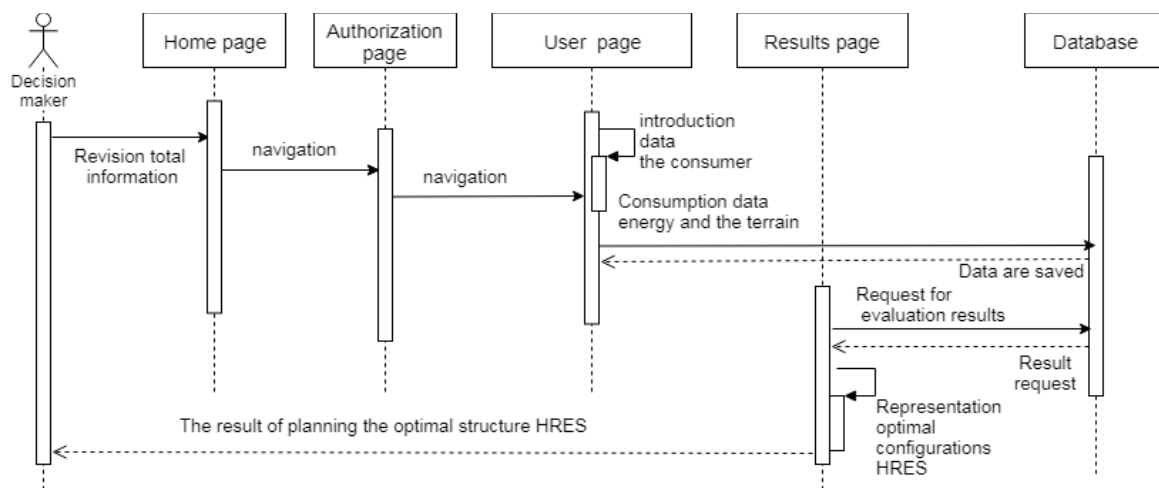


Figure 2 - Sequence Diagram

The module of experts. This module is presented in the form of a separate web-page – an expert panel. After authorization, the expert carries out the weighting of the criteria and assesses the range of changes in linguistic variables in accordance with the scenarios.

An expert assessment is divided into three steps. Step 1 – Weighing the criteria for wind turbine installations. Step 2 – for solar panels. Step 3 – for storage batteries. The interface of this part of the web-application is presented in Fig.3.



Figure 3 –Interface of Expert Panel

Operator module. After authorization, the operator is given the opportunity to manually enter information about the worst weather conditions in the area and installation of RES from which it is planned to build the power system. This feature is implemented for cases where data collection in the automatic mode is impossible.

The web application modules are implemented using PHP5, HTML5, CSS3, JavaScript, MySQL.

## CONCLUSIONS

The developed information system is a web-based application for decision support when planning the structure of the energy system with the use of renewable energy sources. It allows decision maker to choose the optimal HRES configuration. The direction of future research is the development of a module for assessing the expert's reliability.

## REFERENCES:

- [1] Homer Energy. [Online]. Available: [http://vde.nmu.org.ua/ua/lib/HOMER\\_20Energy.pdf](http://vde.nmu.org.ua/ua/lib/HOMER_20Energy.pdf). [Accessed: May 5, 2018].
- [2] RETScreen [Online]. Available: <http://www.nrcan.gc.ca/energy/software-tools/7465>. [Accessed: May 5, 2018].
- [3] PVSyst. [Online]. Available: <http://www.pvsyst.com/en/>. [Accessed: May 5, 2018].
- [4] Solaris-PV. [Online]. Available: <http://www.accasoftware.com/en/solar-pv-system-design-software/>. [Accessed: May 5, 2018].
- [5] Hybrid2 The Hybrid System Simulation Model. [Online]. Available: <https://www.nrel.gov/docs/legosti/old/21272.pdf>. [Accessed: May 5, 2018].
- [6] O., Shulyma, V. Shendryk, A. Bohachov. Informatsiina tekhnolohiia upravlinnia ob'ektamy alternatyvnoi enerhetyky (in Ukrainian) Technology Audit and Production Reserves, vol. 1 (2), pp. 17-22, 2015.
- [7] O., Shulyma, P. Davidsson, V. Shendryk, A. Marchenko. The architecture of an information system for the management of hybrid energy grids. Annals of Computer Science and Information Systems, vol. 6, pp. 281-288, 2015.

**SESSION 2**  
**PROJECT MANAGEMENT**

# Convergence of Agile and Traditional Methodologies in IT Projects

Juliya Kasianenko

Sumy State University, Ukraine, kasianenkojuliya@gmail.com

**Abstract** – The project management methodology is a set of methods, techniques, procedures, rules and practices used through the project life cycle. Despite the growing popularity of agile approach, there are still debates as for its efficiency with the following reasons: the high cost of frequent changes, partial deliverables may not be used, participating in several projects at the same time etc. The traditional approach also have some benefits and drawbacks. The aim is to consider the agile and traditional approaches, its peculiarities and conclude if there is a need and possibility to combine both approaches in one methodology.

**Keywords** – methodology, project management; agile project management; traditional project management, IT projects.

## I. INTRODUCTION

The great deal of attention has been devoted to the agile methodologies since the creation of Agile Manifesto in 2001. According to it the interaction, working product, collaboration with customers, flexibility are the essential values of the successful development of software projects. It was stated that agile thinking is crucial for success in the 21<sup>st</sup> century. According to Van Bennekum, agile is holistic and applicable everywhere in business and life – he uses it as a concept wherever he is and whatever he does. Additionally, Highsmith believes that it is necessary to use agile principles in every project that faces uncertainty.

Agile methodologies are widely used especially in IT projects. The modern IT projects are characterized by the high rate of uncertainty and require the high rate of flexibility. However, it is needed to highlight, that strict deadlines and customer expectations have a high influence on the project development process and might lead to the use of mixed approach in order to ensure the project execution within the project constraints such as time, budget and quality.

Along with the benefits and support of the agile methodologies by authors, traditional approaches are the subject of criticism. Taking into consideration the high cost of late changes in IT projects performed in multi-cultural teams and multi-projects environment, from our point of view, there is the need to consider the convergence of the methodologies that supports and

protects both customer and company interests as well as bring more flexibility and determination of the project results.

## II. IT PROJECTS PECULIARITIES

The following trends in information technologies can be outlined as the following: development of software products, introduction of information systems, development and modernization of computing capacities and data transmission infrastructure. The main difference between IT projects and projects implemented in other field, for example, construction or production is that the project management in IT deals with intangible results within the information field. In addition, IT projects have a number of factors that affect the success of the project.

In the process of IT project management, the project management, faces the need to address unique technological issues related to hardware, operating system, software, problems with databases, etc.

Due to the fact that the result of the IT project is intangible - it can not be measured in conventional units of measurement, project requirements and scope of project should be as detailed as possible. Unlike construction and production projects, IT projects do not have standard costs for typical operations, and the typical operations are different within the framework of similar projects.

Given that information technology is an element of gaining a competitive advantage, many organizations are trying to implement information systems as quick as possible, without conducting detailed planning and setting the task of informatization, that has a very negative impact on the project results.

The features of IT projects in comparison to the projects implemented in other fields, are characterized by the increased complexity and a higher degree of risk. The complexity of the IT project depends on the number of factors.

The organizational scope for projects of information systems implementation determines the number of separate organizational units in which the information system will be implemented. The indicator almost linearly affects the cost of work. The size and structure of the organizational volume affects both the choice of methods for managing the organizational volume, ways of coordinating work, and other aspects of the project.



Functional scope of the project, characterizes the set of functionalities of the information system that are part of the implemented solution. The functional scope for projects of information systems implementation is the most important indicator for assessing the complexity of the project, it is the main part of the content of the project, describes the object of automation - the business processes of the organization. Its change entails a proportional change in other project indicators, such as methodological volume, integration volume, project cost, etc.

Another feature of the IT projects is the need to transfer data from inherited systems (in case this is not primary automated) and to define rules for working with data in the implemented system. Since the purpose of the information system is to work with data, the quality of the system as a whole depends on the quality and completeness of data transfer. For projects implementing information systems, cleaning, standardization and data transfer is a key stage of the work, which largely determines the success of the project.

The implemented information system may contain data that represent the trade secret of the organization, or data that, according to the law, must be protected, for example, financial data or personal data. Work to ensure information security, if included, will increase the cost and complexity of the project. such works are in themselves quite complex and must take into account the provision of information security at the level of workplaces of users, servers and storage sites, at the level of the information and computer network.

These main factors affect the complexity of IT projects for the information systems implementation are present in all projects of this type, but the final set of factors, their weight and criticality depend on the specifics of each particular project. The management of the IT project must take into account factors that determine the complexity of the project, exert control over them, using management tools specific for this type of project.

### III. AGILITY AND THE USE OF AGILE METHODOLOGIES

The advent of Agile project management approach is characterized by the tight connection with software development and software engineering fields. It is needed to highlight, that the brunch of agile approaches include lean approach, extreme approach, adaptive approach. The common characteristic is the adaptability to changes during the project lifecycle. According to DeCarlo adaptability is the key characteristic even more important than predictability of traditional approach. Highsmith states that agility is based on some business principles such as continuous innovation, product adaptation, shortening delivery times, adjustment of people and process and reliable results.

The process of project management can be assessed by considering requirements and specification, project scheduling, team work and the client collaboration. On top of that we have to keep in mind that agile is more than just following the pure process, it is about communication and collaboration inside the team. Communication is definitely one of the project success factors. The most common complaint within the organizations is the lack of communications. Moreover, one out of five projects fail due to unsuccessful communication strategy. Thus, decision making process during the whole project management process is followed by formal as well as informal communication between team and customers. Considering the range of communication methods such as electronic, written, oral, visual, non-verbal and its importance it is definitely the significant factor in project success. Particularly in the agile approaches the team members are much more involved in decision making process which is followed by communication as well.

Agile approaches contain a particular amount of uncertainty and risks, so the specific knowledge are required to complete the project within the constraints taking into account the changes appeared during project implementation. Comparing the reasons for adopting agile in 2016 and 2017 the biggest change were observed especially in accelerate software delivery (75% compared to 69% in 2016), enhancing delivery predictability (46% compared to 30% in 2016), improving IT/Business alignment (49% compared to 42% in 2016), and reducing project cost (24% compared to 18% in 2016). [5] The full list of reasons for adopting agile approach is given below:

- accelerate software delivery 75%;
- enhance ability to manage changing priorities 64%;
- increase productivity 55%;
- improve business/IT alignment 49%;
- enhance software quality 46%;
- enhance delivery predictability 46%;
- improve project visibility 42%;
- reduce project risk 37%;
- improve team morale 28%;
- improve engineering discipline 25%;
- reduce project cost 24%;
- increase software maintainability 18%;
- better manage distributed teams 17%;

The number of different applications of agile project management methodologies have been developed. The 12 State of Agile report gives the following range of mainly used Agile methodologies:

- SCRUM 56%,
- Hybrid (mixed methodologies) 14%,
- ScrumBan 8%,
- Scrum/XP Hybrid 6%,
- Kanban 5%.

Thus, SCRUM, ScrumBan and Scrum/XP Hybrid are the most common methodologies. As for the Agile techniques, the companies use the following:

- daily standup 90%,
- sprint planning 88%,
- retrospectives 85%,
- sprint/iteration review 80%,
- short iterations 69%,
- release planning 67%,
- planning poker/team estimation 65%.

It is essential to remember that agile without structure can cause chaos, particularly in large complex distributed projects where planning control and coordination are critical factors on the way of developing successful project. However, structure without agility can lead to rigidity, particularly when a big amount of discovery, learning and changes are involved in the project.

#### IV. TRADITIONAL APPROACHES

The variety of different projects from the various fields and different sizes apply project management approaches. The goal of the traditional project management approach is considered to be optimization and efficiency through following the detailed project plan and as a result complete the project within the planned time budget and scope.

Modern projects became more and more complicated, the number of interrelations and tasks increases as well as the number of changes, particularly in the IT projects. In the traditional approaches, the tasks relations are linear and cannot reflect the dynamic of modern business environment. Nevertheless, traditional approach is still considered to be the foundation of all modern approaches and it is the methodology which is used in big projects.

The traditional project management methods are widely applied in the following industries: R&D, governmental, transportation, construction, higher education or power. Although, the application of traditional approaches in the IT sphere is very low, the percentage of using both Agile and Traditional approaches is slightly higher.

In order to make the project successful, there are a number of phases that need to be followed, these include:

- initiation,
- planning,
- monitoring and controlling,
- execution.

These steps are referred to as a project life cycle. The traditional project management methods are focused on WBS, Gantt Charts, detailed budgeting, including the earned value technique EVT. Furthermore, according to PMI the project managers have to control the following areas:

- Project Integration Management,
- Project Scope Management,
- Project Time Management,

- Project Cost Management,
- Project Quality Management,
- Project Human Resource Management,
- Project Communications Management,
- Project Risk Management,
- Project Procurement Management,
- Project Stakeholder Management.

Following the traditional approach, the product can be completely described at the planning phase, so the requirements are clear and the change rate is low in comparison to the agile approaches where the requirements might be collected during the whole process of project development as well as the number of improvements might appear on continuous basis. Moreover, documentation is required unlike in the agile approaches. Sufficient documentation preparation surely have a positive influence on the project management process and helps to avoid miscommunication and chaos in the project scope. The project quality is planned in time in details in the traditional approach. However, in agile approach ongoing control is performed along with the achieving of sub-results with respect to the clients expectations. On top of that, communication between team and clients is limited in the traditional approaches what can impact on revealing the clients' needs at the initiation and planning phases and decrease the number of changes, but this also impose a risk on the results verification and tailoring the deliverables according to the clients' requirements.

The Waterfall approach to systems analysis and design was the first established modern approach to building a system. This method was originally defined by Winston W. Royce in 1970. It quickly gained support from managers because everything flows logically from the beginning of a project through the end, (Jonasson, 2008). Sources differ when it comes to the specific steps in the Waterfall process (Jonasson, 2008). However, the basic underlying logic and steps present themselves in each interpretation. The Waterfall method is illustrated on the Figure 1.

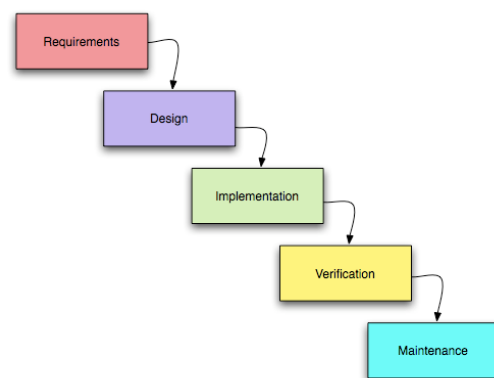


Figure 1 – The Waterfall Development Methodology

There are the number of advantages of waterfall model, such as:

- the model is simple, easy to understand and use;
- the model is easy to manage – each phase has specific deliverables and a review process;
- all software deliverables are completely clear;
- phases do not overlap and are processed and completed at a time;
- the model works well for smaller projects where requirements are absolutely clear;
- meticulous record keeping;
- a customer presence is not strictly required after collection the requirements except for reviews approval etc.;
- the client knows what to expect as the scope, cost and timeline for the project is strictly planned;
- The client have a definite idea of what to expect in the end;
- regardless of turnover, well-prepared documentation decrease the impact on the project;
- the software can be designed completely and carefully.

This provides a better software design with less likelihood of the “piecemeal effect,” a development phenomenon that can occur as pieces of code are defined and subsequently added to an application where they may or may not fit well. However, there are still some disadvantages of the described model:

- no working software is produced until the final stage;
- high risk and uncertainty;
- not suitable model for ongoing projects;
- not suitable for the projects with continuously changing requirements;
- relies heavily on initial requirements;
- in case of requirement error appearing or any change required, the project has to start from the beginning with all new code;
- testing is performed at the end, thus the probability of delay is often very high;
- the client’s evolving needs are not taken into account, in case of any changes needs, the project will be completed late and with the higher budget;
- the high risk of customer dissatisfaction with the delivered software product as all deliverables are based on previously documented requirements so the customer may not see what will be delivered until it is almost finished. By that time, changes can be expensive to implement.

Thus, traditional project management emphasis on linear processes, comprehensive documentation, spends high time on upfront planning; all requirements prioritization is fixed for the lifetime of the project, and works in managed organization. Traditional project management is adverse to changes and follows a formal

change management system. The Return on Investment is after the project is closed and the customer inputs or the involvement in the project may vary depending on the project lifecycle.

Although the limitations of each methodology exist, the organization and project team have to fully understand the project scope and context, tools and techniques and pluses and minuses of each approach to apply the perfect combination of the known methodologies.

## CONCLUSIONS

Both approaches have their benefits and drawbacks, thus the assessment regarding the unified and better approach is not sufficient. Mostly it is necessary to use both approaches and mix the tools generally used in one or the other cases in terms of project category and project characteristics.

Even though IT projects are complex with great amount of uncertainty and with the great number of changes during execution process they are forced to be developed qualitatively and in very fast pace, thus there is a need to combine the agility with clearly defined traditional approach. Depending on the project phase initiation, planning, execution, control and completion the appropriate tools from traditional of agile approach can be used. As the result, through defining the essential features of the IT projects and considering its success factors on the every stage of the project development there is a challenge to develop customized methodology. This is a basis for the further research within the project management context.

## REFERENCES:

- [1] M. Špundaka, “Mixed agile/traditional project management methodology – reality or illusion?”, *Procedia-Social and Behavioral Sciences*, vol. 119, pp. 939-948, 2014.
- [2] Doug DeCarlo, *eXtreme Project*. Jossey bass, 2004.
- [3] J. Highsmith. *Agile project management*. Boston, MA: Addison-Wesley, 2004.
- [4] A. Rasnacisa and S. Berzisa, “Method for Adaptation and Implementation of Agile Project Management Methodology”, *Procedia Computer Science*, vol. 104, pp. 43-50, 2017.
- [5] VersionOne Inc, “The 12th annual state of the Agile Report”. [Online]. Available: collab.net. [Accessed: April. 12, 2018].
- [6] S. Spalek, “Traditional vs. Modern Project Management Methods. Theory and Practice”, in *Proceedings of 21st International Scientific Conference “Smart and Efficient Economy: Preparation for the Future Innovative Economy”*, Brno, Czech Republic, 2016, pp. 499-506.
- [7] T. Dingsmyr, S. Nerur, VenuGopal Balijepally, and N. B. Moe, “A decade of agile methodologies: Towards explaining agile software development”, *The Journal of Systems and Software*, vol. 85, pp. 1213– 1221, 2012.
- [8] SCRUMstudy, “Comparison between Agile Vs. Traditional Project Management”, 12December 2013.
- [9] A.A.Adetokunbo and B.A. Adenowo, “Software Engineering Methodologies: A Review of the Waterfall Model and ObjectOriented Approach”, *International Journal of Scientific & Engineering Research*, vol. 4, no. 7, 427-434, 2013.
- [10] Aljaž Stare, “Agile Project Management in Product Development Projects”, *Procedia of Social and Behavioral Sciences*, vol. 119, pp. 295 – 304, 2014.

**SESSION 3**  
**E-LEARNING TECHNOLOGIES**

# Game Model of Blended Learning in a Unified Learning Environment of the Sumy State University

Inga Vozna, Serhey Shapovalov<sup>1</sup>  
Sumy State University, Ukraine, <sup>1</sup>s.shapovalov@cs.sumdu.edu.ua

**Abstract** – Under consideration is a blended learning game model used in the studying process of Mathematical Logic and Theory of Algorithms course. As the execution environment a single PC-based platform by Sumy State University is used. The single learning environment includes an eLearning system, an open content platform and a blended learning platform.

**Keywords** – blended learning, game model, unified learning environment, mix, eLearning, lectured

“Tell me and I will forget, show me and I may remember; involve me and I will understand.”

Confucius

## I. INTRODUCTION

One of the problems higher education institutions are facing today is searching for new forms of learning [1]. Rethinking of teaching methods as to assigning more time to students for independent work, is in the air. At the same time, students have to retain their skills and lore while teachers have to check knowledge and feedback.

Addressing this problem suggests application of a variety of teaching methods [2].

Along with that, there is another problem to be addressed, which is providing a range of tools to include development of information and communication technologies, mostly internet-based. Such tools would provide appropriate learning environment or a training platform, knowledge management tools and methodologies for individual research.

Therefore, it is important to address this problem in its entirety in the context of ensuring both technological and methodological solution capabilities.

Here we are going to look into application a blended learning game model used in the studying process of Mathematical Logic and Theory of Algorithms course as an example. The course is implemented on a single PC-based platform by the Sumy State University.

## II. TECHNOLOGY AND MEANS OF COMMUNICATION

Sumy State University has created its own information and communication platform (ICP), which

integrated all the teaching resources of the university [3]. It provides virtual communication tools for creating virtual classes, managing educational and training projects.

The teacher manages the creation of educational content at his own discretion.

The presented example is the main content page for the course "Mathematical logic and theory of algorithms" displayed in ICP (Figure 1).

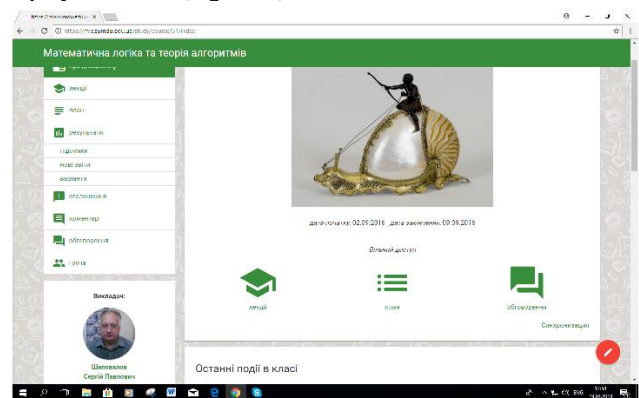


Figure 1. Course Mathematical logic and theory of algorithms in ICP. Home page

The core information and computer environment for content development is called MIX and has home page shown in Figure 2.



Figure 2. MIX platform for blended learning. Landing page

The designer starts to create and manage the training material for the selected training courses at the page shown in Figure 3.

The Sumy State University has developed a complex of integrated learning platforms of proprietary design. All platforms are interconnected.

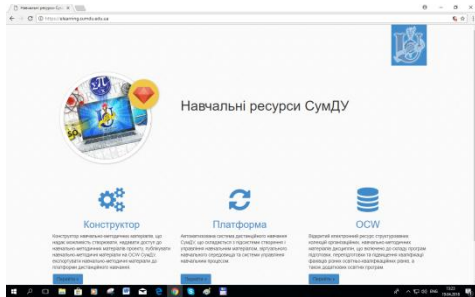


Figure 3. Open educational resources page of the Sumy State University

This complex includes:

- Distance Learning Platform (DL), (Figure 4) – dl.sumdu.edu.ua;
- Open Educational Resources Platform (Open CourseWare, OCW) – structured collections of organizational, learning and educational documents on different disciplines – ocw.sumdu.edu.ua;
- Open editor for teachers – elearning.sumdu.edu.ua;
- Blended learning platform (MIX), consisting of a subsystem for creating and managing educational materials (Figure 6), a virtual learning environment and a learning management subsystem – mix.sumdu.edu.ua and others.

Following are the options available for a teacher to download his teaching materials and to use them in the educational process:

- certification of a distance-learning course using the DL platform;
- import of the course from DL to the editor, where one can supplement and change the materials;
- publication of materials on the open OCW platform of the Sumy State University;
- publication of teaching materials on the MIX platform.



Figure 4. Distance learning platform

The Sumy State University started a project to develop a model of mixed education. The aim of the project is to transform the traditional ways and approaches of teachers to the learning process in general, and to the classroom activities in particular.



Figure 5. Platform for open courses

Teachers became able to model their learning systems using a variety of learning technologies. Any educational process assumes the analysis of educational and methodological activities, which is necessary to ensure positive dynamics of learning (Figure 7). Carrying out such type of analysis will improve the educational process as a whole.

Any educational process involves an analysis of teaching and learning activities that is necessary to ensure a positive learning dynamics. Only after conducting such type of analysis, the technical means should be considered in terms of the benefits that they can give (or not give) to the learning process. Almost in any case a combination of means is needed.

Using the packages of independent multimedia programs alone will be insufficient, because none of them can properly support language activities, which are an integral part of the university training. However, comparing of different learning tools shows how to integrate a range of tools for the most effective use of the benefits from each one.

Figures 8, 9 show the information about the trainees and the progress of each student being trained.



Figure 6. Subsystem for creating and managing educational material



Figure 7. Choice of study

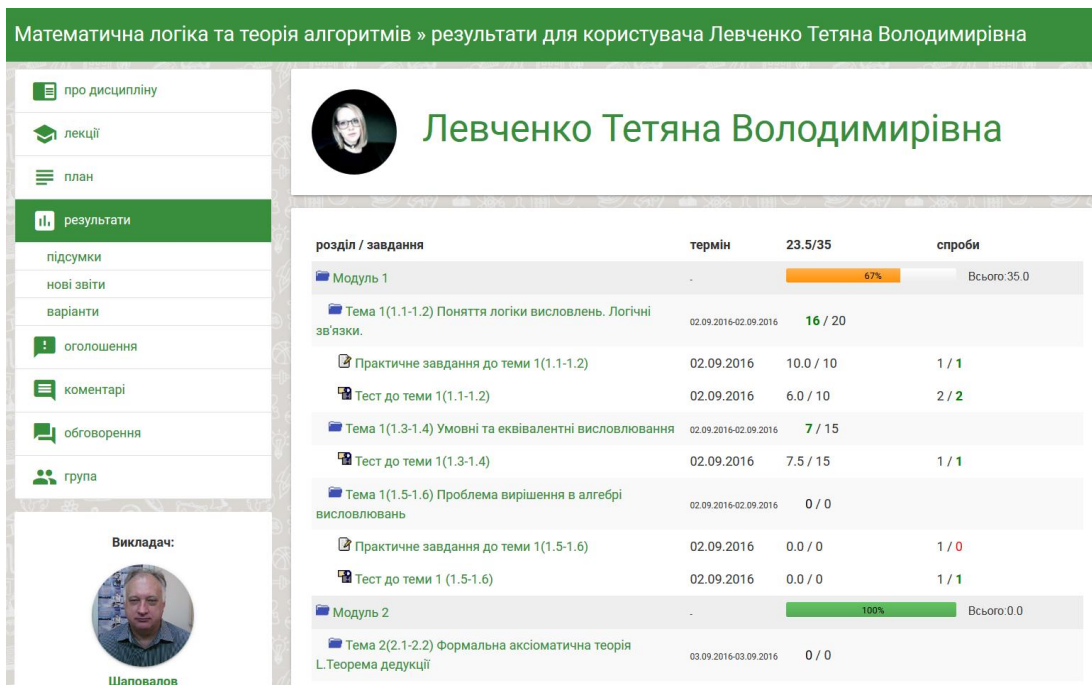


Figure 8. Viewing the academic progress of a particular student

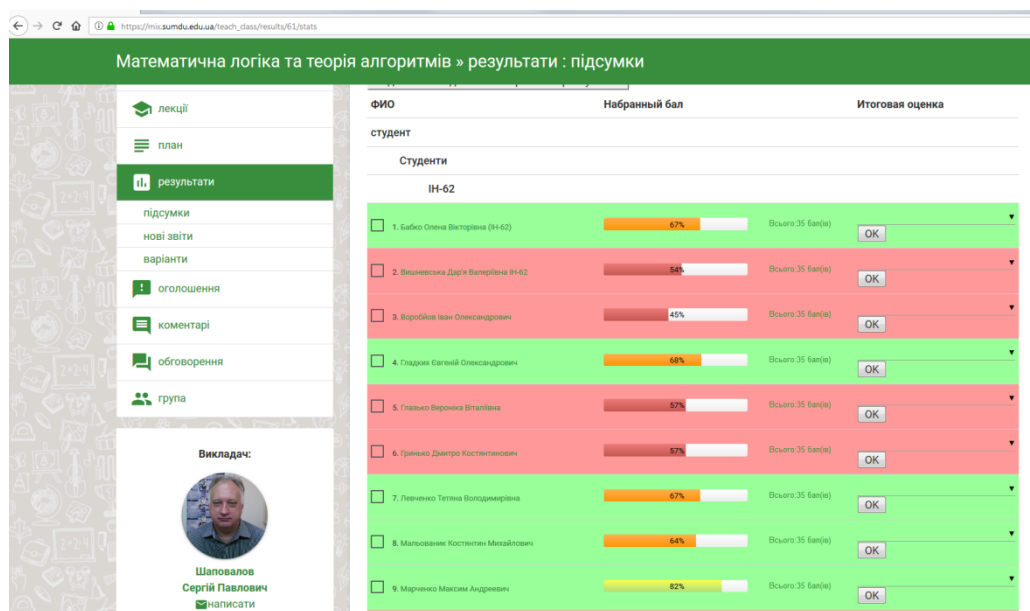


Figure 9. Viewing student group progress

### III. GAME MODEL OF BLENDED LEARNING AND ITS IMPLEMENTATION

The game model of blended learning provides an environment for activating student learning and provides an opportunity to meet their different needs and learning styles. It motivates students by offering them an interactive competitive environment.

Application of the game models in the educational process is the subject of many studies, for example, [4-6]. "Regardless of age or economic, ethnic, or social background, people understand the language of play" (Azriel, [4]). Games offer a medium for students to explore and interrogate information in a fun and interactive way.

The development of the concept of games as a teaching methodology for the first time began, as stated

in [5], in the mid-1940s. Since then there has been many experiments and researches done concerning game models in learning.

The type of game discussed in this article refers to competitive games. The general scheme of this game is shown in Figure 10.<sup>1</sup>

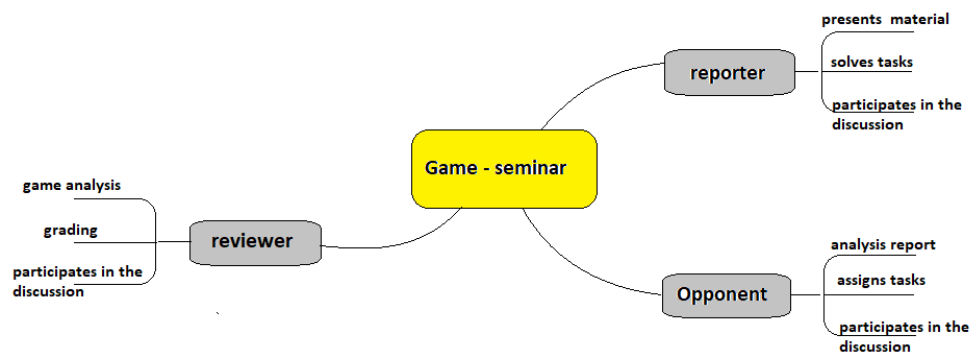


Figure 10. Distribution of roles and actions

The general conceptual scheme of the game model of blended learning is shown in Figure 11. Studying through this model, students go through two phases.

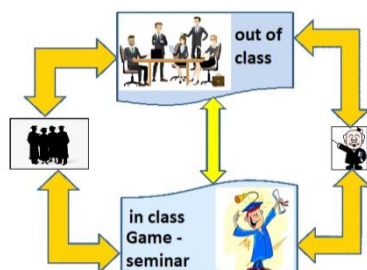


Figure 11. Game model of blended learning

The first stage is distance learning (outside the classroom) in which students independently study the course materials. The teacher obtains some administrative rights and directs the mastery of knowledge. He opens and closes tests and simulators, sets and checks open tasks, answers students' questions on the bulletin board and makes his announcements. In the second stage, learning (in the classroom) is a very action - a game. A group of students is divided into two teams, one of which acts as an orator and the second as an enemy. The teacher in the game plays the role of expert reviewer. The most important role in this technology belongs to the final discussion, in which the students together analyze the course and results of the game, the course of educational and gaming interaction. All actions of students on the development of knowledge are assessed in points and are highlighted on the site.

The results of the game appear in a double plan - in quality both as a game and as educational-cognitive result.

## CONCLUSIONS

The second-year students of special computer science of Sumy State University conducted the tests for the game model of blended learning. Students perceived the game as an attractive learning technology.

The conducted analysis of student achievements and the qualitative feedback show the positive impact of the game model on students. Students are more involved into the learning process and improve their results.

An anonymous student survey showed that 67% of respondents would have chosen a game model of blended learning compared to the traditional form of learning.

There were no negative comments in the questionnaire, although individual recommendations were given to improve the model.

It is important to recognize that although the game model of blended learning has some success in comparison with the traditional form of learning, it requires more labor both from the students and from the teachers.

## REFERENCES:

- [1] G. Light, R. Cox and S. Calkins, *Learning and Teaching in Higher Education: The Reflective Professional*. Thousand Oaks, CA: Sage Publishing, 2009.
- [2] B. Allan, *Blended Learning Tools for Teaching and Training*. London: Facet publishing, 2007.
- [3] A.V. Vasiliev, V.O. Lyubchak and Y.O. Zuban, *IT support for the activities of the Innovation University: the experience of Ukrainian higher education*. Sumy: Sumy State University, 2016.
- [4] J. Azriel, M. Erthal and E. Star, "Answers, Questions, and Deceptions: What Is the Role of Games in Business Education?", *Journal of Education for Business*, vol. 81, no. 1, pp. 9-13, 2005.
- [5] D. Kirkland, F. O'Riordan "Games as an Engaging Teaching and Learning Technique: Learning or playing?" September, 2010. [Online]. Available: [http://icep.ie/wp-content/uploads/2010/01/Kirkland\\_et\\_al.pdf](http://icep.ie/wp-content/uploads/2010/01/Kirkland_et_al.pdf)
- [6] Ruben B., "Simulations, Games, and Experience-Based Learning: The Quest for a New Paradigm for Teaching and Learning", *Simulation and Gaming*, vol. 30, issue 4, pp.498-505, 1999.



# Development of a Technique for the Creation of Virtual Simulators of Artillery Weapons

Natalia Fedotova<sup>1</sup>, Tetiana Yasinska<sup>2</sup>

Sumy State University, Ukraine, <sup>1</sup>fna\_2000@ukr.net, <sup>2</sup>santadachshund@gmail.com

**Abstract** – In this scientific work, the disadvantages of the system of education were analyzed on the example of the students of the military department of the SSU, the classification of modern virtual trainers was developed, the structure of the simulator was proposed for effective and valuable training of the student, an example of work with a training simulator for studying the structure and working with a gate 122 mm howitzers D-30 was shown.

**Keywords** – simulator; multimedia simulator complex; technique; metasystem.

## I. INTRODUCTION

Military training is an integral part of the formation of a holistic and economically stable state. But due to the reduction of personnel and military educational institutions in 2006 - 2011 some worsening occurred, namely, the material component of institutions was destroyed. Therefore, the question arose about modernizing the military education of the country by changing the material and technical base and introducing the necessary number of simulators, virtual simulators, information systems in the course of studying. This is the basis for the relevance of this research work "Development of a technique for the creation of virtual simulators of artillery weapons". The purpose of the work is to improve the quality of training of military specialists, based on the use of computer information systems. To achieve the goal, the following tasks were considered: the proposed information technology for the organization of the educational process for the training of military specialists, the architecture of the multimedia training complex (MTC) was developed, the methodology for designing the visual models of the subject area of the simulator was developed. The object of the study in this paper is the computerized training of military specialists on the example of cadets and students of the Department of Military Training of the SSU. The scientific novelty of this research work consists in the proposed information technology for the modernization of the training process at the military department of the SSU.

## II. ANALYSIS OF TEACHING METHODS

At the moment, there are two modes of training in military training: passive and active. Passive is the

mastering of lectures on the exploitation of a particular weapon. But there are a number of reasons why this regime is not effective, for example, in some military schools it is not possible to provide the students with the necessary training tools, with the help of which a future military specialist would acquire practical skills, while learning the previously acquired theoretical knowledge. Also, the technology considered in the curriculum is expensive, most often it is absent, which impedes its effective use, cadets of the military department have an opportunity to get acquainted with the model of the weapon only using graphic images. Therefore, the question arises about the use of an active mode of training, which is a full-fledged work with the model of a real object, using a virtual simulator. The cadet performs a certain set of actions: acquainted with the structure of the weapon, passes tests for checking knowledge and performs practical tasks. Students of the Department of Military Training of our university together with the teachers of the section of the ITP have created a number of simulators, which gives them the opportunity to reveal their professional potential [1][2]

## III. CLASSIFICATION OF VIRTUAL SIMULATORS

To create our simulator, we conducted a study of existing virtual simulators, which resulted in this classification, which is presented on Fig. 1. Conditional classification can be divided into two groups: the first group is classified by purpose, by the internal structure and by the type of the implemented study scenario, the second by the functions and type executed [3].

In order to meet the educational process needs of military educational institutions, it is proposed to create simulators that contain the following elements: questioning systems, game scripts, simulation complexes.

## IV. TECHNIQUE FOR DEVELOPING MTC

MTC (Multimedia Simulator Complex) is a system for learning using the means of input and output information, which includes a theoretical material, a block of testing and a block of practical tasks in which the transfer and exchange of information take place in a dialogue mode. The scheme of organization of the educational process using this complex is presented on Fig. 2. This simulator provides for the use of both passive and active mode of training [4].

To develop the MTC, a technique for developing computerized 3D models was developed, consisting of the following steps:

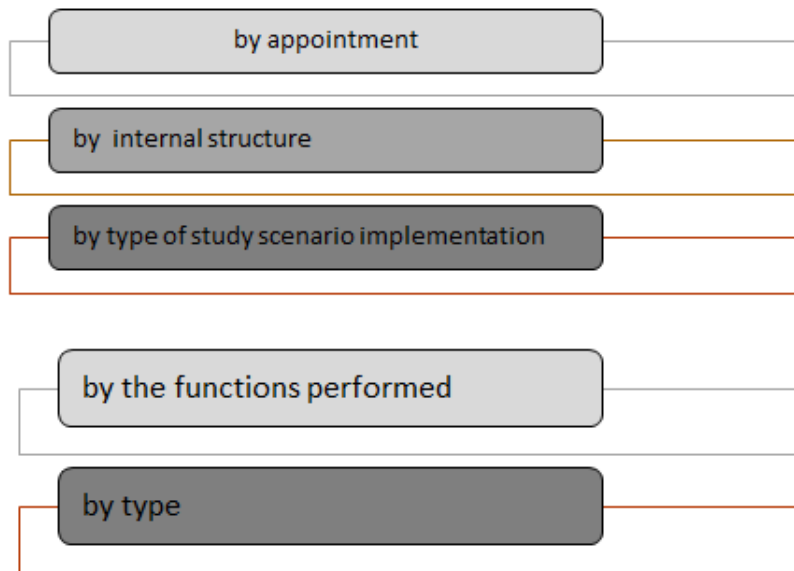


Figure 1. Classification of virtual simulators

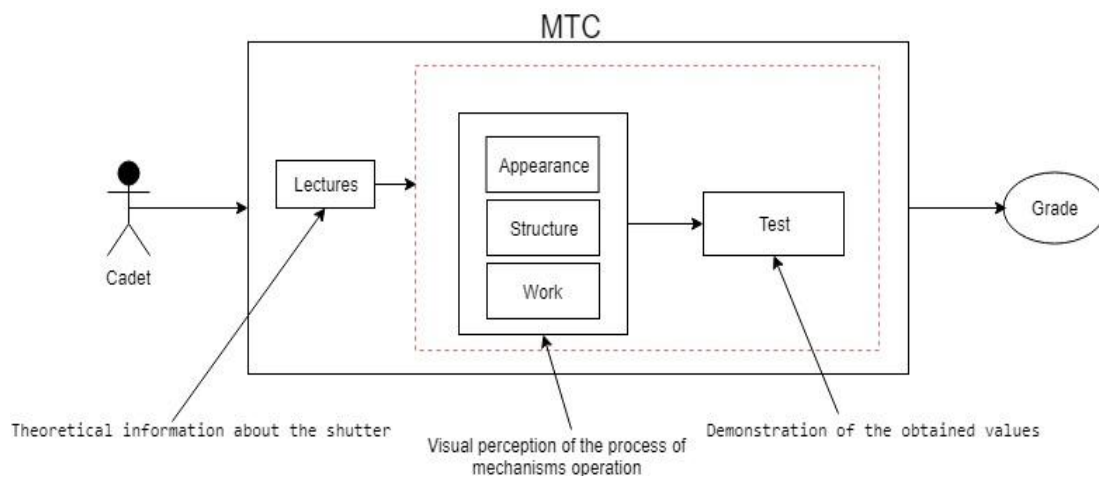


Figure 2. Scheme of educational process organization using MTC

1. defining goals and objectives
2. software choices
3. defining methods for creating a 3D model
4. designing
  - 4.1 determining the level of detail of the 3D model
  - 4.2 decomposition
5. visualization
6. ready software product

#### V. CONCEPTUAL MODEL OF THE SHUTTER 122MM HOWITZERS D-30

When creating a virtual simulator for the study of the structure and operation of the gate 122 mm howitzers

D-30, the methods of simulation-numerical simulation were used and a series of work stages was performed: the study of the main elements for the construction of the simulation shutter mechanism; execution of a virtual model graphic solution using special software that provides the opportunity to simulate materials and lighting as much as possible; writing a software code for a virtual model that describes the physics of the interaction of gate mechanisms. Given that the object of simulation consists of a large number of structural elements, the process of constructing its model is based on the system approach. The shutter itself is considered as a metasystem consisting of subsystems of elements and connections between them. As a result of the analysis in the form of a node tree, a model is constructed which maximally

reflects the model structure of the shutter. The conceptual model of the research object is presented on Fig. 3 [5]

The created complex has a common software shell, which includes separate structural parts, of which the first three have the form of animated videos and a block of test tasks. Structural parts of the MTC are shown on Fig. 4.

#### VI. EXAMPLE OF MTC

To develop this multimedia complex, software products such as Autodesk Inventor were used to create 3d models of all components of the shutter and Adobe Flash to create a shell for testing and adding an interactive part. The main window of the program is shown on Fig.5.

At the initial stage of work with this simulator, the cadet / student should review all the videos that are theoretical material to the tests. If he did not view all the videos or viewed, but using the rewind mode, the button to open the test portion is not active. To begin testing, the

user must enter his data. Then the student is tested using the interactive elements contained in the MTC. After completing the test, the user is given the opportunity to view their answers and find mistakes. The teacher receives the result as a non-editable file. Examples of steps for entering the input data and displaying the correct answer are shown in Fig.6 and Fig.7.

#### VII. PRACTICAL SIGNIFICANCE

The practical significance of this research work is that a new approach was proposed to the organization of the educational process at the military departments with the use of modern information technologies and a number of training complexes were created, according to the proposed methodology, to ensure the improvement of the quality of military education.

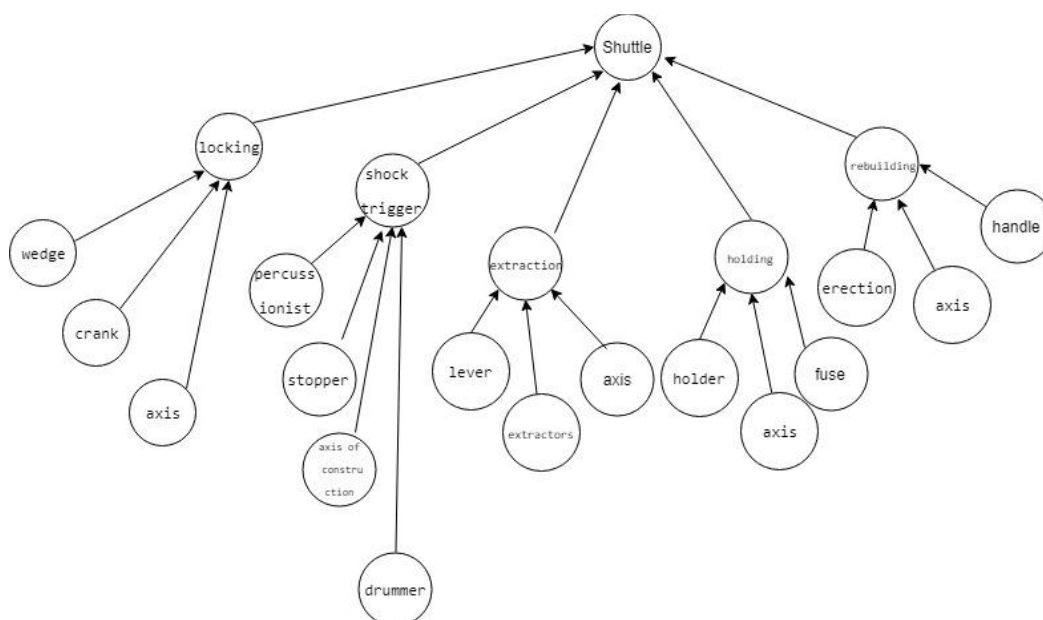


Figure 3. Conceptual model of the shutter 122mm howitzers D-30

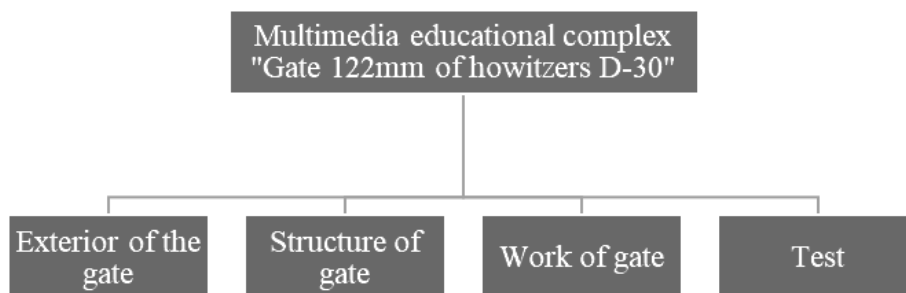


Figure 4. Structural parts of the MTC

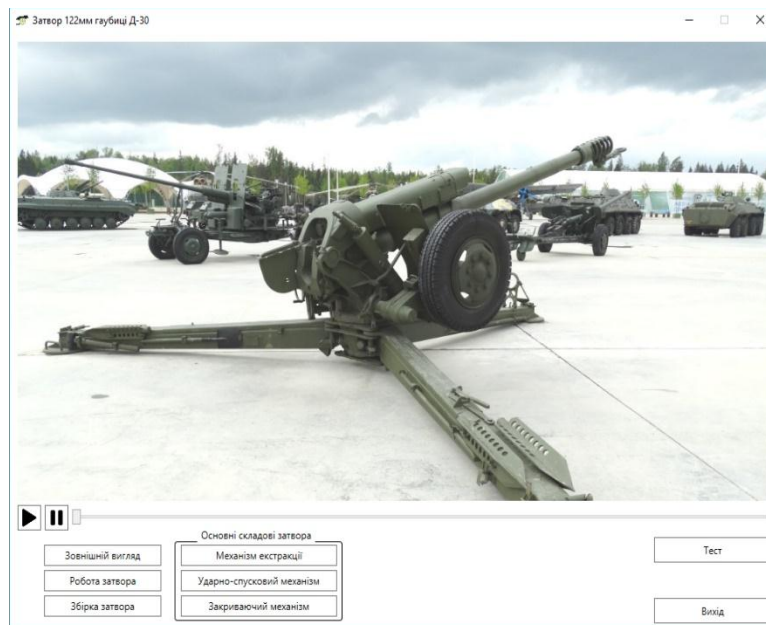


Figure 5. The main window of the programm

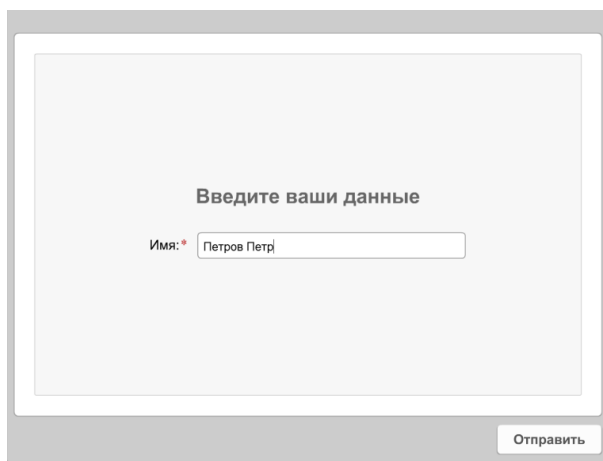


Figure 6. The stage of entering the user's input

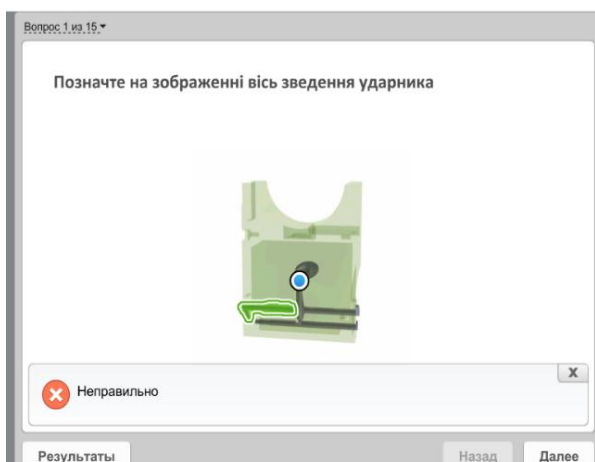


Figure 7. The response checking phase

## CONCLUSIONS

Upon completion of the research work, the following results were obtained: namely, an approach to modernizing the training process at the military department of the university was proposed, a list of stages for the creation of computer simulators of military use was determined, video materials were developed that demonstrate the operation of the shutter 122 mm of howitzers D-30, compiled a list of test tasks for assessing the level of mastering the training material, in accordance with the proposed methodology, MTC has been created to study the shutter operation.

## REFERENCES:

- [1] H.I. Vavilova, Designing of virtual simulators. M.: Logos, 2012. 127 p.
- [2] O.L. Dzubenko. "Theory and Methods of Education Professional". [Online]. Available: [http://sconference.org/publ/nauchno\\_prakticheskie\\_konferencii/pedagogicheskie\\_nauki/teoriya\\_i\\_metodika\\_professionalnogo\\_obrazovaniya/12-1-0-2255](http://sconference.org/publ/nauchno_prakticheskie_konferencii/pedagogicheskie_nauki/teoriya_i_metodika_professionalnogo_obrazovaniya/12-1-0-2255) [Accessed: April. 30, 2018].
- [3] N.N. Filatova and N.I. Vavilov, "Designing multimedia simulators based on scenarios of presentation of knowledge". [Online]. Available: [http://ifet.s.iiee.org/russian/depository/v3\\_i4/html/3.html](http://ifet.s.iiee.org/russian/depository/v3_i4/html/3.html) [Accessed: April. 30, 2018].
- [4] "Application vIRTUALNYH in laboratories of technical education". [Online]. Available: <https://www.allbeton.ru/article/450074.html> [Accessed: April. 30, 2018].
- [5] "Model predmetnoy region". [Online]. Available: <http://www-sbras.nsc.ru/Report2006/Report321/node14.html> [Accessed: April. 30, 2018].

# Electronic Military Training Device

Kateryna Savytska, Viktoria Kudrytska, Natalia Fedotova<sup>1</sup>  
Sumy State University, Ukraine, <sup>1</sup>n.fedotova@cs.sumdu.edu.ua

**Abstract** – Leading countries of the world are increasingly paying particular attention to improving the military training system with the involvement of various types of training systems and simulation facilities for combat conditions. This is due to the need for a high level of preparedness of personnel with simultaneous reduction of material and financial costs. This approach today is extremely relevant for Ukraine, especially in the context of the current situation in the country.

**Keywords** – multimedia simulator; PTRK; model.

## I. INTRODUCTION

Earlier, large-scale field training exercises were needed to train personnel to engage in joint military operations, in which participated in units from the battalion level and above all NATO member states military units and the Warsaw Pact, as was the case with the Cold War era. To conduct such maneuvers spent huge funds, and their preparation and organization took a lot of time. With the advent of modern training systems and means of simulation of the situations of combat situation, the situation in the field of organization and training personnel has dramatically changed. The anti-aircraft missile systems are stationary mechanical simulators that imitate not only the results of these weapons but also the sound and visual effects that arise during their time use.

In the structure of the PTRK, as a rule, includes a launching unit with the aiming and launching equipment and an antitank guided missile. Depending on the type of control system that is used. There are types of PTRK of the first, second and third generation. In the structure of the PTRK, as a rule, includes a launching unit with the aiming and launching equipment and an antitank guided missile. Depending on the type of control system that is used. There are types of PTRK of the first, second and third generation. The first generation PTRK was developed in the 50's, is presented on a target using the "three-point" method: the first point is the eye of the gunner (crossing the visor), the second point is the rocket on the trajectory, and the third is the target.

The second generation PTRK, which appeared in the 1960s, differed, first of all, by a semi-automatic guidance. The operator combines and holds the target mark for the purpose of the entire flight time of the rocket, and the observation of the missile and the development of

corrective commands is carried out by the automatic equipment

The third generation of PTRK realized the principle of "shot and forgot". The operator must only bring the launcher to the target and carry out the shot, and the head of self-direction PTRK in the flight itself carries out the support of the target and gives it a rocket.

## II. MAIN RESEARCH

The anti-tank missile system 9P149 "STURM-S" refers to the complexes of the second generation. Figure 1 shows the appearance of the machine.

The 9P149 product is designed to combat moving and stationary armored and small-scale terrestrial targets (DFT, DZVT), as well as low-speed airborne low-flying targets.

The 9P149 product allows shooting from the place, from short stops and afloat, in conditions of direct optical visibility at any time of day and year. The product is able to overcome water obstacles.

The product 9P149 is a self-propelled start-up unit with automatic charging from the PU.

As a basic machine, a lightweight, multipurpose traction unit (MT-LB) was used, modified to accommodate the equipment and equipment of the product 9P149 and has an index-product 49.

The location of the power transmission units in the front of the body of the base machine, the control and engine compartment is saved as on the base tractor. On-site cargo platform formed a combat division. Passage along the combat vehicle on the right side between the control division and the combat unit is blocked by a partition with a hatch and assigned to the control division. Thus, all compartments are separated by partitions.

The anti-tank guided missile 9M114 is designed to destroy modern tanks and other armored targets, as well as small-scale terrestrial (TDT, DZVT, etc.) and low-flying airplanes (helicopters, low-speed aircraft) for enemy targets. The 9M114 is used for firing from the combat vehicle 9P149, as well as from the 9K113 complex, which is located on the Mi-24 helicopter [1].

Anti-tank guided missiles must always be operational and ready for immediate use, which is ensured by strict compliance with the operating rules.

The storage and preservation of anti-tank guided missiles in troops is organized in accordance with the requirements of the current governing documents.

Personnel who are well-known in the structure, the principle of operation of an antitank guided missile, as well as a technical description and instruction manual on

the operation of this type of antitank guided missile are allowed to operate the shells.

Disassembly of the PLD in the troops is prohibited.



Figure 1. The appearance of the machine

The location of the power transmission units in the front of the body of the base machine, the control and engine compartment is saved as on the base tractor. On-site cargo platform formed a combat division. Passage along the combat vehicle on the right side between the control division and the combat unit is blocked by a partition with a hatch and assigned to the control division. Thus, all compartments are separated by partitions.

The anti-tank guided missile 9M114 is designed to destroy modern tanks and other armored targets, as well as small-scale terrestrial (TDT, DZVT, etc.) and low-flying airplanes (helicopters, low-speed aircraft) for enemy targets. The 9M114 is used for firing from the combat vehicle 9P149, as well as from the 9K113 complex, which is located on the Mi-24 helicopter.

Anti-tank guided missiles must always be operational and ready for immediate use, which is ensured by strict compliance with the operating rules.

The storage and preservation of anti-tank guided missiles in troops is organized in accordance with the requirements of the current governing documents.

Personnel who are well-known in the structure, the principle of operation of an antitank guided missile, as well as a technical description and instruction manual on the operation of this type of antitank guided missile are allowed to operate the shells.

Disassembly of the PLD in the troops is prohibited.

In case of detection of malfunctions in the course of conducting routine work, the PTCR seem to be established in the order of storage bases (arsenals).

When entering the warehouses of military units from storage bases (arsenals), with the help of KPM 9B94,

entrance checks are made of 3% of the OPCP (but not less than 6 units) from each batch. In case of defects detection, the entire batch is checked.

When sending the PPTC from the warehouse to the unit, it is necessary to conduct regular work in full, if since the last inspection passed more than three months.

During the receipt of the PTP from the units into the warehouse, each projectile is inspected in the full scope of the scheduled work.

The transfer and loading of the 9A687 encapsulation with the projectile should be carried out by two persons.

It is forbidden to transfer the closure with a projectile lid down, bend, throw and drag it when loading and unloading.

In case of accidental falling of a projectile in the enclosure 9A687 from height to 0,5 m the projectile is allowed for further exploitation.

In the event of accidental falling of a projectile in an enclosure with a height of more than 0,5 m and without closure from any height the projectile to be stored and used in troops is forbidden, it is subject to destruction in the established order.

Persons who have been trained in safety engineering are allowed to work with the PTKR. When using the PTCR, you must comply with all the safety rules that are installed for the operation of ammunition.

It is forbidden to use the PTCR:

- carry out routine inspections directly in storage facilities and at launch units;
- carry out work related to routine inspections, faulty tool;

- use a military projectile for educational purposes;

Attending third parties during inspections and routine inspections.

Regular work when storing a projectile in warehouses and under a canopy should be carried out only in a specially equipped place and at a distance of at least 50 meters from the storage facilities and residential buildings, with the military part of the projectile being directed in a safe direction.

It is forbidden to find people from the front and rear end of the projectile when checking it.

Do not touch the electric contacts of the projectile.

In the event of a drop in a projectile in a purchase from a height of more than 0.5 meters, the following procedure for handling it shall be established:

Carefully, without punches and canting, to load a shell in the enclosure into a car body on a special foam rubber stomach with a thickness of 50 millimeters or sand, sawdust or other bulk material with a thickness of not less than 200 millimeters, a battle piece towards the side board; transport the projectile to a disposal site at a speed not exceeding 15 km/h at the nearest safe distance along the highway or on a level road/

On a place destined for destruction of a projectile, the closure with a projectile shall be transferred from the body of the vehicle, with all safety precautions, and destroy the projectile without removing it from the enclosure, by undermining it.

Storage of the product should include all types of work necessary to keep it in constant alert. Short-term storage of the product is allowed both in indoor unheated premises, and in field conditions (under a canopy or open areas). Regardless of the type of storage and type of storage, the product is stored in the cover, all hatches must be tightly closed [2].

### III. ADOBE FLASH PLAYER

The development of an elearning tool is done with the Adobe Flash software. The purpose of the work is to develop an electronic educational tool, which will display the theoretical material and practical tasks implemented on it, on the basis of which knowledge testing will take place. Adobe Flash Professional (AFP) is a multimedia platform used for authoring vector graphics, animations, games and rich Internet applications that can be viewed, played, or run in Adobe Flash Player.

The scope of Flash is different, it can be games, websites, presentations, banners and just cartoons. When you create a product, you can use media, sound and image files, you can create interactive interfaces and complete web applications.

Adobe Flash is an environment for creating Flash Platform applications.

The scope of Flash is different, it can be games, websites, presentations, banners and just cartoons. When you create a product, you can use media, sound and

image files, you can create interactive interfaces and complete web applications using PHP and XML.

Adobe Flash is an environment for creating Flash Platform applications, along with other tools (environments): Adobe Flash Builder, Flash Development Tool (FDT), and others.

Flash files have .swf extensions and require Adobe Flash Player to be viewed, which can be installed as a browser plug-in. Flash Player is distributed free of charge through the Adobe website. Output files with the .fla extension are created in the Macromedia Flash development environment, and then compiled into an understandable Flash Player format - .swf.

At the heart of Flash lies the vector morphing, that is, the smooth "flow" of one keyframe into another. This allows you to make rather complex cartoon scenes by asking only a few key frames for each character.

The second "whale" Flash is complete programming. Flash uses the ActionScript programming language, which is syntactically similar to JavaScript. The latest version of the language (ActionScript 3.0) is a fully featured, object-oriented language.

Using vector graphics, along with software code, provides small file sizes for Flash, enabling you to save traffic and reduce page load times. But to display text, video alternatives can provide better performance and consume less CPU resources than Flash objects, for example, using transparency or display of images[3].

The main drawback of Flash-applications is the excessive demand for processor resources. Lack of power on the computer can affect the performance of the operating system as a whole, or lead to distortion of the results of the Flash application associated with the display of animation or time calculation. Sometimes this is due to the disadvantages of software responsible for processing Flash components, or the low quality of the Flash application itself.

Another important drawback is that it is not always possible to launch the Flash application, or it is associated with some difficulties (for example, you need to install a plugin or upgrade it to the latest version). Some users (or system administrators across the entire network) disable the browser's ability to download content that is processed by plug-ins or downloaded in frames for the purpose of information security (due to possible content threat, such as interception of the clipboard), saving system resources, or for salvation from annoying ads.

On the Internet, you can find sites that are completely designed as a Flash application (all content, as well as navigation elements). Usually these sites are dedicated to games, design studios, personal pages and other sites that aim to impress the visitor with the beauty and unusual nature of the implementation. Large portals and information resources are trying to avoid the use of Flash (with the exception of the insertion of advertising banners, the impossibility of which does not cause inconvenience to users).

Using Flash to accommodate textual information prevents its indexing by search engines. However, there are many ways to solve this problem. One way to solve this problem is to use text in HTML format, in the footer of the page[4].

ActionScript is a scripting programming language that allows you to program Adobe Flash-based clips and applications. ActionScript, like JavaScript, is based on ECMAScript, a script language standard, so syntax is very similar in both languages. But while DOM JavaScript interacts with the browser window, HTML documents and forms, ActionScript DOM works with flash clips that can include animation, audio, video, text, and event handling. ActionScript is a scripting programming language, which allows you to program Adobe Flash-clips and applications. ActionScript, like JavaScript, is based on ECMAScript, a script language standard, so syntax is very similar in both languages. But while DOM JavaScript interacts with the browser window, HTML documents and forms, ActionScript DOM works with flash clips that can include animation, audio, video, text, and event handling[5].

SWFs can contain animations or applets with varying levels of interactivity and functionality. The picture stored in this format is scaled without visible distortion, the video is small, it is faster loading and playing the video file.

With ActionScript, you can create interactive multimedia applications, games, websites.

#### IV. AUTODESK 3DS MAX

The machine, which shown in Figure 1 was created using Autodesk 3Ds max.

Autodesk 3Ds max is a professional 3D computer graphics program for making 3D animations, models, games and images. It is frequently used by video game developers, many TV commercial studios and architectural visualization studios. It is also used for movie effects and movie pre-visualization. 3Ds Max supports both import and linking of DWG files. Improved memory management in 3Ds Max 2008 enables larger scenes to be imported with multiple objects [6].

Modeling 3ds Max has extensive tools for creating various in form and complexity of three-dimensional computer models of real or fantastic objects of the surrounding world using a variety of techniques and mechanisms, including the following:

- polygonal modeling;
- modeling based on inhomogeneous rational B-splines (NURBS);
- modeling based on portions of Bézier surfaces;
- modeling using built-in libraries of standard parametric objects and modifiers.

Autodesk 3ds Max is primarily based on polygon modeling. Polygon modeling is used extensive in game design because it offers a highly specific control over individual polygons that make up the model. This feature

of polygon modeling also allow for greater optimization of the model [7].

After the 3D model is created, it can be exported to Adobe Flash Player. The 3ds Max editor allows you to import and export files and objects. This makes it possible to get geometric forms with other programs, such as Adobe Photoshop, Adobe Illustrator, AutoCAD, etc.

When exporting files, the program converts scenes with the extension .max to universal formats .3ds or .dwf. Since these formats are less susceptible to the transformation, some of the information about the scene in the transformation may change or disappear. The editor of 3ds Max 2013 warns about this when converting.

#### CONCLUSIONS

In the process of studying the issue of military training, we can say with confidence that this is a very important for Ukraine and other countries. This is due to the need for a high level of preparedness of personnel with simultaneous reduction of material and financial costs. Military simulator allows in electronic form to study the nuances of work with military equipment and shooting.

The development of an elearning tool is done with the Autodesk 3Ds max and the Adobe Flash software. The purpose of the work is to develop an electronic educational tool, which will display the theoretical material and practical tasks implemented on it, on the basis of which knowledge testing will take place.

3ds Max is a full-featured professional software system for working with three-dimensional graphics. With it, we can create an object for a military simulator, for example a tank. Adobe Flash Player creates an application for simulate the operation of military equipment. Thanks to the 3Ds Max and Adobe Flash Player collaboration, you can get a quality Training Device.

#### REFERENCES:

- [1] Operator training simulator for antitank guided missile "konkurs". [Online]. Available: [http://simulator.ua/files/tp/konkurs\\_en.pdf](http://simulator.ua/files/tp/konkurs_en.pdf). [Accessed Apr. 7, 2018].
- [2] O.M. Droban, B.S. Fedor, P.M. Evdokimov, AA Zvonko, and O.G. Yakubovsky, Building and operation of antitank rankat complex 9p149, 2015. p. 245
- [3] K. Grover, Flash CSS. Practical Guide, 2011.
- [4] . Lotte and K. Rotondo, Adobe AIR. A Practical Guide to the Environment for Flash and Flex Desktop Applications, 2009.
- [5] Adobe Creative Team. ActionScript 3.0 for Adobe Flash Professional CS5 Classroom in a Book. Adobe Press, 2011, p. 432.
- [6] "What is 3ds max? An introduction". [Online]. Available: <http://www.tutorialboneyard.com/3ds-max-introduction/>. [Accessed Apr. 7, 2018].
- [7] "3D modeling, animation & rendering using autodesk 3Ds Max". [Online]. Available: <http://www.goalsinfocloud.com/courses/3d-modeling-animation-rendering-using-autodesk-3ds-max/>. [Accessed Apr. 7, 2018].



# Information System for Account of Students Work

Maria Mova<sup>1</sup>, Volodymyr Nahorny<sup>2</sup>, Olga Zagovora<sup>3</sup>, Maksym Portiano<sup>4</sup>

<sup>1,2</sup>Sumy State University, Ukraine, <sup>2</sup>v.nahorny@cs.sumdu.edu.ua

<sup>3</sup>GESIS Leibniz Institute for the Social Sciences, Germany, Olga.Zagovora@gesis.org

<sup>4</sup>Fanshawe College of Applied Arts and Technology, Canada

**Abstract** – The goal is to create an information system for an account of students' work. The system should be easily accessible and provide complete information for learning process. To realize this idea were chosen Android OS, programming language Java, development environment Android Studio and three-level client-server architecture.

**Keywords** – information system, Android OS, mobile application, account of students work.

## I. INTRODUCTION

Significant changes in the information sphere of human activity lead to a significant reduction in the effectiveness of outdated approaches to controlling the student's educational process. That is why today it is expedient to develop a special system that will allow to automate the process of accounting of students' work and make it faster, more convenient and reliable.

For this purpose, the task was to develop an information system for student records. It is decided to develop this system as a mobile application with the ability to integrate with a web-based analogue. This is due to the fact that the number of smartphone users grows year by year and the proportion of users among young people aged 18 to 29 years is about 70%. Such popularity can be explained by the need for modern people to always be in touch and have access to the right information.

## II. STUDENT WORK INFORMATION SYSTEM

### A. Application requirements

A functional requirement, in software and systems engineering, is a declaration of the intended function of a system and its components. Based on functional requirements, a programmer determines the behavior (output) that a device or software is expected to exhibit in the case of a certain input.

The application must ensure that below functions:

1. Support for several types of users: students and teachers.
2. The presence of a personal user profile and the ability to edit personal data.
3. Ability to review the list of disciplines and their regulations (students).
4. Ability to view group list (students).

5. The ability to view current progress and attending classes (students).
6. Ability to create discipline and its regulations (teachers).
7. Ability to create a lesson (teachers).
8. Ability to register current student attendance and student visits in real time (teachers).
9. The possibility for teachers to perform active operations of the system (editing, adding and deleting records in the database).

### B. Choice of mobile operating system

According to 2018 market share of mobile operating systems in Ukraine as illustrated in figure 1 [1]. The most popular OS among the major OS is Android which holds 76%, next popular is iOS with 13% share and then comes the Windows.

Implement a mobile application was decided on the basis of the Android platform because this platform will provide access to the largest number of users.

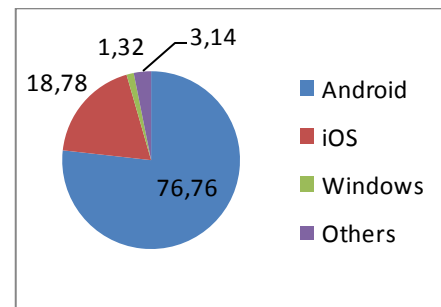


Figure 1. Market share of mobile operating systems in Ukraine

### C. Choice of programming language

The languages you might consider learning for Android development include [2]:

- Java. Java is the official language of Android development and is supported by Android Studio;
- Kotlin. Kotlin was recently introduced as a secondary “official” Java language. Kotlin is free and open source. But Kotlin is very young. It was officially released in 2016.
- C/C++. C++ is also an another programming language used to develop Android Apps. But

sometimes the Native Development Kit is not as efficient as SDK since it doesn't support all API frameworks. Thus targeting the native layer is not completely possible using NDK. Therefore preferring Java over C++ will be a better option.

- **C#.** C# is alternative to C or C++. It's supported by some very handy tools like Unity and Xamarin which are great for game development and for cross-platform development.
- **Lua.** Lua is an old scripting language that was originally designed to supplement programs written in more robust languages like C, VB.NET, etc. That being said, Lua can be used as a main programming language in certain cases, and Corona SDK is one great example. With Corona, you can create rich apps using Lua that can be deployed on Windows, Mac, Android, iOS, and even Apple TV and Android TV.
- **HTML + CSS + JavaScript.** For development on HTML, CSS and JavaScript use the PhoneGap Build environment or, in a more specialized case, Adobe Cordova.

Of the programming languages presented, Java was chosen. Advantages of using Java are it gives high performance, complete security of your program because it uses its own run-time environment without any interaction with the system OS, official documentation is presented for this language and a huge amount of different teaching materials.

#### D. IDE

Let's consider two main programming environments for writing Java applications: Eclipse and Android Studio.

Android Studio provides the fastest tools for building apps on every type of Android device. Android Studio is the official integrated development environment (IDE) for Google's Android operating system. Android Studio supports a number of programming languages, e.g. Kotlin, Java, C++ etc.

Eclipse is a powerful, open source, integrated development environment (IDE) that facilitates the creation of desktop, mobile, and web applications. Eclipse is a highly versatile and adaptable tool. As such, you can use many types programming languages and plugins like Ruby, C++, Java, PHP, COBOL etc.

It was chosen the software environment of Android Studio because it is now the main development environment for developing applications for Android.

The Android Studio main window is identified in figure 2.

#### E. Three-level client-server architecture

Three-level client-server architecture was chosen to implement the application, which will allow simultaneous access to the system of many users.

Client-server architecture is architecture of a computer network in which many clients (remote processors) request and receive service from a centralized server (host computer) [3].

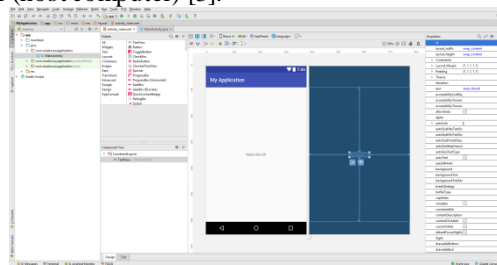


Figure 2. The Android Studio main window.

Three-level client-server architecture is special type of client-server architecture consisting of three well-defined and separate processes, each running on a different platform:

1. The user interface, which runs on the user's mobile (the client).
2. The functional modules that actually process data. This middle tier runs on a server and is often called the application server.
3. A database management system (DBMS) that stores the data required by the middle tier. This tier runs on a second server called the database server.

Three-level client-server architecture as illustrated in figure 3.

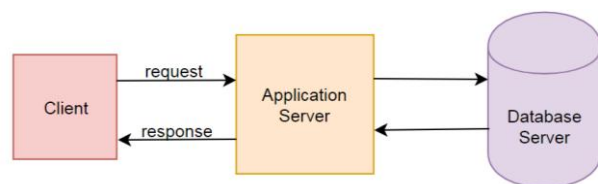


Figure 3. Three-level client-server architecture

#### F. Design

For the Android application design, the Material Design principles were used.

Material Design principles [4]:

1. Material is the metaphor

A material metaphor is the unifying theory of a rationalized space and a system of motion. Our material is grounded in tactile reality, inspired by our study of paper and ink, yet open to imagination and magic.

2. Surfaces are intuitive and natural

Surfaces and edges provide visual cues that are grounded in our experience of reality. The use of familiar tactile attributes speaks to primal parts of our brains and helps us quickly understand affordances.

3. Dimensionality affords interaction

The fundamentals of light, surface, and movement are key to conveying how objects interact. Realistic lighting shows seams, divides space, and indicates moving parts.

#### 4. One adaptive design

A single underlying design system organizes interactions and space. Each device reflects a different view of the same underlying system. Each view is tailored to the size and interaction appropriate for that device. Colors, iconography, hierarchy, and spatial relationships remain constant.

#### 5. Content is bold, graphic, and intentional

Bold design creates hierarchy, meaning, and focus. Deliberate color choices, edge-to-edge imagery, large-scale typography, and intentional white space create immersion and clarity.

#### 6. Color, surface, and iconography emphasize actions

User action is the essence of experience design. The primary actions are inflection points that transform the whole design. Their emphasis makes core functionality immediately apparent and provides waypoints for the user.

#### 7. Users initiate change

Changes in the interface derive their energy from user actions. Motion that cascades from touch respects and reinforces the user as the prime mover.

#### 8. Animation is choreographed on a shared stage

All action takes place in a single environment. Objects are presented to the user without breaking the continuity of experience even as they transform and reorganize.

#### 9. Motion provides meaning

Motion is meaningful and appropriate, serving to focus attention and maintain continuity. Feedback is subtle yet clear. Transitions are efficient yet coherent.

User experience (UX) design [5] is the process of creating products that provide meaningful and relevant experiences to users. This involves the design of the entire process of acquiring and integrating the product, including aspects of branding, design, usability, and function.

Prototypes for application screens as illustrated in figures 4-6.

### G. Mobile application development

The client is a mobile application. The application server is web server. As a database, MySQL is used.

To administer the MySQL database, phpMyAdmin is used. phpMyAdmin is a free and open source administration tool for MySQL.

To interact with the server in Android was using the library Volley.

Volley [6] is an HTTP library that makes networking for Android apps easier and most importantly, faster.

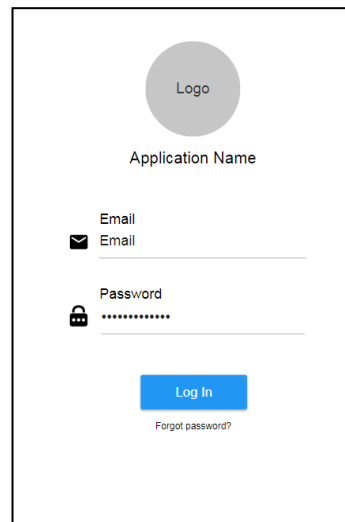


Figure 4. Login

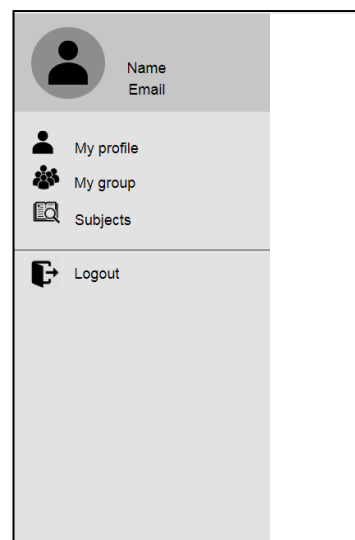


Figure 5. Navigation drawer

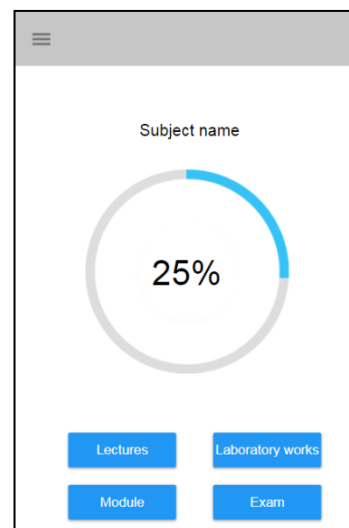


Figure 6. Subject

Volley offers the following benefits:

- Automatic scheduling of network requests.
- Multiple concurrent network connections.
- Transparent disk and memory response caching with standard HTTP cache coherence.
- Support for request prioritization.
- Cancellation request API. You can cancel a single request, or you can set blocks or scopes of requests to cancel.
- Ease of customization, for example, for retry and backoff.
- Strong ordering that makes it easy to correctly populate your UI with data fetched asynchronously from the network.
- Debugging and tracing tools.

Basic diagram of Volley as illustrated in figure 7.

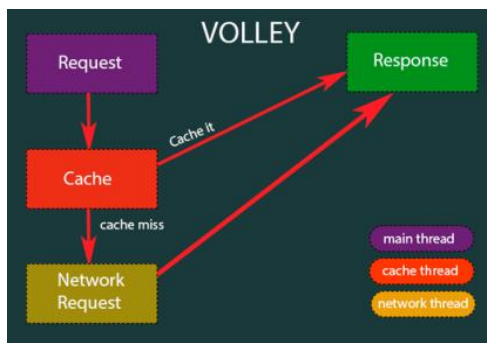


Figure 7. Basic diagram of Volley

The default request classes already included in Volley library are String request, JSON request, and image request [7].

String Request is type of request is used to retrieve data from a server as a String.

Volley provides the following classes for JSON requests:

- JSONArrayRequest – a request for retrieving a JSONArray response body at a given URL.
- JSONObjectRequest – a request for retrieving a JSONObject response body at a given URL, allowing for an optional JSONObject to be passed in as part of the request body.

JSON (JavaScript Object Notation) is a lightweight data-interchange format [8].

Format JSON as illustrated in figure 8.

Image Request is type of request is used to load images in Android.

```
{ "users":[
  {
    "firstName":"Ray",
    "lastName":"Villalobos",
    "joined": {
      "month":"January",
      "day":12,
      "year":2012
    }
  },
  {
    "firstName":"John",
    "lastName":"Jones",
    "joined": {
      "month":"April",
      "day":28,
      "year":2010
    }
  }
]}
}
```

Figure 8. Format JSON

## CONCLUSIONS

The information system for the account of students' work in the form of a mobile application for OS Android is developed.

The developed mobile application has the following functions: the use of a database to store information system, support for several types of users, the ability to register the current progress and student visits to students in real time, the ability to review the rules, the ability of one of the types of users to use the active system options (editing, adding and deleting entries in the database).

## ACKNOWLEDGMENT

The work was performed as part of the state budget scientific research theme of the Sumy State University, "Models and information design and management technologies in complex systems" (state registration number 0115U001569).

## REFERENCES:

- [1] J.Lee, "To Build an Android App, You Need to Learn These 7 Programming Languages", makeuseof.com, Mar. 6, 2017. [Online]. Available: <https://www.makeuseof.com/tag/build-android-app-programming-languages/>. [Accessed Apr. 2, 2018].
- [2] Statista, "Market share of mobile operating systems in Ukraine from 2010 to 2018", Statista, 2018. [Online]. Available: <https://www.statista.com/statistics/669506/market-share-mobile-operating-systems-ukraine/>. [Accessed Mar. 28, 2018].
- [3] A. Berson, Client/Server Architecture. New York, US: McGraw-Hill (Tx), 1992.
- [4] T. Schropfer. Material Design. Basel, CH: Birkhäuser Architecture, 2010.
- [5] A. Cooper, R. Reimann, D. Cronin, and C. Noessel. About Face: The Essentials of Interaction Design. Hoboken, NJ: Wiley, 2014.
- [6] Developers, "Transmitting Network Data Using Volley", Developers, 2018. [Online]. Available: <https://developer.android.com/training/volley/index.html>. [Accessed Apr. 7, 2018].
- [7] A.Saini, "Volley Tutorial With Example In Android Studio", June, 2017. [Online]. Available: <http://abhiandroid.com/programming/volley>. [Accessed Mar. 30, 2018].
- [8] JSON, "Introducing JSON", JSON, 2017. [Online]. Available: <https://www.json.org/>. [Accessed Apr. 7, 2018].

# Information Technology for Modeling of Human-Machine Interactions

Evgeniy Lavrov<sup>1,2</sup>, Nadiia Pasko<sup>3</sup>, Yan Voitsekhovskiy<sup>1</sup>, Ruslan Plaks<sup>1</sup>, Galyna Mihalevska<sup>4</sup>

<sup>1</sup>Sumy State University, Ukraine, <sup>3</sup>Sumy National Agrarian University, Ukraine,

<sup>4</sup>Khmelnitsky National University, Ukraine

<sup>2</sup>prof\_lavrov@hotmail.com, <http://www.famous-scientists.ru/13333>

**Abstract – Questions of ergonomic support in automated systems are considered. The problem of reducing human operator errors is considered. Human-machine interaction is described using functional networks. The method of Professor A. Gubinsky is used. Information technology for assessing the reliability of human-machine interaction was developed. Possible application of the results is described.**

**Keywords – automated system, information technology, effectiveness, distribution of functions, man-operator, ergonomics, algorithm of functioning**

## I. INTRODUCTION

From 50 to 80% of the accidents in production systems of different types, more than 64% of accidents in the sea fleet and 80% in aviation caused by errors of the human operator [1,2].

In many cases, operators work under stress [3-5].

The aim of Research in the field of designing human-machine systems (HMS) is to reduce erroneous reaction of the human operator, to ensure acceptable conditions of work and to adapt the machine to the man-operator [6-8].

The effectiveness of ergonomic research essentially depends on the ability to formalize the interaction between man and machine and to receive prompt assessment of options for activities to solve the optimization problem [8-10].

## II. RESULTS

Statement of the problem. To develop an interactive system for simulating human-computer interaction in HMS.

As a basic methodology was selected the functional-structural theory of HMS [10].

We obtained new mathematical models to ensure rapid computer simulation [11-13]:

- system analysis for HMS (component and morphological);
- language for discrete human-computer interaction's description;

- automatic assessment of the reliability of the functional network;
- select optimal variant of human-computer interaction in different formulations for different tasks of ergonomics:
- to determine the degree of automation,
- calculation of the number of operators,
- the choice of activity structures and methods of operations,
- the distribution of functions between operators,
- defining activities ergonomic quality assurance systemothers.

We have implemented a computer technology, that allows to simulate interactively the HMS:

- to describe;
- to evaluate;
- to optimize.

interaction of operators with software and hardware tools in HMS.

Evaluation of the functional network is based on the technology of functional structures typing and function network folding.

The structure of the software is shown in Fig. 1. (FN-functional activity network, TFS-type functional structure of activities, TFU is a typical functional activity unit).

The software allows you to describe the activities of a human operator in a dialogue mode.

It is possible to store data on the reliability of the individual operator's actions.

The system "knows how" to perform a syntactic analysis of activities. Automatic recognition of TFS is provided.

This allows automatic reduction of the functional activity network.

Thus, the assessment of the reliability of the operator's activity is carried out automatically.

There is a bank of optimization models for human-machine interaction.

Optimization can be done on the "work graph" or on the "event graph".

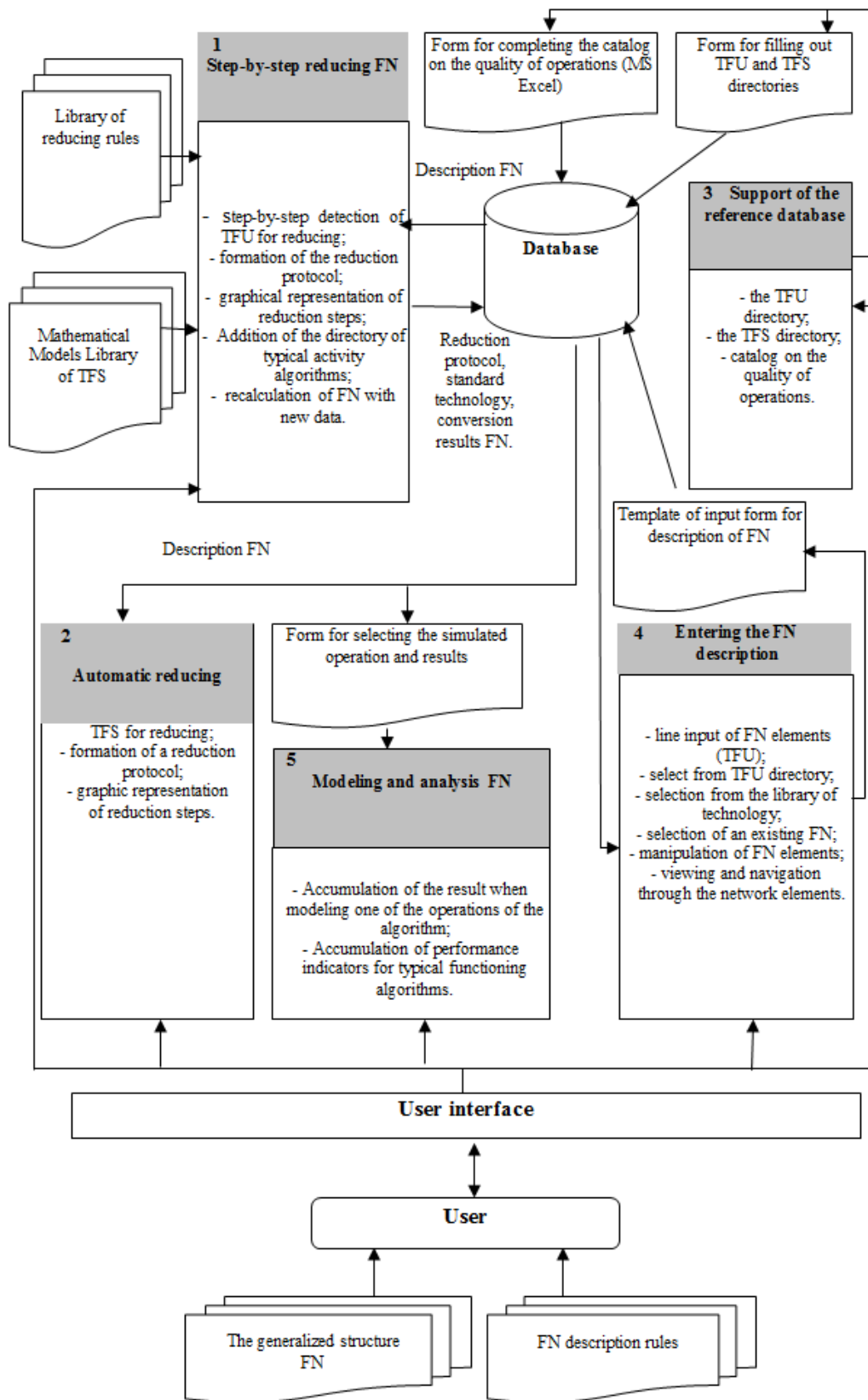


Figure 1. The structure of the software

It is possible to model the process of human-machine interaction with various initial data.

All models can be stored in a special database. This makes it possible to use the created models to develop new systems.

It is possible to assess the working conditions in the workplace of human operators.

This allows to calculate the initial data on the reliability and timing of individual actions of the human operator, taking into account the impact of working conditions and harmful factors.

For this, special correction factors are calculated.

The system allows to find ergonomic reserves to increase the efficiency of automated systems.

This reduces the number of errors, accidents and improves the health of people.

### III. APPROBATION

The results are used in the design and operation of processes in different systems [10-13]:

- Flexible manufacturing systems;
- Safety management systems;
- Incident Management Systems;
- Telecommunication systems;
- Call-Centers and Contact-centers;
- E-learning systems.

and in the learning processes for courses:

- “Modeling of processes and systems”.
- “Ergonomics”.
- Decision theory.
- Simulation of human-machine interaction

in universities:

- Sumy National Agrarian University.
- Sumy State University.
- Kharkiv Engineering and Educational Academy.
- National University of Biological Resources and Nature Management (Kiev).

### IV. DIRECTIONS FOR FURTHER RESEARCH

In the future, it is necessary to conduct new scientific research in such areas:

- Modeling of errors of different types.
- Functional networks with queues.
- Study of the operational-tempo tension of operators.
- Creation of real decision support systems that work in real time.

### REFERENCES:

- [1] P. Rothmorea, P. Aylwardb, and J. Kamon, “The implementation of ergonomics advice and the stage of change approach”, *Applied Ergonomics*, vol. 51, pp. 370–376, 2015. doi: 10.1016/j.apergo.2015.06.013
- [2] T. A. Bentley, S. T. T. Teo, L. McLeod, F. Tana, R. Bosua, and M. Gloet, “The role of organisational support in teleworker wellbeing: A socio-technical systems approach”, *Applied Ergonomics*, vol. 52, pp. 207–215, 2016. doi: 10.1016/j.apergo.2015.07.019
- [3] Y. Wang, L. Zheng, and T. Hiu, “Stress, burnout and job satisfaction: case of police force in China”, *Public Personnel Management*, vol. 43, no. 3, pp. 325–339, 2014. doi: 10.1177/0091026014535179
- [4] F. De Felice and A. Petrillo, “Methodological Approach for Performing Human Reliability and Error Analysis in Railway Transportation System”, *International Journal of Engineering and Technology*, vol. 3, no.5, pp. 341–353, 2011.
- [5] P. C. Cacciabue, “Human error risk management for engineering systems: a methodology for design, safety assessment, accident investigation and training”, *Reliability Engineering & System Safety*, vol. 83, no. 2, pp. 229–269, 2014. doi: 10.1016/j.res.2003.09.013
- [6] A. N. Anokhin, “Otechestvennaja jergonomika i jergonomicheskoe soobshhestvo: sostojanie i napravlenija razvitiija”, *Chelovecheskij faktor: problemy psihologii i jergonomiki*, vol. 1, no. 68, pp. 4–15, 2014. (in Russian)
- [7] J. Dul, R. Bruder, P. Buckle, P. Carayon, P. Falzon, and W. S. Marras, “A strategy for human factors/ergonomics: developing the discipline and profession”, *Ergonomics*, vol. 55, no. 4, pp. 377–395, 2012. doi: 10.1080/00140139.2012.661087
- [8] A. Anokhin and A. Ivkin, “Evaluation of ecological interface design for supporting cognitive activity of nuclear plant operators”, In Proceedings of the 5th International Conference in Applied Human Factors and Ergonomics 2014 and the Affiliated Conferences, 2014, pp. 260–270.
- [9] A. Anokhin, I. Gorodetskiy, V. Lvov, and P. Paderno, “Education and professional development of ergonomists in Russia”, In Proceedings of the 5th International Conference in Applied Human Factors and Ergonomics 2014 and the Affiliated Conferences, 2014, pp. 1017–1024.
- [10] A.N. Adamenko, A.T. Asherov, and E.A. Lavrov, *Information controlling human-machine systems: research, design, testing, Reference book*, Gubinsky A.I. and Evgrafov V.G., eds, Moscow: Mashinostroenie, 1993, pp. 528. (in Russian).
- [11] E. Lavrov, N. Pasko, A. Tolbatov, and V. Tolbatov, “Ergonomic Reserves for Improving Reliability of Data Processing in Distributed Banking Systems”, In Proceedings of 2nd International Conference on Advanced Information and Communication Technologies-2017 (AICT-2017), 2017, pp. 79-82.
- [12] E. Lavrov, N. Pasko, A. Krivodub, and A. Tolbatov, “Mathematical models for the distribution of functions between the operators of the computer-integrated flexible manufacturing systems”, In proceedings of the XIII-th international conference tcset’2016 “Modern problems of radio engineering telecommunications, and computer science”, 2016, pp. 72-76.
- [13] E. A. Lavrov and N. B. Pasko, “Podhod k formirovaniju banka optimizacionnyh modelej dlja raspredelenija funkcij mezhd operatorami ASU”, *Vostochno-Evropejskij zhurnal peredovyh tehnologij. Serija «Matematika i kibernetika – fundamental’nye i prikladnye aspekty»*, vol. 1, no. 49, pp. 46-50, 2011. (in Russian).

# The Simulator on the Topic “Land And Doig Method” of the Distance Learning Course “Optimization Methods and Operations Research”: Development and Software Realization

Oleksandr Syvokin<sup>1</sup>, Oleg Iemets<sup>2</sup>

Poltava University of Economics and Trade, Ukraine, <sup>1</sup>syvokin94@gmail.com, <sup>2</sup>yemetsli@ukr.net

**Abstract** – The report presents the simulator on the topic “Land and Doig Method”. In the report, the relevance and scope of the simulator is justified.

**Keywords** – simulator, digital programming, land and doig method, PLP, PILP.

## I. INTRODUCTION

For the tasks of integer programming, one can apply an algorithm that was proposed by A. N. Land and A. G. Doig in 1960. This approach represented a scheme, which in a generalized form was later called the “method of branches and bounds”. The simulator was created by the algorithm previously developed by the authors. [1].

The simulator is created for teaching the Land-Doig method. In the development of the simulator, examples have been used for the topic “Land and Doig's Method” from the course “Methods of Optimization and Operations Research” [2].

In works [3-16] the simulators for the distance course “Methods of optimization and research of operations” and others are considered. As in the distance course on “Methods of optimization and operations research” of PUET and on the Internet (in the Ukrainian segment), there is no simulator on the subject “Land and Doig method”, therefore the creation of such a simulator is relevant.

The practical value of the work is that this simulator will teach the student the method of Lend and Doig for the problems of integer programming, which in turn will allow to learn how to maximize profits in the production of a product or minimize the cost of resources.

This simulator allows to ease the burden on the teacher, since students can learn from the program in their classroom, as well as remotely at home, by mastering the Land and Doig method, having the relevant knowledge and skills in solving problems by simplex-method and M-method.

In the implementation of the simulator, the development environment of NetBeans IDE 8.2 was

used. When writing the program, a high-level programming language Java was used.

This program can be presented either as a Java applet to facilitate its placement on the website of distance learning, or in the form of .exe file for download and installation on the computer. This is also good for computers that do not have JVM and JRE installed, which will allow you to open Java applications without a development environment. Installing the program bypasses the register and avoids conflicts with antivirus programs, which is also a positive.

## II. MAIN RESEARCH

The simulator is divided into two parts. Initially, the student is given the opportunity to answer the test tasks (Fig. 1) in the first three steps, which gradually lead to the beginning of tasks solving and give an idea of what the next tasks will be related to. The second part is a task for solving examples and filling tables (Fig. 2). This will allow the student to consolidate his knowledge of solving problems using the simplex-method and the M-method, as well as to some extent to study the method of Land and Doig.

It also provides guidance if the wrong data are entered in the table (Fig. 3). This makes it easy to find the error, and also saves time for the transition to the answer to the following tasks that go immediately after checking for the correctness of the data entered in the table (Fig. 3).

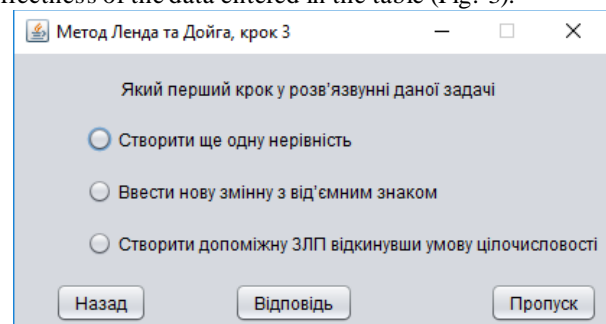


Figure 1. An example of a test task.



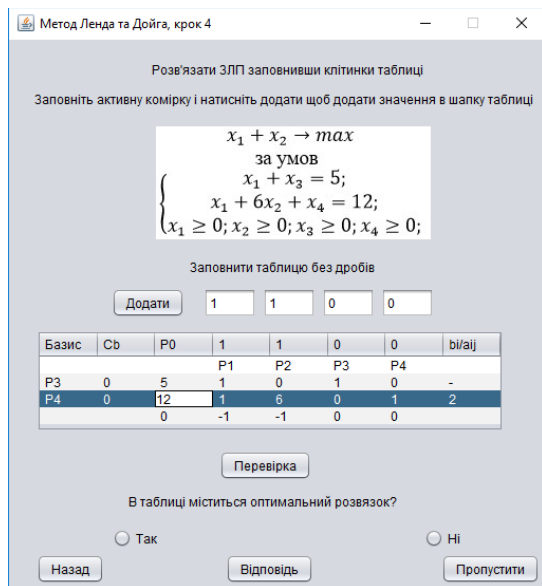


Figure 2. An example of filling the table of data.

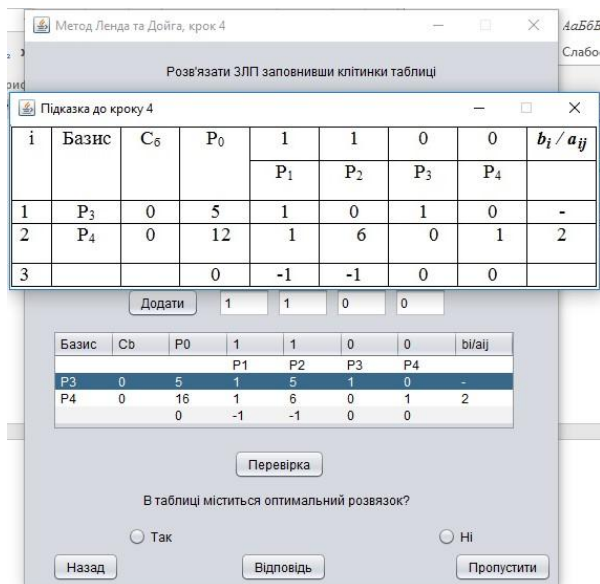


Figure 3. Prompt if error is in the table

The ability to move from step to step, without solving any tasks that allow the student to continue his education from the place where he stopped last time is added to the simulator.

Below is an algorithm for the simulator.

First, the student can see a window with the name of the simulator, its authors and a button to start (Fig.4).

Step 1. A window opens to the student, asking: "For which tasks the method of Land and Doig is used", and four variants of the answer are offered a) PLP; b) PILP, c) PDLP; d) PNP. If the answer is selected, the student clicks on the "Reply" button. If the answer is correct, then b) the transition to the next step. If the answer is not correct, an informative window about the incorrect

answer appears, the student closes it and makes the appropriate conclusions, chooses the correct answer.



Figure 4. Startup simulator window.

Step 2nd. Before the student opens the next window, which shows the task (Fig. 5).

$$x_1 + x_2 \rightarrow \max$$

за умов

$$\begin{cases} x_1 + x_3 = 5; \\ x_1 + 6x_2 + x_4 = 12; \\ x_1 \geq 0; x_2 \geq 0, x_3 \geq 0, x_4 \geq 0; \\ x_1, x_2, x_3, x_4 - \text{цілі.} \end{cases}$$

Figure 5. The task in step 2.

The screen asks if the task can be solved by the Land and Doig method.

After the answer is checked and, go to the next step. If the answer is incorrect, that is, "no", an informative window about the infidelity of the answer that needs to be closed and the correct option appears. If the answer is yes, then the next step is moving.

Step 3rd. The student is asked a question that relates to the task given in step 2: "What is the first step in solving this problem?" Three variants of the answer are given: 1) to create another inequality; 2) enter a new variable with a negative sign; 3) create an auxiliary PLP by rejecting the condition of the integer. After selecting the correct (3rd) option, the student moves to the next step of the simulator, if the choice is not correct, the student remains in the third step until he chooses the correct answer.

Step 4. The student is given the opportunity to enter data in a standard table according to a given example, which is visible on the screen. After filling the table, the student clicks the "Check" button to check that the spreadsheet is correct. If the entered data are correct, the informative window "data entered correctly" opens. Then you have to answer the question: "Does the table contain an optimal solution?". If the answer is correct (no, then), go to the next step. If the answer is not correct,

an information window will appear about it. The student is forced to draw conclusions and give the correct answer.

Step 5-6. In these two steps the student continues to enter the data into new simplex tables based on the simplex table completed in step 4. After verifying the correctness of the entered data in the table, the student will also have to answer the questions posed below the table, as in step 4. When the student answers the questions, (for Step 5, "What is a guideline?" For step 6, "Has the optimality criterion been fulfilled?"), He will be able to move on to the next step.

Step 7. In the new window, you must enter a solution to the LP (namely: "x1, x2, x3, x4, fmax"), which the student solved within 4-6 steps. After entering the data, the student clicks the "Reply" button, and if the data entered is correct, moves to the next step. In case of non-fidelity of the entered data, an error window opens and the student is forced to enter the correct ones.

Step 8. At this step, you must break the task into two by the rule of the method. After the data entered by the student, if they are correct, we will proceed to the next step. In case of non-fidelity, we remain at this step.

Step 9. After the task is divided into two, the student is given the task of solving PILP1 (Fig. 6). After the data is entered and checked, the student responds to the question that is found below, namely: "Does the table contain the best solution?". After the correct answer is selected ("no"), go to the next step.

$$\begin{aligned} x_1 + x_2 &\rightarrow \max \\ x_1 + x_3 &= 5 \\ x_1 + 6x_2 + x_4 &= 12 \\ x_2 + x_5 &= 1 \\ x_1, \dots, x_5 &\geq 0 \end{aligned}$$

Figure 6. The task in step 9.

Step 10-11. In these two steps the student continues to enter the data into new simplex tables, based on the data of the simplex table, considered at the 9th step. After verifying the accuracy of the data entered in the data table, the student will also need to answer the questions posed below the table, as in step 9 (for step 10: "Select a guiding column"; for step 11 " Does the table contain the optimal solution? ") When a student answers questions, he will be able to move on to the next step.

Step 12. A window appears where you need to enter "x1, x2, x3, x4, F1max" for the LPP, which the student solved in steps 9-11. After entering the data, the student clicks the "Check" button and in case of the correct data entered, the student answers the question "Should I break the task into two?". If a student chooses the correct answer "no", he is moving on to the next step. If he does not select the correct answer, an error window opens. And the student is forced to choose the correct answer.

Step 13. We begin to solve the PILP2 (Fig. 7). After the data is entered into a table and a check is made, the student responds to a question that is found below, namely: "What column is a guide?". After choosing the correct answer go to the next step. If the answer is incorrect, the student remains at the step 13.

$$\begin{aligned} x_1 + x_2 + (-M)x_6 &\rightarrow \max \\ x_1 + x_3 &= 5 \\ x_1 + 6x_2 + x_4 &= 12 \\ x_2 - x_5 + x_6 &= 2 \\ x_1, \dots, x_6 &\geq 0 \end{aligned}$$

Figure 7. The task in the step 13.

Step 14-15. At these two steps, the student continues to enter data into new simplex tables based on the results of the simplex table at step 13. After verifying the correctness of the entered data in the table, the student will also need to answer the questions posed below the table. When the student correctly answers the questions (for step 14: "Which line is the guideline?" For step 15: "Has the optimality criterion been fulfilled?"), He will be able to move on to the next step.

Step 16. In the window that opens, you must enter "x1, x2, x3, x4, F2max" for the PLP, which the student solved within 13-15 steps, as well as compare the values of the target functions. After entering the data, the student clicks the "Check" button and in case of the correct data entered, the student answers the question "Should I break the task into two?". If a student chooses the correct answer "no", he is moving on to the next step. If he does not select the correct answer, an error window opens. The student is forced to choose the correct answer.

Step 17. The student must compare the value of the target functions F1max and F2max by entering the "<", ">" or "=" sign. If the student chooses the correct sign, namely ">", then the window appears at the bottom of the page with the conclusion: "Thus, the solution of the original problem is the solution of PLP1 (F1max > F2max) (Fig. 8). At this step, the work on the simulator is completed.

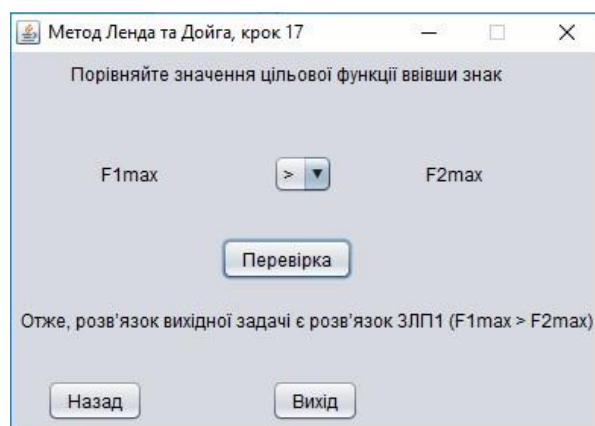


Figure 8. The final step in the solution of the problem.

The designed simulator is intended to be introduced into the educational activities of PUET.

### CONCLUSIONS

In the course of the work, the goal was achieved - a simulator was created that would be used in teaching students to solve problems using the method of Land and Doig, based on the knowledge gained during the study of the simplex method, the M-method. It is planned to create other simulators for those topics for which they were not created. It is also planned to translate simulators into English, for wider use in the world.

### REFERENCES:

- [1] O. Y. Syvokin and O. O. Iemets, "Development and program realization of the simulator on the topic "Land and Doig method of" the distance course", In proceedings of Computer Science and Applied Mathematics (CS&AM-2018), 2018. (in Ukrainian).
- [2] O. O. Iemets and T. O. Perfonova, "Methods of optimization and operations research". [Online]. Available: [http://elib.puet.edu.ua/action.php?kt\\_path\\_info=lm.web.view&fDocumentId=670571](http://elib.puet.edu.ua/action.php?kt_path_info=lm.web.view&fDocumentId=670571). [Accessed: April. 30, 2018]. (in Ukrainian).
- [3] O. O. Iemets, eds., "Computer Sciences and Systems Science (CS&SS-2015): Materials of the VII All-Ukrainian Scientific and Practical Conference on International Participation", Poltava : PUET, 2017.
- [4] K.G. Glynsky and O. O. Iemets, "Simulator on the topic "Dual simplex method" of the distance learning course "Methods of optimization and operations research", in From Effective Governance to Effective Economics: a collection of scientific articles of the masters of the Faculty of Economics and Management based on the results of scientific studies 2012-2013 academic year. Poltava: PUET, 2013, pp. 220-223. (in Ukrainian).
- [5] A.M. Koinash. "Development of the simulator software on the topic "Simplex-method" of the distance learning course "Methods of optimization and operations research", in Proceedings of the V science-practice conference "Computer Science and Systems Science, CS&SS, 2014. pp. 153-154. (in Ukrainian).
- [6] O. S. Ovcharenko, "Algorithmization and software of the simulator on the theme "Jordan-Gauss method" of the distance learning course "Methods of optimization and research of operations", in Proceedings of the V science-practice conference "Computer Science and Systems Science, CS&SS, 2014, pp. 231-233. (in Ukrainian).
- [7] S. M. Samovyk, "Development of the algorithm and software of the simulator on the topic "Hungarian method in the task of assigning" the distance learning course "Methods of optimization and operations research", in Proceedings of the V science-practice conference "Computer Science and Systems Science, CS&SS, 2014, pp. 273-275. (in Ukrainian).
- [8] A.B. Smymov, "Algorithmization and programming of the simulator on the topic "Ford-Falkerson method" the distance learning course "Methods of optimization and operations research", in Proceedings of the V science-practice conference "Computer Science and Systems Science, CS&SS, 2014, pp. 287-289. (in Ukrainian).
- [9] S.I. Usolcev, "Development of the algorithm, program and study simulator on the theme "Modified simplex method" the distance learning course "Methods of optimization and operations research", in Proceedings of the V science-practice conference "Computer Science and Systems Science, CS&SS, 2014, pp. 309-310. (in Ukrainian).
- [10] V.R. Soingar, "Software for simulator "Matrix Games" the distance learning course "Methods of optimization and operations research in Proceedings of the V science-practice conference "Computer Science and Systems Science, CS&SS, 2014, pp. 294-296. (in Ukrainian).
- [11] M. Y. Stavkovy, "Development of the simulator on the topic "Method of analysis of hierarchies" the distance learning course "Methods of optimization and operations research", in Proceedings of the V science-practice conference "Computer Science and Systems Science, CS&SS, 2014, pp. 294-296. (in Ukrainian).
- [12] V. A. Kuchugur, "Algorithmization and program implementation of the simulator on the topic "The Brown-Robinson Method" the distance learning course "Methods of optimization and operations research", in Proceedings of the V science-practice conference "Computer Science and Systems Science, CS&SS, 2014, pp. 172-173. (in Ukrainian).
- [13] M. P. Kryklia, "Development of algorithm for the simulator on the topic "Graphical method of solving linear programming problems", in Proceedings of the V science-practice conference "Computer Science and Systems Science, CS&SS, 2014, pp. 163-165. (in Ukrainian).
- [14] A. V. Dushynska, "Development of the program-simulator on the topic "Duality in linear programming" the distance learning course "Methods of optimization and operations research", in Proceedings of the VI science-practice conference "Computer Science and Systems Science, CS&SS, 2015, pp. 143-146. (in Ukrainian).
- [15] K. V. Gryshyna, "Development of software simulator on the topic "Potential methods for transport task" the distance learning course "Methods of optimization and operations research", in Proceedings of the V science-practice conference "Computer Science and Systems Science, CS&SS, 2014, pp. 76-79. (in Ukrainian).
- [16] O. V. Sokol and O. O. Chemenko, "Development of the simulator on the topic "Normal algorithms" of the distance learning course "Theory of algorithms", in Proceedings of the VIII science-practice conference "Computer Science and Systems Science, CS&SS, 2017, pp. 247-252. (in Ukrainian).
- [17] S. V. Pedorenko and O. O. Iemets, "Development of the program-simulator for the M-method for the distance learning course "Methods of optimization and operations research", in Proceedings of the VI science-practice conference "Computer Science and Systems Science, CS&SS, 2016, pp. 226-230. (in Ukrainian).
- [18] V. V. Kylnyk, "The algorithm of the simulator on the topic "Sorting by the method of mixing", in Informatics and Systems Science (ISS-2017): materials of the 9th All-Ukrainian Scientific and Practical Conference on International Participation 2017, pp. 135 – 137. (in Ukrainian).
- [19] D. A. Marchenko, Development of the program-simulator on the topic "Transitional polyhedra", in Informatics and Systems Science (ISS-2017): materials of the 9th All-Ukrainian Scientific and Practical Conference on International Participation 2017, pp. 190 – 192. (in Ukrainian).
- [20] O. O. Poterailo and O. O. Iemets, "Programming of the study simulator for the gradient method of optimization of nonlinear functions of the distance learning course "Methods of optimization and operations research", in Proceedings of the VIII science-practice conference "Computer Science and Systems Science, CS&SS, 2017, pp. 235-225. (in Ukrainian).

# Peculiarities of Distance Learning Technologies Application in Training Future Officers for the Armed Forces of Ukraine

Olena Pavlichenko<sup>1</sup>, Yurii Kalynovs'kyi<sup>2</sup>, Halyna Serhieieva<sup>3</sup>

<sup>1</sup>Ivan Kozhedub Kharkiv National University of Air Force, Ukraine, <sup>2,3</sup>Yaroslav Mudryi National Law University, Ukraine  
<sup>1</sup>elena\_aleksandro@ukr.net, <sup>2</sup>kalina\_uu@ukr.net, <sup>3</sup>g.a.sergeyeva@gmail.com

**Abstract** – The article analyses intrinsic features and principles of distance education. In particular, the following principles of distance education are regarded as the basic ones: humanitarianism; priority of a pedagogical approach while planning educational process; pedagogical appropriateness of using new information technologies; selection of education content; initial educational level; conformity of education technology with specific learning goals; ensuring security of information circulating in the system of distance learning; use of visual aids; systematic and sequential nature; scientific nature; simplicity; communicativeness; mobility. The study examines peculiarities of a distance learning management system MOODLE utilization for military specialists' training at Ivan Kozhedub Kharkiv National University of Air Force, exemplified by an academic subject "History of Ukraine and Ukrainian Culture".

**Keywords** – distance education, information and communication technologies, innovative education, traditional education, distance course, the MOODLE platform, cadets' individual work.

## I. INTRODUCTION

The increase in informatization of modern society has had a profound influence on the process of learning. Distance learning, a method of study based on information and communication technologies extensively used in many countries of the world, is among the recent teaching tools, which are being intensively introduced into education space nowadays.

Use of distance modes and methods for teaching academic disciplines facilitates individualization of the process of prospective specialists' professional becoming, motivates them to complete individual assignments, forms information culture, inclines towards mastering innovative techniques to acquire and use knowledge. Current capabilities of distance learning fully meet social and

professional demand with regard to the future specialists' qualification.

It is necessary to note that issues related to introducing distance learning modes and methods into the domestic educational process are covered in publications by such Ukrainian scientists as Unhurian [10], Veremchuk [11], Blazhuk [2], etc.. The current state and peculiarities of introduction of distance education to professional training of future military specialists for the armed forces of Ukraine was the subject of the scientific works by Hudym [5], Artabaiev [1], Stasiuk [8], Oksiuk [1], Andrushchenko [1], etc.

Considerable attention to the above-mentioned issues is also given in foreign scientific studies. Berankova and Zerzanova [4], namely, showed good prospects of applying distance education methods to effective introduction of the military terminology course. Whereas, in the scientific research by Cao and Shen [3] the problem of professional training of technicians for military equipment by means of a distance learning system are analysed. The works by Tung, Huang, and Keh [9] reveal specific character of training and retraining of Taiwan Armed Forces officers with the use of distance learning technologies. Miller and Tucker [7] in their study, report the findings about applying distance learning technologies to military leaders training aimed at developing their critical thinking, intercultural competences as a basis for making decisions under high psychological stress. Researchers Kovanda and Zuna, [6] discovered advantages and disadvantages of implementing a distance learning programme based on the MOODLE platform for Czech Armed Forces senior military officers at the Defence University of Brno.

At the same time, the problem of applying distance learning technologies to training future officers for the Armed Forces of Ukraine, particularly the MOODLE platform, has not been studied appropriately yet. Precisely this goal our scientific research will be devoted to. To achieve the aim mentioned above, we intend to carry out a number of research objectives: to outline principal

distinctions between traditional and distance education, revealing the nature of the latter in more detail; to show innovation of applying distance courses to training prospective officers, particularly on the MOODLE platform; to identify problems which, owing to various factors, students of military schools are faced with while using information and communication technologies for distance learning; to determine factors which impede the development of distance education in today's Ukraine.

## II. METHODS

To carry out a comprehensive study into the scientific issue in question, it is reasonable to employ a set of general scientific, philosophical and special methods, where the leading one is a dialectical method, which enables to infer cause-and-effect relationships and identify contradictions regarding distance learning introduction to domestic military education. Attention should be paid to a comparative method, with the help of which we will discover peculiarities of various distance learning technologies and compare traditional and innovative approaches to future officers training. Of fundamental importance is also a system method, which lies in considering distance education in the context of common institutional logic of the whole teaching and educational process at a higher military education institution. This method constitutes a basis for identifying interrelation between methods of distance and traditional education. Employing communicative and discourse approach, we can see mechanisms of information exchange between lecturers and cadets, how the latter master new terminology and make progress in choosing and selecting information necessary for their self-study.

## III. RESULTS AND DISCUSSION

The fundamental distinction between distance education and its conventional forms is that the basis of the latter is formed by independent cognitive activity of a learner. It is important that subjects of cognition not only absorb certain knowledge, but also learn to acquire it on their own, process information, master learning styles, which could be used later on within the conditions of continuous self-studying.

The characteristic features of distance education are: flexibility, as students study where and whenever convenient to them rather than attend scheduled classes in the form of lectures and training workshops; modularity, which means that distance education is based on a modular approach and each course provides insight into a certain field of knowledge; cost-effectiveness, as experience of domestic non-governmental centres shows their expenses

approximately amount to 60 % of expenses involved in training specialists in the case of full-time study; the new role of a lecturer whose functions are as follows: coordination of cognitive process, course design and correction, advice on drawing up individual education plan, etc.; special-purpose control forms such as on-line interviews, tests, practical assignments and course works are used as forms of control; use of specialized technologies: distance education technology is a set of methods, forms and tools of interaction in the process of independent but controlled acquisition of certain amount of knowledge [10].

One can assess quality of adoption and application of distance education using such characteristics as: the outcomes (level of knowledge acquisition, ability to apply stored knowledge to practice, academic achievements, individual learning process, flexible tutorials); accessibility for all segments of the population (students, businessmen, disabled persons, personnel of the armed forces, and prisoners have equal opportunities to study on-line); resource intensity (no need for attending lectures and training workshops, financial expenditure, material resources, lecture-rooms, lecturers, etc.); expeditiousness (amount of time necessary to acquire knowledge, to inform students, etc.); democratic relation "lecturer – student"; packaged software; advanced education technologies [11].

Proceeding with previously mentioned theoretical speculation, we turn our attention to practical aspects of the distance learning technologies implementation in the military education system of Ukraine. In particular, integration of the military education system of the Ministry of Defence of Ukraine into the state-run system of education on the basis of the common legislation and regulatory framework presupposes application of recent learning technologies based on high-performance computer and telecommunication technologies. The point at issue is such mode of military specialists' training as distance learning, which can be introduced into education process owing to internationally known learning management system MOODLE ("Modular Object-Oriented Dynamic Learning Environment").

Modern peculiarities of cadets' training at higher military education institutions of the Ministry of Defence of Ukraine are determined primarily by complicated military-political situation in Ukraine and on its borders that implies periodic special tactical training exercises of cadets involving field days, interruptions in their study for various periods of time.

In the circumstances of constant interruptions in studies, the need for distance education becomes urgent, and, moreover, taking into consideration further tasks of strengthening combat power and capabilities of the Armed Forces of Ukraine, introduction of distance learning into the

practice of cadets' training at higher military education institutions of the Ministry of Defence of Ukraine is rapidly moving from the sphere of theory to its direct practical realization.

In our opinion, distance learning courses for cadets of higher military education institutions of the Armed Forces of our country should become commonly used and undergo constant improvement; that will greatly increase the level of knowledge gained and provide good opportunities for self-education of future officers. Distance learning may also be helpful in retraining and advance training of militaries in various professional directions and at different qualification levels. Distance learning should be an essential component of military training of students studying at higher education establishments according to reserve officers programme (especially those who are beyond the bounds of a military education institution), in preliminary training and testing candidates for entry to military education institutions, when training foreign militaries; for providing civilian education to militaries.

Defining terminology of the research, we note that a distance course (hereinafter - DC) is an integrated and structured set of information, electronic learning resources and educational and methodological tools necessary to master learning material according to the programme of an academic discipline, available by means of the Internet (local network) using Web browser and/or other software tools at the disposal of a user.

Taking into account specific character of a distance mode of study, e-learning materials of a DC should encompass the following components: general information on the academic discipline; a set of recommendations for cadets on sequence and terms allocated to study both the material of the discipline in whole and its individual components; methodic guidelines on the performance of various types of academic work, practical tasks and individual work; general criteria for assessment of knowledge, abilities and skills which should be acquired by a cadet in the process of training; forms and terms of interactive communication of cadets and a teacher who implements the functions of active learning; a set of assignments and tests to check and assess knowledge, abilities and skills of cadets while studying the discipline (formative assessment) and at the end of their studies (summative assessment); examples of the correct answers to all tasks for checking cadets' knowledge; web-resources and software support necessary to perform practical types of academic work; system of "student – student" interactive communication.

The content of a distance course should be organized in accordance with the programme of the respective academic discipline. Learning material of an academic discipline is

divided under a modular approach; modules should include educational materials on the topics.

Distance summative assessment of knowledge, abilities and skills is conducted in accordance with the academic programme of the discipline in the form of a credit or exam by means of information and communication technologies provided that the learner is authenticated.

This part should include: chat for consultations on preparation for summative assessment; list of questions to prepare for summative assessment; methodic guidelines to undertake a summative assessment assignment; an example of the correct answer to summative assessment assignment; methodic guidelines on summative assessment test performance; list of final testing questions; test to examine knowledge of a discipline; a sample of an examination card or credit assignment, and a pattern of completing an examination or credit assignment.

These requirements for the organization of distance courses should be obligatory taken into account when developing an academic discipline as a distance course. When analysing successful experience of implementing distance learning, we note that teaching staff of Ivan Kozhedub Kharkiv National University of Air Force are actively using the MOODLE platform in teaching process today.

The use of the distance learning technologies can be exemplified by the course "History of Ukraine and Ukrainian Culture". This platform, in view of the peculiarities of the education institution, is currently being used as a complementary tool in class. A wide range of multimedia materials (documentary films, on-line excursions, audio fragments, presentations), hosted on a web site and used in class, makes the process of studying "History of Ukraine and Ukrainian Culture" far more interesting and effective.

Furthermore, the platform ensures cadets' individual work. Each of them receives an individual creative assignment (preparation of multimedia materials about their native cities), completes it at a convenient time and sends to a teacher for review. The distance course in question also includes supplementary self-study materials for cadets (extended presentations on the topics, further reading, current researches, and documentary films). Knowledge assessment involves testing, which, at the discretion of the teacher, the cadets can take both in writing and with the aid of the platform. If testing is by means of the platform, the results are automatically registered by the system.

All teaching aids available on the platform have been developed with consideration of modern ways of presenting information, which increase the degree of its perception and learning. The offered training tools are divided into two groups: multimedia (text, presentations, photomaterials, video clips, and audio fragments) and hypermedia, which

are the most promising forms of combining heterogeneous information; a nonlinear form of arrangement of the material (hypertext, hypertext links), which makes it possible to use a considerable amount of reference information and different learning paths.

It is worth noting, that among the assignments, designed for cadets' individual work, there are both traditional tasks, that is, those which should be carried out according to a certain algorithm, and innovative ones which completion is based on cadets' personal experience and requires their creative approach. Conventional assignments are completed according to the instructions specified in the task: to fill in a comparative table, to study primary sources according to the plan, to compile a chronological table on the specified topic. Among the creative assignments, which are given to cadets, there is a task to collect material for presentation on the history of their native cities using multimedia. On completion, the assignment is sent to the teacher for checking and then is to be defended before the audience.

Thus, in the current conflictogenic military-political situation in Ukraine to train the cadets of the higher military education institutions of the Ministry of Defence of Ukraine, it is necessary to design and implement distance courses of academic disciplines to ensure continuous and quality education of future military specialists. The distance course of an academic discipline must be designed in conformity with the teaching package of the discipline for full-time study; it must contain a set of certain components to ensure that cadets acquire all of the skills and abilities in this discipline.

#### CONCLUSIONS

Summing up the findings of the research, we can state that distance education differs from traditional one, primarily, in its greater mobility, flexibility and accessibility. Distance learning provides great facilities for cadet's self-education and, consequently, creates opportunities for social adaptation, mental and intellectual development, etc. Use of the MOODLE platform at Ivan Kozhedub Kharkiv National University of Air Force, as it has been demonstrated by our research, enabled development of graded courses with various structure and of different levels of complexity for cadets' learning, which made their study considerably more active and optimal. Introduction of the distance course "History of Ukraine and Ukrainian Culture" facilitated improvement of the cadets' individual work. At the same time not all the cadets and lecturers are equally ready to use technologies of on-line training. This problem is caused by age peculiarities, different level of technical competence of learners and educators, wrong mental attitude, etc. Insufficient level of teachers' skills and the problem of their training and retraining, undue bureaucracy with respect to distance

learning methods implementation, underfunding of distance technologies development, late renewal of computer equipment and resource base, a partial solution of the problem with providing cadets and professors with constant access to the Internet are among objective reasons impeding more dynamic introduction of distance learning technologies into the domestic Ukrainian education space.

#### REFERENCES:

- [1] Yu.Z. Artabaiev, O.H. Oksiiuk, and V.M. Andrushchenko, "Prospective systems of distance learning for military higher education establishments," *Suchasni informatsiyni tekhnolohiyi u sferi bezpeky ta oborony*, vol. 3, pp. 45-48, 2009. (in Ukrainian)
- [2] O.A. Blazhuk and N. Ostrovs'ka, "Distance education is a modern form of higher education," *Naukovyy visnyk Uzhorod's'koho natsionalnoho universytetu. Seriya : Pedahohika. Sotsialna robota*, vol. 30, pp. 28-30, 2014. (in Ukrainian)
- [3] J.H. Cao and Y. Shen, "Military Professional Distance Training Based on E-Learning Technology," in *Proceedings of International conference on economic, education and management (ICEEM 2011)*, Macao, China, 2011, pp. 305-308.
- [4] I. Cechova, J. Berankova, and D. Zerzanova, "E-learning technologies: military terminology course development," in *Proceedings of 18th International Conference on Knowledge-Based Organization: economic, social and administrative approaches to the knowledge-based organization*, Sibiu, Romania, 2012, pp. 888-89.
- [5] V.I. Hudym, and B.I. Shuneych, "The state of distance learning introduction at the higher military school of Ukraine," *Visnyk Kyivs'koho natsionalnoho universytetu imeni Tarasa Shevchenka. Viys'kovo-spetsial'ni nauky*, vol. 28, pp. 41-44, 2012. (in Ukrainian)
- [6] P. Kovanda, and P. Zuna, "Moodle supported distance learning study program: project of the officers' career education," in *Proceedings of the Conference of Information and Communication Technology in Education, Roznov pod Radhostem, Czech Republic, 2012*, pp. 131-138.
- [7] J.W. Miller and J.S. Tucker, "Addressing and assessing critical thinking in intercultural contexts: Investigating the distance learning outcomes of military leaders," *International journal of intercultural relations*, vol. 48, pp. 120-136, 2015. doi: 10.1016/j.ijintrel.2015.07.002.
- [8] V.V. Stasiuk, "Humanization and humanization of military education system: contemporary requirements," *Visnyk Natsionalnoho universytetu oborony Ukrainy*, vol. 1(32), pp. 155-160, 2013. (in Ukrainian)
- [9] M.C. Tung, J.Y. Huang, H.C. Keh, and S.S. Wai, "Distance learning in advanced military education: Analysis of joint operations course in the Taiwan military," *Computers & education*, vol. 53(3), pp. 653-666, 2009. doi: 10.1016/j.compedu.2009.04.003.
- [10] L.M. Unhurian, H.V. Chemetska, and I.A. Naumenko, "Distance learning of extramural students of pharmaceutical faculty at a medical university," *Medychna osvita*, vol. 3, pp. 95-97, 2013. (in Ukrainian)
- [11] A. Verenchuk, "Problems and prospects of distance learning at higher education institutions," *Problemy pidhotovky suchasnoho vchytyelya*, vol. 7, pp. 319-325, 2013. (in Ukrainian)

**SESSION 4**  
**INFORMATION SYSTEMS ENGINEERING**



# Automated Recording and Processing of Lecture Attendance Data Using RFID Student Cards

Alexander Kalashnikov<sup>1</sup>, Hongwei Zhang, Misko Abramriuk  
Sheffield Hallam University, United Kingdom  
<sup>1</sup>a.kalashnikov@shu.ac.uk, <http://tinyurl.com/lokvnb>

**Abstract**—Aspects of the ongoing project to automate the recording of students' lecture attendance using RFID cards and automate processing of the records without a university wide infrastructure are presented and discussed.

**Keywords** – students' attendance monitoring, RFID cards, automated data recording, automated data processing.

## I. INTRODUCTION TO THE PROJECT

### A. Project Motivation

Often when completing an assignment in the lab a student would ask “I do not understand how does X work, could you please explain?” Sometimes this X was discussed in detail during a full lecture. A short answer would leave the student unsatisfied, but going through the lecture for just one student would leave little lab time available to the others.

With the lecture attendance register on hand, the lecturer could reply “We have dedicated to X a whole lecture, which you unfortunately have missed. Here in the lab I have Y minutes per a student on average. If I could present X in Y minutes, I would do it at the lecture at the first place. Please have a look at the lecture handouts first, some additional resources are mentioned at the learning portal for the module, and I will be happy to have a short discussion afterwards”. This would not only help the student who missed the lecture, but also be fair to the students who were present for the topic in lecture theatre.

### B. Reasons for not monitoring students' attendance

There are good reasons *not* to monitor students' attendance. First, any monitoring procedure would take time and attention away from the lecture's content. It would also unnecessary reduce flexibility for time planning of some generally well attending students, lowering their satisfaction. Additionally, some students may believe (rightly or not) that they can cover the lecture material on their own, especially if they perceive (again, rightly or not) the lecture's content or the presenter as not addressing their needs sufficiently. Finally, students at this education level are believed to be responsible adults who do not need their attendance to be policed.

### C. Reasons for monitoring students' attendance

Most people would agree that better attendance leads to better attainment. Surprisingly, convincing studies on this topic with solid evidence are difficult to come by. A strong correlation between attendance and attainment is evident from a reliable subject neutral dataset, collected from several school children of various ages [1]. Whether attendance is the only or the most important factor in greater attainment cannot be judged purely from the correlation data. Both attendance and attainment could be simultaneously influenced by some fundamental, difficult to quantify causes such as student ambition, background, attitude *etc.*

Another reason for introducing attendance monitoring is related to the students' behaviour as a group. When some students can consistently get away with non-attendance, it lowers the morale and motivation of the attending students.

There are several instances when attendance monitoring at an educational establishment is required by law in the UK. Parents are held legally responsible for their children's attendance at school, and can be fined by local authorities for non-attendance of more than five days or arrival at the school after the close of registration [2]. The number of parents fined was approximately 90,000 in the academic year 2015/16 [3]. Truancy cases are continuously contested in courts (nearly 20,000 cases in England in 2015 alone) with the most famous case being that of Jon Platt. He, despite several preceding court rulings in his favour, was eventually found guilty of truancy by the Supreme Court at a total cost to the taxpayer of £140,000 [4]. Another example of legally required attendance monitoring concerns students outside of the European Economic Area who arrive in the UK on Tier 4 (general) student visas [5]. The educational institutions, which sponsor these students by issuing unconditional offers to study, must monitor the students' attendance. An institution may lose their sponsoring status (and thus the ability to admit new foreign students) if they fail to produce attendance records on request. Finally, attendance monitoring is required for higher degree apprenticeships currently being developed. Students' learning time is paid for by their employer, and

proper timekeeping becomes an obligatory part of the contract among the parties involved.

Another consideration is fairness. In most workplaces not attending a scheduled activity is misconduct. If tuition fees are paid out of public or sponsored funds, not attending scheduled learning activities could be considered misconduct as well. Even if funding comes from a commercial student loan or family funds, most future employers or customers would like to have assurances that their surgeons or pilots spent all the scheduled hours learning their trade.

#### D. Methods to monitor students' attendance

At present paper registers are used at our university (SHU). The register contains a photograph of each student and signature fields for scheduled sessions. Sometimes the register is passed from one student to another and some sign for themselves and also for their friends. If each student comes to the lectern to sign in person, a substantial time is required.

Alternatively, the lecturer can call every student and mark the register herself/himself. This procedure becomes time-consuming for large classes where attendance rates are usually lower compared to the small group classes where each student is visible to the teacher clearly. The lecturer could do random checks a few times during the term and use the sample data, for example, to allocate additional questions at oral examinations. However oral exams have become a rarity, thus this approach has lost its appeal to lecturers.

The paper register could also be kept by appointed student representatives, but this approach is not 100% accurate either.

A reduction in time, along with increased accuracy of recording, could be expected if automated tools quickly identified any student in the class. One can envisage simply taking a picture of the class and identifying all the students present by some means of artificial intelligence. However, such a scenario is unlikely to become viable in the near future, and a biometric or unique token identification is used instead. Biometric sensors (such as fingerprint or iris) require more time to operate compared to the alternative, and might raise privacy concerns.

On the other hand, each student is supplied with a student ID card, which is used for borrowing books from the library, accessing premises, *etc.* Common student ID cards use barcodes (ubiquitous for goods in shops) or RFID (radio frequency identification) tokens, or both (for example, SHU cards). Barcode readers require close optical proximity to the card and particular viewing angles; RFID tokens are non-sensitive to the orientation, thus should operate faster.

The token identification system can be deployed over the entire institution (like the University of Westminster, UK) but deploying and maintaining the required infrastructure in every teaching room may be rather costly. Alternatively, a lecturer could have a standalone

reader with sufficient memory to take records in the teaching room and process them later in her/his office.

#### E. Aim of the project

To develop a quick, easy to use procedure for recording student attendance at lectures using their RFID cards, and later automate the processing of the collected data. Hardware should be fully integrated and autonomous, and any software should not be tied to a particular platform or operating system and require minimal effort to install, maintain and use.

## II. PASSIVE RFID BASICS

Passive RFID cards were developed to facilitate small frequent payments, e.g., for transport fares [6]. The cards are based on an electronic circuit with an antenna that absorbs energy from an RFID card reader, located in close proximity, to power the card. The antenna is also used to return requested data to the reader by altering levels of energy, absorbed from the reader.

There are three distinct operating frequency bands utilised with RFID cards of different types. Moreover, within every frequency band there are several non-compatible data exchange protocols introduced by different manufacturers and later standardised internationally (such as ISO 14443 standard with type A and type B cards [6]).

The most commonly used RFID cards seem to be Mifare Classic cards, compliant to ISO 14443 type A specifications. These cards feature globally unique four-byte card numbers, allowing the use of approximately 4.3 billion unique cards. More recent RFID cards utilise considerably longer addresses, allowing more cards to be used globally. The address is not hidden or encrypted, thus can be read by any compatible RFID reader. There are reports that addresses of some Mifare cards can be set at will [7]; therefore, these addresses themselves should not be used with anything related to security or value.

In addition to their unique addresses, the cards have encrypted read-write non-volatile memory, where, for example, personal or financial information can be stored securely. The level of security, offered by common RFID cards, is trusted by transport operators for fare collection (Oyster cards are used by Transport for London), defence ministries for granting access to restricted premises (Common Access Card in the USA), leading financial institutions for authorising payments (Visa and Master Card) and many governments (biometric passports). The details of encryption are proprietary for obvious reasons (security by obscurity), and RFID card issuers are not keen to make any of these public.

Widespread use of RFID technology brought about the associated risk of card misuse or fraudulent access to stored information. At the consumer level there is the possibility that fake purchase authorisation can be initiated by a fraudster, equipped with a mobile RFID payment terminal. Cautious RFID card users who do not

want to risk a contactless payment taken covertly can place a piece of aluminium foil, acting as an electromagnetic energy shield, into their wallets. This makes the interaction of a covert RFID payment terminal with their cards more difficult. To limit the possibility of accessing sensitive information, the US Department of Defence ordered over 4 million sleeves to shield the Common Access Cards of military personnel in 2010 [8]. Because of potential vulnerabilities, any RFID reading equipment should be sufficiently trusted not to read, alter or erase any information stored on the card. An additional complication in the case of SHU cards is their use as a payment card at the university. This increases both the cards' vulnerability to security breaches and potential liability for altering or damaging the card by faulty or malicious RFID card reader. If damage occurs during attendance recording liability needs to be met by the card's manufacturer, SHU card vendor, RFID reader manufacturer, or the software/firmware developer. The university's or lecturer's liability can be avoided completely if all transactions are conducted by an external RFID card vendor, maintaining and operating the complete RFID card infrastructure, but this is costly. We believe the use of open source hardware and software for the reader, that can be audited independently if required, can be an assurance of the reader's safety for all relevant stakeholders (students, RFID card vendor, and university administration).

Using the above considerations we concluded that using student RFID cards as unique tokens for attendance monitoring is viable, but the associated development should be based on open source designs as much as possible to avoid potential security breaches and excessive liabilities.

### III. PAST PROGRESS AND OUTCOMES

The viability of the project was established in the academic year 2015/16, when a student as his final year project developed a device, which used workplace issued RFIDs to lock and unlock equipment for safety purposes. It was found that an inexpensive MRF522 module could reliably read the SHU RFID card numbers. An MSc student completed the design of an Arduino-based RFID card reader as part of his project in January 2017. Since then the reader has been used to record attendance at lectures for a significant number of students (50–60). The development was reported at the annual SHU teaching and learning conference in June 2017 [9]. In December 2017 a departmental teaching enhancement grant was awarded to support this project.

Continuous use of the card reader led to hard learned lessons:

- the register is best taken at the end of the session when the students want to leave the room as quickly as possible; taking it at the start of the class is not convenient as some students will be late; taking it in the

middle requires more time as students tend to get their card out only when the lecturer is in front of them;

- make sure students leave the room after swiping their card without attempting to swipe a card for another student; the best place to take the register is near the room's exit, if students need to come to the lectern or somewhere else it causes overcrowding;

- students occasionally forgot their cards, damaged the card's antenna, or did not have valid cards at all. There must be a paper register for such students to write their names, which is to be processed manually;

- the prototype device consisted of three pieces—the device enclosed in a splash-proof case, a power bank, and USB A to micro USB B cable. Occasionally one piece of this kit was forgotten in the office and automated registration could not take place;

- processing the collected data in real time was very complicated due to time pressure and difficulty in developing intuitive software to handle the required procedures;

- some students wanted to swipe their wallet but the presence of other RFID cards tended to confuse the reader as collision detection was not easy to implement in the firmware.

These experiences reinforced the need to have a single piece of equipment that is lightweight, battery operated and can be held single-handed.

### IV. RECENT PROJECT DEVELOPMENTS

The teaching enhancement grant was awarded to develop a better integrated front end electronic RFID reader for raw data collection without inconveniences (a single piece of equipment) and design flaws (despite the MFRC522 module reliably operated with 5V Arduino Uno digital output pins, this voltage level was not specified by the manufacturer of the integrated circuit - IC). Additionally, the back end software to aggregate the collected data into a common Excel spreadsheet needed to be developed.

#### A. Custom built RFID logger

A few options for the sensing IC and associated module were considered, namely MFRC522 (reads type A tokens), PN532 (type A and B), and THM360 (type A and B). Any of these ICs can be used to build proximity readers, which read tokens at short distances. A much better range, desirable for this application, was observed in CR95HF transceivers, which read not only types A and B tokens but also support longer range ISO standards for vicinity readers. There are code examples and hardware options available to implement an RFID reader based on this IC (for example, various software, hardware and firmware options listed on the product page [10] and third party compact PCB module [11]). This transceiver must be complemented by a microcontroller (MCU) in order to build a standalone device. We selected STM32F103 MCU from the same manufacturer STMicroelectronics.

LiFeO<sub>4</sub> rechargeable cell with a compatible charger powers the design. This cell chemistry was selected as its voltage is safe to use with both the reader and MCU ICs without any additional circuitry, allows more charge cycles and the cells are generally safer to use compared to cells utilising other energy denser Li-ion chemistries.

### B. Off-the-shelf RFID reader

This option was only attempted for completeness but was found quite robust and convenient for the application. Some of the purchased RFID readers cost as little as £10 and operated as a standard keyboard without the need to download, install or utilise any proprietary drivers, software or software development kits. Of the three readers purchased one operated perfectly with every computer tried, one occasionally put an extra new line character inside the read number, and one was found operating at the frequency of 125 kHz. Sometimes readers of the same design are offered for either 125 kHz (labelled ID) or 13.56 MHz (labelled IC). Two ICs were found inside the 13.56 MHz readers. Unfortunately, their labels (SYC8P1213 and SYC8P1223) were unknown to Google. We believe these readers can be used safely as they do not seem to do anything except read RFID card numbers. However, having control over the reader's firmware (i.e. operating a custom fully developed reader) is a more reliable option to avoid potential liability and data leakage.

The best reader was tested alongside the prototype and recorded the same card numbers in the right sequence. The card numbers were reported in decimals to an Android-based smartphone through a USB on-the-go (OTG) cable. The phone was running a text editor application and the file with the recorded numbers was saved, then emailed to the office computer for later processing. An added convenience of the off-the-shelf reader was an audio beep in addition to flashing LEDs when the number is read successfully.

### C. Back end software

We aimed to develop software that can be deployed on a variety of computers (including single board computers like the Raspberry Pi) and operating systems, hence selected Python as the programming language. The attendance records are commonly kept using Excel spreadsheets and can be accessed from custom Python code using a free and open source *openpyxl* library written in pure Python (thus there were no dependencies on other modules). The developed code operates as follows. The user copies the recorded numbers and pastes these into a new column of the spreadsheet following previously processed data. This column is copied into a reserved column by the software, then cleared. Each entry in the copied column is checked against each entry in the valid RFID numbers column, which must be prepared in advance for every class. If the entry is not valid then an error message is generated. Otherwise the entry is placed

in the same row where the valid number was found. At the end of processing the numbers for every session will be placed in the same row and blank cells will represent absences. This custom Python code was tested with the data sets recorded over many lecture sessions, and operated correctly for the class size of approximately 60 students.

## CONCLUSIONS

We believe that encouraging student attendance at lectures by keeping a record is beneficial for learning, the morale of the class, and society. To reduce the time and effort required in keeping attendance records, we are developing hardware and software for individual automated use by lecturers. It reads RFID student card numbers in the class and processes the stored data in an Excel spreadsheet. At this stage we have fully prototyped a custom RFID reader and successfully used an off-the-shelf RFID reader with the developed software for classes of 60 students attending 12 lecture sessions for two academic years.

## ACKNOWLEDGMENT

This project progressed to its present stage with the help of a grant awarded from the departmental teaching enhancement fund, which is gratefully acknowledged.

## REFERENCES:

- [1] The link between absence and attainment at KS2 and KS4, 2012/13 academic year. Research report by Department for Education, Feb 2015. [Online]. Available: <http://tinyurl.com/y98lg8od>. [Accessed: April. 30, 2018].
- [2] Poor attendance at school - penalty notices, Tameside Metropolitan Borough: Education. [Online]. Available: <http://tinyurl.com/y9xbx8pe>. [Accessed: April. 30, 2018].
- [3] 90,000 parents fined for taking children out of school in term-time, The Guardian, 7 Jul 2016. [Online]. Available: <http://tinyurl.com/zzxjpu5>. [Accessed: April. 30, 2018].
- [4] Isle of Wight term-time holiday dad Jon Platt guilty, BBC news, 23 Jun 2017. [Online]. Available: <http://tinyurl.com/y7rfnal0>. [Accessed: April. 30, 2018].
- [5] Tier 4 (General) student visa, GOV.UK: visas and emigration. [Online]. Available: <http://tinyurl.com/oh8nrwe>. [Accessed: April. 30, 2018].
- [6] J.Wehr, IS THE DEBATE STILL RELEVANT? An in-depth look at ISO 14443 and its competing interface types, Secure ID News, 01 Jul 2003. [Online]. Available: <http://tinyurl.com/y8sqjhan>. [Accessed: April. 30, 2018].
- [7] M.Balboa, Arduino RFID Library for MFRC522 # Security. [Online]. Available: <http://tinyurl.com/o9kr7ra>. [Accessed: April. 30, 2018].
- [8] Z.Martin, Defense Department order RF shields from National Laminating, Secure ID news, 29 Nov 2010. [Online]. Available: <http://tinyurl.com/o9kr7ra>. [Accessed: April. 30, 2018].
- [9] R.Canhye and A.N.Kalashnikov, Improving consistency of student experiences by encouraging their attendance using automated attendance recording, presented on 29 June 2017. [Online]. Available: <http://tinyurl.com/ycaev6wq>. [Accessed: April. 30, 2018].
- [10] CR95HF NFC reader IC, product page. [Online]. Available: <http://st.com/en/nfc/cr95hf.html>. [Accessed: April. 30, 2018].
- [11] BM019: Serial to Near Field Communication (NFC) module, product page. [Online]. Available: <http://solutions-cubed.com/bm019/>. [Accessed: April. 30, 2018].

# Hierarchical Algorithm of the Machine Learning for the System of Functional Diagnostics of the Electric Drive

Anatoly Dovbysh, Victoria Zimovets<sup>1</sup>  
Sumy State University, Ukraine, <sup>1</sup>zc.vika@gmail.com

**Abstract** – The hierarchical algorithm of machine learning in the framework of the information-extreme intellectual technology, which is based on maximizing the system information capacity in the process of its training, is considered. The algorithm is implemented on the example of the machine learning of the system of functional diagnostics of the electric drive of a mine winder.

**Keywords** – machine learning, information-extreme intellectual technology, functional diagnostics, information criterion, electric drive.

## I. INTRODUCTION

Methods of the functional diagnostics, in contrast to the test, do not require a temporary equipment withdrawal from the operation. In addition, functional diagnostics can predict the dynamics of the development of individual damage and evaluate the residual equipment resources [1, 2]. Intelligence information technologies are widely used to improve the accuracy and reliability of the functional diagnostics systems [3, 4]. But most of the known methods of data mining do not take into account the junction of the recognition classes, arbitrary initial conditions of the technological process cycles, and are sensitive to increasing the alphabet power of the recognition classes. One of the promising ways to increase the functional efficiency of functional diagnostics systems is to use the ideas and methods of the so-called information-extreme intellectual technology (IEI-technology) of the machine learning [5].

The article deals with the hierarchical information-extreme algorithm of the machine learning for the system of functional diagnostics of an air clutch electric drive.

## II. FORMALIZED PROBLEM SETTING

Let the structure of the recognition classes be  $\{X_{h,s,m}^o | h = \overline{1, H}; s = \overline{1, S}; m = \overline{1, M}\}$ , where  $H$  is the number of tiers of the hierarchical structure;  $S$  is a number of strata on  $h$  tier;  $M$  is the number of classes of recognition in the  $S$  stratum. According to the results of the periodic registration of the diagnostic features in the process of all technological cycles of moving the skip

according to the speed chart, the input learning matrix has been formed for each recognition class  $\|y_{h,s,m,i}^{(j)} | i = \overline{1, N}, j = \overline{1, n}\|$ , where  $N$  is the number of recognition attributes in the vectors-implementation of the recognition class  $X_{h,s,m}^o$ ;  $n$  is the number of vectors-implementation, which is equal to the number of periods of the information reading from the sensors on all the technological cycles of the air clutch.

A structured vector of training parameters of the system of functional diagnostics is defined to recognize the vectors-implementation class  $X_{h,s,m}^o$ :

$$g_{h,s} = \langle x_{h,s,m}, d_{h,s,m}, \delta_{K,h,s,m,i} \rangle, \quad (1)$$

where  $x_{h,s,m}$  is a binary averaging vector-implementation, which vertex determines the geometric center of the hyperspherical container of the recognition class  $X_{h,s,m}^o$  in the binary space of diagnostic signs;

$d_{h,s,m}$  is the code distance that defines the radius of a hyperspherical container of class  $X_{h,s,m}^o$ ;  $\delta_{K,h,s,m,i}$  is a parameter that equals a half of the symmetric field of the acceptance tolerances of the  $i$  diagnostic sign of an averaged vector-implementation  $x_{h,s,m}$  of the recognition class  $X_{h,s,m}^o$ .

A limitation is applied:

$$d_{h,s,m} \in [0; d(x_{h,s,m} \oplus x_{h,s,c}) - 1],$$

where  $d(x_{h,s,m} \oplus x_{h,s,c})$  is a code distance between the averaged vector-implementation  $x_{h,s,m}$  of the recognition class  $X_{h,s,m}^o$  and vector-implementation  $x_{h,s,c}$  of the nearest neighbor recognition class  $X_{h,s,c}^o$ ;

$$\delta_{K,h,s,i} \in [0; \delta_{E,h,s,i} / 2],$$

where  $\delta_{E,h,s,m,i}$  is the operation tolerances field of the  $i$  diagnostics sign of an averaged vector-implementation  $X_{h,s,m}$  recognition class  $X_{h,s,m}^o$ .

At the stage of machine learning it is necessary to optimize the parameters of the vector (1) by the average information criterion

$$\bar{E}_{h,s} = \frac{1}{M} \sum_{m=1}^M \max_{G_E \cap G_d} E_{h,s,m}(d_{h,s,m}), \quad (2)$$

where  $E_{h,s,m}(d_{h,s,m})$  is an information criterion for optimizing the machine learning parameters of the functional diagnostics system to recognize the vector-implementation of class  $X_{h,s,m}^o$ , which value is calculated at each step of machine learning when the radius of the container of the recognition class changes;  $G_E$  is working (admissible) area of definition of the information criterion of optimization;  $G_d$  is permissible range of radius change for the hyperspherical container of the recognition class.

At the exam stage, that is, directly in the operational mode of the functional diagnostics, a decision is made on the affiliation of the current recognized vector to one of the recognition classes of a given hierarchical structure.

### III. CATEGORY MODEL OF THE MACHINE LEARNING

The input mathematical description of the system of functional diagnostics will be presented in the form of a structure

$$\Delta_B = \langle G, T, \Omega, Z, Y, X; f_1, f_2, \rangle,$$

where  $G$  is a set of input factors;  $T$  is a set of moments of the information interception time;  $\Omega$  is a space of diagnostic signs;  $Z$  is a space of possible technical states of the object of diagnostics;  $Y$  is a set of vectors-implementation of the recognition classes, which forms an input learning matrix;  $X$  is a binary training matrix;  $f_1$  is an operator of the formation of an input learning matrix  $Y$ ;  $f_2$  is an operator of the formation of a binary learning matrix  $X$ .

The categorical model of machine learning of the functional diagnostics system with a hierarchical data structure is presented in the form of a directed graph (Fig. 1). Thus, the edges of the graph are the operators, which represent the corresponding sets to the other, which are used in the machine learning process.

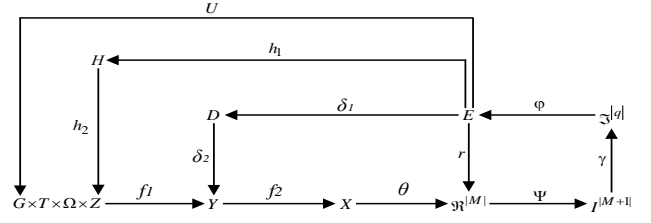


Figure 1. Categorical model of hierarchical learning of the functional diagnostics system

In Fig. 1 an operator  $\theta$  displays binary vectors-implementation of the matrix  $X$  to break up  $\mathfrak{R}^{|M|}$  space of diagnostic features on the recognition classes, and the operator  $\Psi$  checks the basic statistical hypothesis about the relevance of implementations to the corresponding recognition class. According to the results of the statistical testing of hypotheses a set of statistical hypotheses is formed  $I^{|S|}$ , where  $S$  is a number of statistical hypotheses, and an operator  $\gamma$  forms a set of exact characteristics  $\mathfrak{S}^{|q|}$ , where  $q = S^2$ . Operator  $\phi$  calculates the set  $E$  of the values of the information criterion for optimizing the parameters of the learning, and operator  $r$  at each step of the learning restores the classes of recognition containers in the radial basis of the signs space. In the categorical model, the contour of optimization of the control tolerances on the recognition signs is closed through the term-set  $D$  – a system of control tolerances of diagnostic signs. The contour, which includes a set  $H$  hierarchical structures, defines the alphabet of the recognition classes, for which the optimal breakdown of the signs is constructed. Operator  $u: E \rightarrow G \times T \times \Omega \times Z$  regulates the machine learning process.

### IV. MACHINE LEARNING ALGORITHM

According to the categorical model (Fig. 1), the information-extreme algorithm for learning the functional diagnostics system of an air clutch drive with a given hierarchical data structure is presented in a form of an integrational procedure of the search for global maximum averaged by the alphabet  $\{X_m^o\}$  classes of the information criterion (2) recognition in the work area of determining its function

$$\delta_{K,h,s}^* = \arg \max_{G_{\delta,h,s}} \{ \max_{G_E \cap G_d} \bar{E}_{h,s}(d) \}, \quad (3)$$

where  $G_{\delta,h,s}$  is valid values for the parameter  $\delta_{K,h,s}$  of the fields of control tolerances for diagnostic attributes of recognition classes of  $s$  stratum of tier  $h$  of the hierarchical structure. Internal procedure cycle (3).

implements the so-called basic algorithm of machine learning. The principal functions of the basic algorithm are the calculation at each step of the training criterion (4), the search for its global maximum and the definition of optimal geometric parameters of the classes of recognition containers. Thus, the optimal radii of the classes of recognition containers are determined by the procedure

$$d_{m,h,s}^* = \arg \max_{G_E \cap \bar{G}_d} \bar{E}_{h,s}(d_{h,s,m}), m = \overline{1, M_{h,s}}, \quad (4)$$

where  $M_{h,s}$  is a number of recognition classes, belonging to the  $s$  stratum of tier  $h$  of the hierarchical structure;

To optimize the machine learning parameters, consider the modified Kullback criterion, which has the form [5]

$$E_{h,s,m}^{(k)} = \{2 - [\alpha_{h,s,m}^{(k)}(d) + \beta_{h,s,m}^{(k)}(d)]\} \times \log_2 \left( \frac{2 - (\alpha_{h,s,m}^{(k)}(d) + \beta_{h,s,m}^{(k)}(d)) + 10^{-r}}{\alpha_{h,s,m}^{(k)}(d) + \beta_{h,s,m}^{(k)}(d) + 10^{-r}} \right), \quad (5)$$

where  $\alpha_{h,s,m}^{(k)}(d)$  is the first-kind error when making classification decisions, calculated in the restoration process of a hyperspherical container of a recognition class  $X_m^o$  with a radius  $d$ ;  $\beta_m^{(k)}(d)$  is a second-kind error;  $10^{-r}$  is a number small enough to be entered to avoid dividing by zero ( $1 < r \leq 3$ ).

A generalized scheme of hierarchical parallel optimization algorithm of control tolerances recognition under extreme training information system functional diagnostics is as follows:

- 1) reset of the tiers gauge of the data structure:  $h := 0$ ;
- 2) initialization the tiers gauge of the data structure:  $h := h + 1$ ;
- 3) reset of the strata gauge of the  $h$  tier of the hierarchical structure:  $s := 0$ ;
- 4) initialization of the tire strata gauge:  $s := s + 1$ ;
- 5) reset the gauge of the parameters control steps of the control tolerances field:  $\delta_{K,h,s} := 0$ ;
- 6) initialization of the gauge of the parameters control steps of the control tolerances field:  $\delta_{K,h,s} := \delta_{K,h,s} + 1$ ;
- 7) implementation of the basic machine learning algorithm, which calculates the maximum value of the information criterion  $\bar{E}_{h,s,m}(d_{h,s,m})$  and in accordance with the procedure (4) determines the optimal geometric parameters of the containers of the recognition classes;

8) if  $\delta < \delta_H / 2$ , then paragraph 6 is implemented, otherwise it is paragraph 9;

9) a maximum value of an averaged alphabet of the recognition classes of the information criterion  $\bar{E}_{h,s}^*$  is calculated; an optimal value of the parameter field of the control tolerances is determined

$$\delta_{h,s}^* = \arg \bar{E}_{h,s}^* ;$$

10) optimal lower  $A_{HK_i}^*$  and upper  $A_{BK_i}^*$  control tolerances for all diagnostic features are calculated:

$$A_{HK_i}^* = y_{h,s,m}^* - \delta_{h,s}^* ; \quad A_{BK_i}^* = y_{h,s,m}^* + \delta_{h,s}^* , \quad (6)$$

where  $y_{h,s,m}^*$  is an optimal nominal value of  $i$  sign of the vector-implementation  $X_{h,s,m}^o$  class;

11) an optimal binary training matrix  $\|x_{h,s,m,i}^{(j)*}\|$  is formed for the optimal system of control tolerances (6) according to the rule

$$x_{h,s,m,i}^{(j)*} = \begin{cases} 1, & \text{if } A_{HK,i}^* \leq y_{h,s,m,i}^{(j)*} \leq A_{BK,i}^* ; \\ 0, & \text{if else ;} \end{cases}$$

12) optimal averaging vectors-implementation  $\{x_{h,s,m}^*\}$  are formed to determine the geometric centers of the containers of the corresponding recognition classes, according to the rule

$$x_{h,s,m}^* = \begin{cases} 1, & \text{if } \frac{1}{n} \sum_{j=1}^n x_{h,s,m,i}^{(j)*} \leq \rho_{h,s,m} ; \\ 0, & \text{if else,} \end{cases} \quad 1$$

Where  $\rho_{h,s,m}$  is the level of selection (quantization) of the coordinates of averaged binary vectors of the recognition classes;

13) if  $s \leq S_h$ , then paragraph 4 is implemented; otherwise, it is paragraph 14 ;

14) Fields of the control tolerances (14) if  $h \leq H$ , paragraph 2 is implemented, otherwise it is paragraph 15;

15) Stop.

Thus, the algorithm of information-extreme machine learning is for adapting the input mathematical description of the system of functional diagnostics to its maximum informational capability.

## V. RESULTS OF MACHINE LEARNING

Implementation of the above-mentioned algorithm was carried out through the example of machine learning of the system of functional diagnostics of the units of an air clutch electric drive. A learning matrix was formed

according to the archival data provided by the ULIS Systems enterprise, which is engaged in the design of the air clutch control systems.

For illustrative purposes, the diagnostics was carried out for the four classes of recognition: class  $X_1^o$  described the functional state of the electric drive "Norma", class  $X_2^o$  - the state of elevated temperature of the first bearing of the electric motor, class  $X_3^o$  - the state of elevated temperature of the windings of the electric motor and class  $X_4^o$  - the state of the elevated temperature of the second bearing of the electric motor.

Hierarchical data structure is shown in Fig. 2.

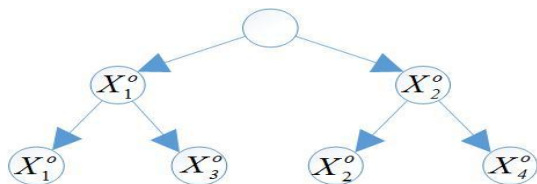


Figure 2. Hierarchical data structure

Fig. 3 shows the graph of the dependence of the information criterion (4) on the parameter  $\delta$  of the control tolerances field of the diagnostic features for the recognition classes of the first stratum of the second-level structure shown in Fig. 2

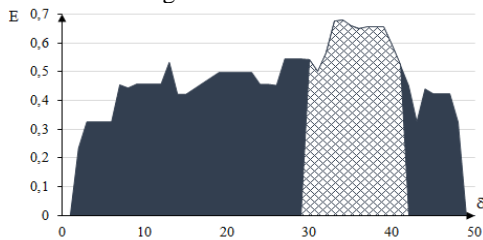


Figure 3. Graph of the information criterion dependence on the parameter of the control tolerances field on the diagnostic features

Analysis of Fig. 3 shows that the maximum value of the averaged criterion (4) in the working (shaded) field is equal to  $\bar{E}_{2,1}^* = 0,68$  by the optimal parameter of the control tolerances field  $\delta_{2,1}^* = \pm 34$  of the percent of the nominal value of the diagnostic sign. Thus, optimal radii of the classes of recognition containers  $X_{2,1,1}^o$   $X_{2,1,3}^o$  respectively equal to  $d_{2,1,1}^* = 28$  and  $d_{2,1,1,3}^* = 37$  (hereinafter in the code units).

Fig. 4 shows the graph of the dependence of the information criterion (4) on the parameter  $\delta$  of the control tolerances field of the diagnostic features for the recognition classes of second stratum of the second-level structure shown in Fig. 2

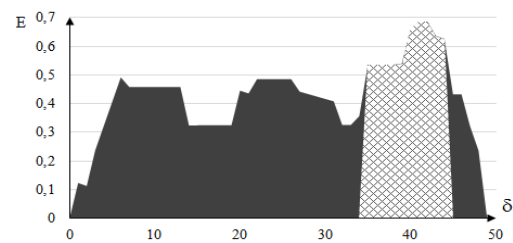


Figure 4. Graph of the information criterion dependence on the parameter of the control tolerances field on the diagnostic features

Analysis of Fig. 4 shows that the maximum value of the criterion (4) is equal to  $\bar{E}_{2,2}^* = 0,69$  by the optimal parameter of the control tolerances field  $\delta_{2,1}^* = \pm 42$ . Thus, optimal radii of the classes of recognition containers  $X_{2,2,2}^o$  and  $X_{2,2,4}^o$  respectively equal to  $d_{2,2,2}^* = 25$  and  $d_{2,1,1,3}^* = 39$ .

According to the results of physical modeling, applying the hierarchical decisive rules, the averaged total probability of correct classification was equal to  $P_i = 0,83$ , and applying the linear solving rules equaled  $P_i = 0,71$ .

## CONCLUSIONS

1. A proposed algorithm of machine learning with a hierarchical data structure allows increasing the asymptotic reliability of the functional diagnostics system of the air clutch electric drive in comparison with the linear solving rules.

2. To construct a non-error-based educational matrix of decisive rules, it is necessary to increase the depth of information-extreme machine learning by optimizing additional parameters of the functional diagnostics system.

## REFERENCES:

- [1] P. S. Bhowmik, S. Pradhan, and M. Prakash, "Fault diagnostic and monitoring methods of induction motor: a review," *Int. J. Appl. Control. Electr. Electron. Eng.*, vol. 1, no. 1, pp. 1-18, 2013.
- [2] O. Andrzej, S. Zygmunt, "Modern Methods Control and Diagnostics of Hoisting-Machines with Application of Artificial Intelligence Methods", *Archives of Mining Sciences*, vol. 1, pp. 217 - 231, 2010.
- [3] Z. Szymański, "Intelligent, Energy Saving Power Supply and Control System of Hoisting Mine Machine with Compact and Hybrid Drive System", *Archives of Mining Sciences*, vol. 60, no. 1, pp. 239 - 251, 2017. DOI:10.1515/amsc-2015-0016
- [4] S. Subbotin, (2013), "The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition", *Optical Memory and Neural Networks (Information Optics)*, vol. 22 no. 2, pp. 97 - 103, 2013. DOI: 10.3103/s1060992x13020082
- [5] A. S. Dovbysh, M.M. Budnyk, V.V. Moskalenko, "Information Extreme Algorithm for Optimizing Parameters of Hyperellipsoidal Containers of Recognition Classes", *Journal of automation and information sciences*, vol. 44 no. 10, pp. 35 - 44, 2012. DOI: 10.1615/JAutomatInfScien.



# Physical and Mathematical Modelling of Pitting Corrosion

Nutthawut Suchato<sup>1</sup>, Roger Light<sup>1</sup>, Richard Smith<sup>1</sup>, Alexander Kalashnikov<sup>2</sup>

<sup>1</sup>Nottingham University, United Kingdom

<sup>2</sup>Sheffield Hallam University, United Kingdom

<sup>2</sup>a.kalashnikov@shu.ac.uk, <http://tinyurl.com/lokvnbt>

**Abstract** – The study of corrosion (especially pitting corrosion) is complicated by the highly random nature and slow rates of the natural corrosion process. We present an experimental setup for enabling accelerated corrosion, procedures for its use to accelerate physical modelling of pitting corrosion, and results of associated mathematical modelling.

**Keywords** – accelerated corrosion, pitting corrosion, physical corrosion modelling, Monte-Carlo corrosion modelling.

## I. INTRODUCTION

Corrosion causes damage to various structures, machinery and goods. Its cost in the US alone was estimated at 276 BUSD in 2002 [1]. Corrosion is a highly random process; therefore, the development of corrosion monitoring instrumentation and procedures requires a good understanding of how the corroded parts change over time and respond to interrogation (e.g., ultrasonic signals). However, the observed in the UK average rates of natural corrosion exceed 0.1 mm/year only in some industrial atmospheres [2], making study of natural corrosion very time consuming. In this paper, we discuss how pitting corrosion can be modelled both physically and mathematically within short time frames.

## II. PHYSICAL MODELLING OF PITTING CORROSION

We developed a setup to accelerate natural corrosion by applying a constant electrical current, which eroded the cylindrical iron bar anode in a typical electroplating setup (fig.1) [3]. The sides of the bar were protected from corrosion by applying lacquer and heat shrink tubing (fig.2). Clear erosion of the anode and accumulation of extra iron on the cathode were observed after the 2.5 hour experiment (fig.3). The anode lost over 1% of its weight (fig.4) or over 1 mm of its length, confirming the suitability of the setup for studying accelerated physical modelling of uniform corrosion [3].

Pitting corrosion is a form of extremely localised corrosion that leads to the creation of small holes in the metal. We modelled it by covering the corroding surface with a protection layer. Corrosion was still able to erode the anode, but this time it progressed non-uniformly only through some weaker spots in the protection layer, even

though the protection layer was notionally uniform [4]. This experimentally observed result needed some further explanation.

## III. MATHEMATICAL MODELLING OF PITTING CORROSION

We modelled corrosion by representing the metal sample as a 2D body consisting of identical square cells. We assumed that if a cell is adjacent to a corroded cell, a probability exists that the former will also corrode. This probability was multiplied by the number of corroded cells that surrounded the metal cell in question. At every Monte-Carlo modelling step, uniformly distributed random numbers [0,1] were generated for all the cells. If the number for a given cell was lower than the total probability of corrosion, the cell was marked as a corroded one.

The modelling results, obtained using a custom MATLAB script, are presented in figs.5 and 6 for various corrosion scenarios (no protection layer and protection layer with defects) and two cell corrosion probabilities. The results showed that the erosion rate is roughly proportional to the corrosion probability and that imperfections of the protection layer caused pits.

## CONCLUSIONS

Accelerated physical modelling of pitting corrosion can be achieved by covering the corroding surface with a notionally uniform protective coating and forcing electrical current through it. Monte-Carlo mathematical modelling confirms validity of this approach.

## REFERENCES:

- [1] “Corrosion costs and preventive strategies in the United States, publication No FHWA-RD-01-156 (2002)”. [Online]. Available: <https://tinyurl.com/ycqfknjg>. [Accessed: April. 30, 2018].
- [2] G.O.Lloyd, “Atmospheric corrosion”, pp. 8. [Online]. Available: <https://tinyurl.com/yavogoce>. [Accessed: April. 30, 2018].
- [3] N.Suchato, S. Sharples and A.N.Kalashnikov, “Design and experimental evaluation of a test rig for ultrasonic monitoring of accelerated corrosion”, In Proceedings of International Conference on Applied Engineering (ICSAE) for Students, 2016, pp.147-152.
- [4] N.Suchato, A.N.Kalashnikov, R.Light and S.Sharples, “Experimental setup setup of continuous ultrasonic monitoring for corrosion assessment”, In Proceedings of Ultrasonics Symposium (IUS), 2017 IEEE International, 2017, p. 4.

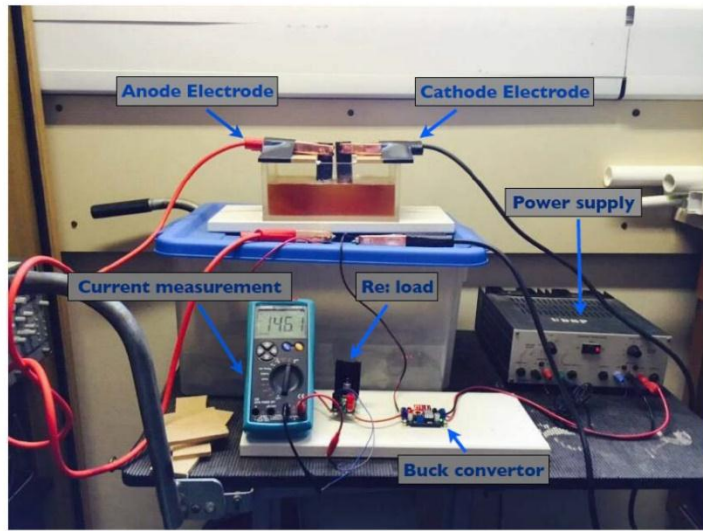
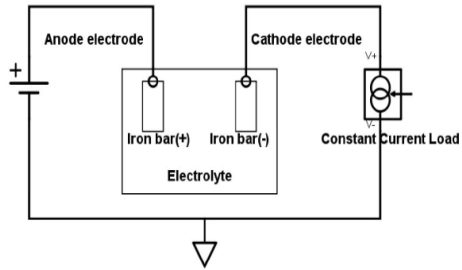


Figure 1. The experimental setup



Figure 2. Preparation of the corroding samples: original iron bar, bar coated with PCB lacquer, coated bar with set heat shrink tubing

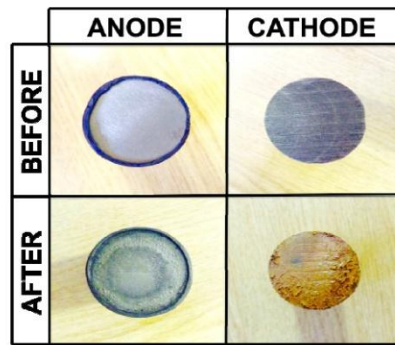


Figure 3. Surfaces of the electrodes before and after the experiment

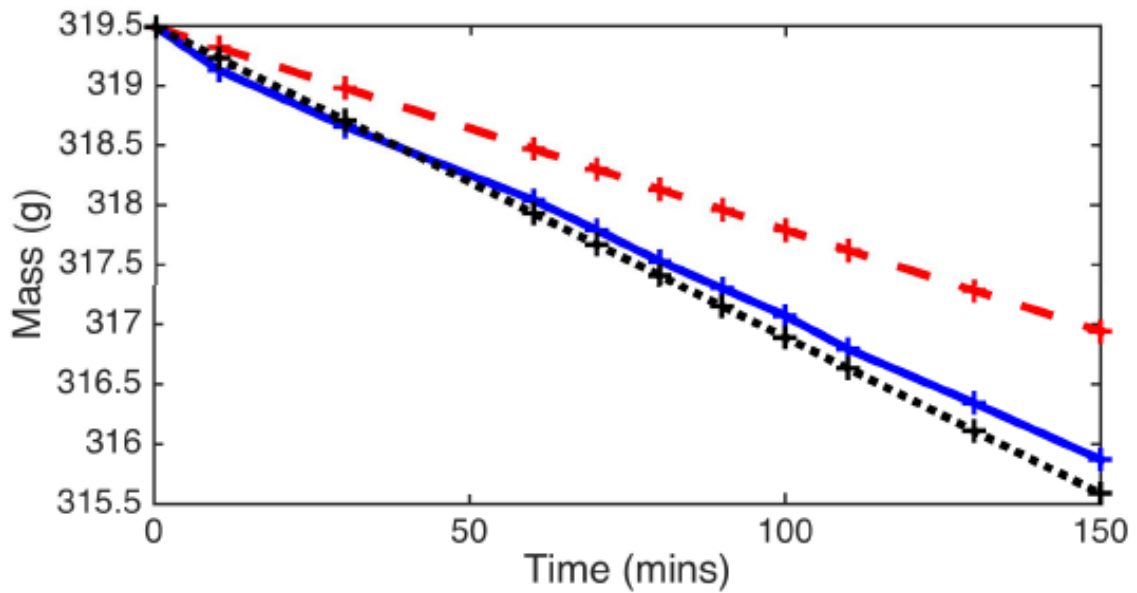


Figure 4. Measured (solid line) and expected (dotted lines for Fe<sup>2+</sup> and dashed lines for Fe<sup>3+</sup>) weight of the corroded sample versus the accumulated corrosion time

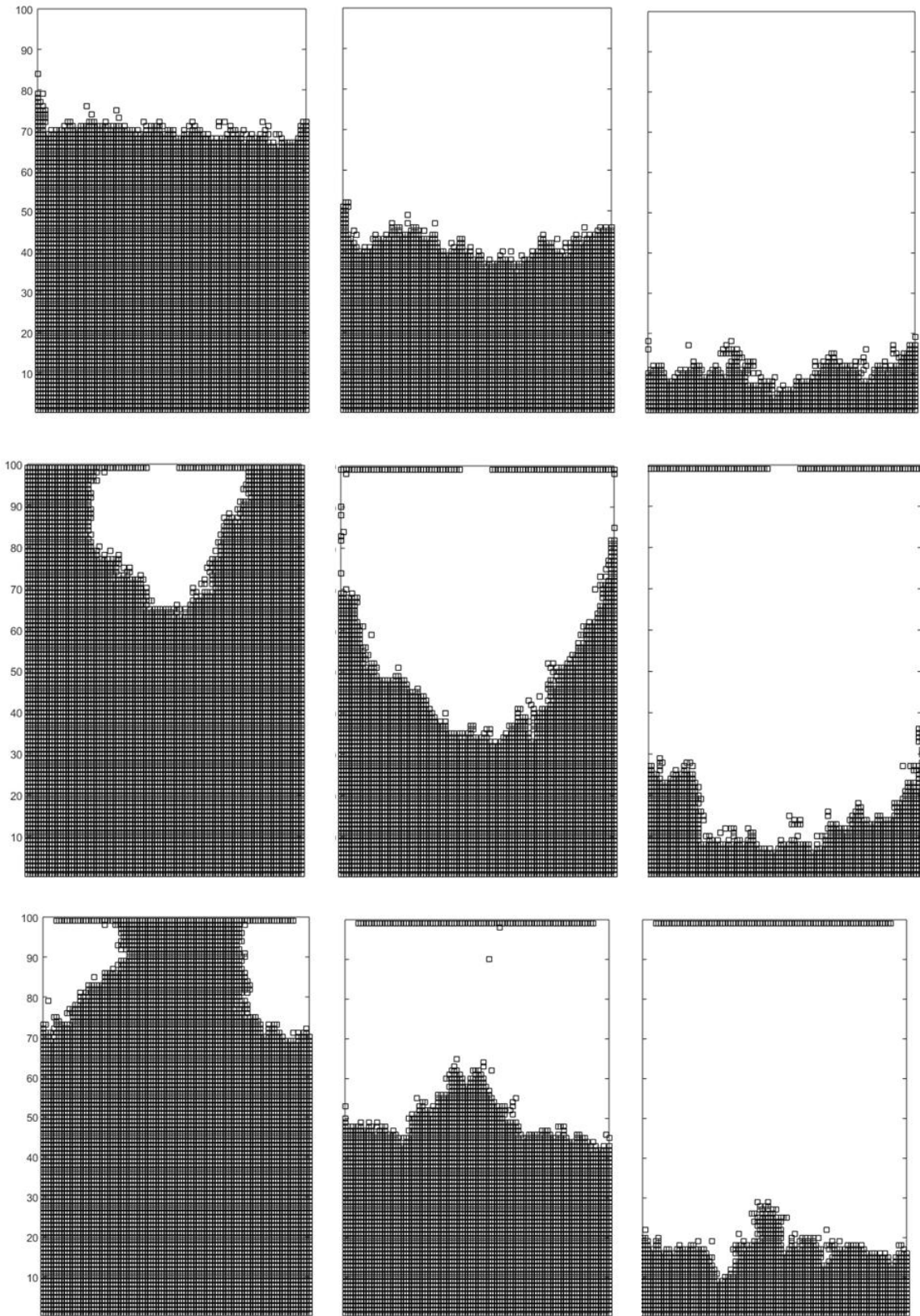


Figure 5. Mathematical modelling of corrosion for single neighbour corrosion probability of 0.04 (upper graphs - uniform corrosion of an untreated sample, centre graphs - corrosion via a small scratch in the protection layer, lower graphs - corrosion via the sides of the protection layer; left, centre and right graphs - corrosion after 100, 200 and 300 simulation steps, respectively)

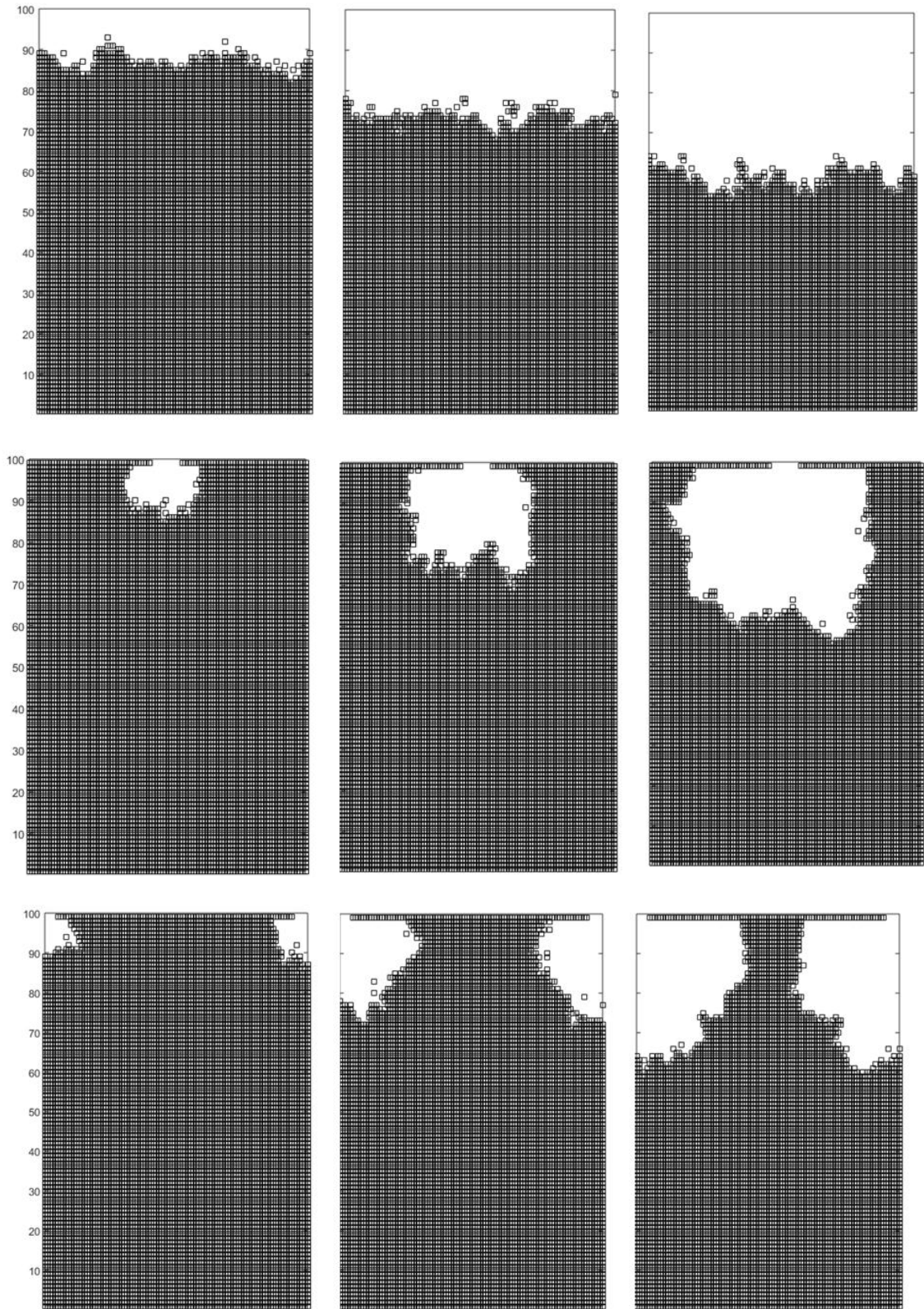


Figure 6. Mathematical modelling of corrosion for single neighbour corrosion probability of 0.02 (upper graphs - uniform corrosion of an untreated sample, centre graphs - corrosion via a small scratch in the protection layer, lower graphs - corrosion via the sides of the protection layer; left, centre and right graphs - corrosion after 100, 200 and 300 simulation steps, respectively)

# Secure Mobile Application Development

Roman Yatsenko, Viktor Obodiak<sup>1</sup>, Valerii Yatsenko  
Sumy State University, Ukraine, <sup>1</sup>v.obodyak@cs.sumdu.edu.ua

**Abstract – The article deals with a problem of mobile applications vulnerability. It explores the most popular attack types and ways to avoid them. There were identified the main security rules of code writing for mobile developers. Recommendations how to protect a mobile application from cracking have been given.**

**Keywords – mobile application; vulnerabilities; attack types; security rules.**

## I. INTRODUCTION

When developing a mobile application, it should be taken into account that the data that this application operates on may be of some interest to third parties. The degree of value of these data varies widely, however, even the most simple private information, for example, the password for entering the application, requires the elaboration of its protection. This is especially important in the light of the spread of mobile applications to all areas of electronic services, including financial, banking, storage and transfer of personal data.

## II. FORMULATION OF THE PROBLEM

Mobile technologies are developing very quickly and at the same time the number of vulnerabilities and methods of hacking is growing. Every year, Apple announces a new iPhone, and Google is pleased users with a new version of Android, but with no less frequency, you can see articles about a new virus or a leak of private user information. Therefore, developers should constantly monitor both new versions of products and the appearance of new types of attacks.

In general, there are three main types of attacks on a mobile application:

1) Decompiling the application file (.ipa files for Apple iOS and .apk files for Google Android) and parsing the locally stored data. These files are archives in which executable files, configuration files, application resources, etc. are stored. If you unpack them and analyze the configuration files, you can often find lines of code that developers forgot to remove from the final version of the product. These lines of code are most often used for debugging during the development period of the application, and they can greatly facilitate the attacker's task of obtaining confidential data or implementing other unauthorized actions. At the testing stage, developers often incorrectly assign access rights and forget to edit them at the final release of the software product, which is

why the attackers have even more opportunities for unauthorized access. Despite attempts by developers of mobile operating systems, unfortunately, getting an application file is quite easy for an attacker and by the fact that the protection of this, the most important at present, level entirely lies on the shoulders of the mobile developer.

2) Interception of data transmitted over the network (MITM-attacks). Most mobile applications are client-server, therefore, constantly transmit and receive large amounts of information. And although modern mobile and web development actively complete the transition to the HTTPS-protocol of communication, nevertheless, you should not rely on a single line of protection in the form of a secure communication channel. The way to store important information on the server is still the safest, but choosing the wrong method or a sequence of data transfer can be a fatal error.

3) Rooting of the device and attack on the application and the algorithms used in it through external debugging tools. An attacker will always be able to get to the very depths of the program code and learn the method used to store, transfer and encrypt data.

The security requirements for mobile applications are not the same as for web applications. It is assumed that the user can work offline so you can not do without local storage of information on the device. For example, for online authorization with the storage of session data and cookies. After the identification data (login and password) have been entered and the application has authorized the user, it stores a special identifier, which is then presented to the server with each request coming from the application. If an attacker received a user ID and the system did not perform the procedures for checking the session IP address or having more than one connection within the session, the attacker could gain access to the system with the rights of the user's account. If these are applications related to Internet banking or to the personal cabinet of the payment system, then it is not difficult to guess the consequences of unauthorized access. Another problem is the inadequate control of client applications in some markets. This is the process of verifying software uploaded to Application stores. Before you go to the App Store, iOS-applications are checked in detail for vulnerabilities and compliance with Apple development standards. Each application installed on iOS must be signed with a special certificate "iOS Developer Program", issued by Apple only after a number of

necessary checks. Such security measures ensure the absence of malicious software in the App Store. While in the Android operating system, applications are not scanned for malicious code before they are uploaded to the Google Play platform. Instead of the preliminary verification procedure, Google has implemented a mechanism for automatically scanning the application store for potentially harmful software. As practice shows, this method of analyzing information security increases the percentage of penetration of malicious applications and their further distribution to end users' devices.

### III. CONSIDERATION OF VULNERABILITIES AND METHODS OF THEIR PREVENTION

Any vulnerability of the mobile application is based on shortcomings in the technology of operating with critical user data. Critical user data includes any data that should not be available to a third party, it concerns the user's personal data (date of birth, address of residence, personal correspondence) and his private data (passwords, credit card details, bank account numbers, order numbers and so on). Each time a mobile application is hacked is a race between an attacker and a developer in using technology and responding in a timely manner to threats and bugs. It is the developer and the technologies that he uses that determine the security of the application and the safety of user data.

In accordance with the classification of the Open Web Application Security Project (OWASP) [1], the main vulnerabilities of mobile devices are: systemic vulnerabilities (architectural solutions of mobile platforms); unsafe storage of data; insufficient security of data transfer protocols; the vulnerability of the authorization and authentication system; weak cryptostability; the vulnerability of the application code; hidden application functionality; improper control over client applications.

The most common mistake is the use of unprotected local storages. This vulnerability occurs very often, it is expressed in the storage of critical user data in unprotected or weakly protected local storages for a specific platform. Even if the application developer sets the level of access to the storage as private, prohibiting its use by other applications, the local storage can still be read from outside – by rooting the device or connecting it to a PC. In this case, the presence of special skills for the attacker is almost not required.

The second, by prevalence, vulnerability can be considered the storage of critical user data in the application code: in static constant strings, in application resources, etc. As examples: storing password salt in a constant or macro, which is used throughout the code to encrypt passwords; storage of a private key for asymmetric algorithms; storage of passwords and logins for server nodes or databases. Particularly it is relevant for applications written for the Android platform, but iOS applications are also subject to this problem. Such a

vulnerability is easily revealed by a third party with basic code decompilation skills. It is important to take into account that the popular method of protecting the source code from decompilation – obfuscation, in no way protects applications from vulnerabilities, or can only partially slow down the search process [4].

In addition to storage, it is also dangerous to transfer critical data to the external environment in an open manner. This vulnerability is expressed in the transmission of data without the use of encryption on any available communication channel with the external environment, whether it is data transfer to a third-party application or transmission to the network. It can be opened indirectly by opening not the application, but its storage, or another application receiving this data. In this case, hacking success is demanding of the skills of the attacker, but only if the storage is protected. Any critical data before going outside the application must be encrypted. Local platform storages are not an application area, they must also receive only encrypted data, otherwise the security of the data can be compromised.

Using encryption of information is still not secure, there can be an application of algorithms with the storage of a private key. Vulnerability is relevant in the case when private information of the algorithm – private key, is stored in the code or resources of the mobile application, which often happens. Even despite the use of encryption in this case for an attacker, there is still a way to find a private key by decompiling, as in the previous vulnerability, and deciphering the algorithm. In mobile development, it is desirable to use only modern symmetric algorithms with a generated random one-time key that have high resistance to brute-force cracking, or to output an asymmetric private key outside the application, or to personalize this key, for example – a private key can be a user input code stored and encrypted in a protected operating system store. To implement this protection, there are many official libraries from Google that easily integrate into applications for both Android and iOS.

A similar vulnerability is the use of an asymmetric algorithm with a private key, known to the server. In this case, the danger may exist on the server side. This vulnerability is of a dual nature. Storage of a private key allows for decryption of user data on the server side. First, it is incorrect from the security point of view, if the server is hacked – the attacker will also have access to private user data, and secondly, it violates the privacy of personal data. The user should always be sure that his personal information is not known to anyone but himself, unless he explicitly gave permission to publish it. Often applications position themselves as protected, but in fact they are not, as they contain the means for deciphering personal information inside themselves. Without the explicit need and explicit permission of the user, most often through a license agreement, neither the application nor the server should have any opportunity to decrypt the

user's private data. The simplest example is that the user's password must go to the server already in the form of a hash, and the hash should be checked, and not the original password, the server absolutely does not need to know the user's password; if the user forgot it, then for such a situation there is a long-established mechanism for recovering the password, including with two-factor authorization of the client for increased security of the recovery procedure [4]. Technically, as in the previous example, you can use the official libraries from Google that are designed to send and receive information from the server securely. Moreover, if you do not need to additionally process data before storing it, you can use ready-made server from Google such as Firebase Realtime Database or Firebase Cloud Firestore [3].

It should also avoid using self-written encryption and protection algorithms. Since this is a direct violation of the Kerckhoffs principle [5]. Expressed in the developer's attempt to invent his personal, not known to anyone, and therefore a super-secure encryption algorithm. Any deviation from existing, repeatedly verified and studied, mathematically proven encryption algorithms in 99% of cases turns into a rapid hacking of such protection. Of course, hacking such an algorithm requires the medium-high skills of the attacker, but does not guarantee complete security. For such solutions, it is necessary to select a suitable algorithm only from well-established and current well-known cryptographic algorithms. In addition, it is also necessary to pay attention to the relevance and timeliness of the chosen encryption algorithm. As the computing power grows exponentially, so the well-functioning algorithms of the last decade can already be susceptible to breaking through brute force on modern high-speed video cards following the progress.

A common mistake of developers is simply ignoring the fact of the presence of rooted or infected devices. Rooted devices are devices where modifications have been made to obtain superuser rights for any operations originally prohibited by the operating system manufacturer. It is performed by the user on his device independently, and not necessarily voluntarily, when the client may not be aware that the device was hacked. Installing the application on a handheld device removes all standard protection for the operating system. Especially this problem is relevant for users of the Android operating system. There is a huge number of special programs called "patchers" that in runtime can substitute data in the application or unlock the limited functions. Such operations are possible only on rooted devices and do not require special skills and technical knowledge from the attacker. In such cases, if it is technically possible for the platform, it is desirable to limit the application if it was recognised that the launch is made on rooted device, or at least warn the user of possible danger and leave the decision of the application start at the discretion of the user.

It is also not acceptable to store critical data in protected storage, but in an open form. Developers are often inclined to store data in protected system stores without additional protection, because system mechanisms are well resilient to hacking. However, their level of durability drops to a minimum in case the device is rooted. Such data should not be used in the application without additional encryption. As soon as the need for open critical data has disappeared – they must immediately be either encrypted or destroyed. Using such a vulnerability also requires a cracker to have high skills, but still retains risks.

There is also not a deliberate transfer of part of the functionality to the built-in web engines. Most often it looks like transferring critical data to the built-in browser, where an external web page is loaded, performing its part of the functionality. The level of protection in this case is sharply reduced, especially for rooting devices, even if a browser is used for a well-known authoritative company, because in this case it does not matter. So, you should not use the built-in browser and built-in web engine in operations with important data. In extreme cases, encrypt them before transmission.

In conclusion, it is worth mentioning the vulnerability associated with reverse engineering algorithms that have an intellectual value. If the development of the application within the company uses some of its own algorithms that can be of high value to potential competitors or hackers, then these algorithms must be protected from unauthorized access. In this case, you can protect yourself from this theft with the help of a banal automatic, or even, depending on the scale of the company, manual obfuscation of the code. In the obfuscated form, the algorithm may no longer represent the primary value for the attacker, or simply dissolve into the sea of obfuscated code.

#### IV. BASIC RECOMMENDATIONS FOR THE PROTECTION OF MOBILE APPLICATION

There are several common points for all mobile platforms that you should follow when developing. Often, a personal password is used to protect personal data: PIN code, fingerprint scan, graphic password, etc., in this case it is important when the application goes to the background or when folding it is necessary to immediately display the input window for this security code, overlapping the entire screen of the application. This eliminates the possibility for an attacker to obtain private information in the event of device theft, while the application is still running and is in sleep mode. Any user code should have a limited number of attempts to enter, for example 5 times, then, in case of failure, the application should automatically be logged off, or completely blocked, depending on the particular application [4]. At present, when using digital codes, it is strongly recommended to use a code length restriction of at least 6 digits [2].

In the case of client-server applications, it is very useful to use a session mechanism with a limited session lifetime. This will prevent the application from idle in an unprotected mode, if the user simply forgot to close it and left the device in the public domain. It should be borne in mind that the duration of the session and its identifier are relevant to the user's critical data, with all the ensuing consequences. One of the successful examples of implementing such a mechanism is getting the absolute value of time from the server after passing the user authorization procedure, in which case the date and time should indicate when the session will become inactive. It is important to remember that the session end date and time should not be generated on the device, this reduces the security and flexibility of the application. In addition, the client-server application should not change the critical data in the local mode. Any action that requires changing important data must be synchronized with the server. The exception to this rule is only the user-defined login code that is set by the user personally and stored in a secure local store.

An important aspect is the competent operation of dates. When you run an application with dates, such as the time to destroy a session, you should not rely on relative time. So, the data transmitted from the server should not contain the date in the form of N seconds / hours / days from the current moment. Due to potentially high delays in transferring data over the network from the mobile application to the server and back, such a synchronization method will have too much error. In addition, an attacker, or simply an unscrupulous user, can simply change the time zone on the device, thus violating the logic of the application's limiting mechanisms. It is always needed to transmit only the absolute value of time. Absolute values should be transmitted using universal ways of exchanging such information, without reference to the time zone of a particular user device. Most often, the best option is the behavior of the application, in which the data is displayed to the user in its local time zone, but their storage and transmission is done in a format not tied to the time zone. Suitable formats for dates and times are either a universal UNIX timestamp stored in a variable of a 64-bit integer sign type, or, in extreme cases, a string in the full ISO-8601 format with a zero time zone. The UNIX timestamp is preferred, it avoids potential errors and problems with converting strings to date and back on different mobile platforms [3].

If we talk about general security principles, then the application should not display private user information in large, bright, well-readable fonts, without explicit need and without a separate user request, to exclude the possibility of reading this data from the device's screen.

Do not blindly trust open source libraries that offer some kind of protection for private user data. The only exception is libraries that are time-tested and used in large corporate projects, for example, built-in encryption

in the open engine of the Realm database. The standard mechanisms of protection of the operating system and publicly available cryptographic algorithms in the overwhelming majority of cases will be more than enough.

It is also better not to use cryptographic libraries with closed source code, even if they are paid. In such solutions, you can not in any way check whether the library is effective, or how honest it is protecting, whether there is a backdoor mechanism, or whether your data is being sent to some third party. For the release builds of applications, data logging to the system console and unprotected files should be disabled. Logs are log files containing records of all events occurring in the mobile operating system, with a high level of detail. On the device, any application during installation can request access rights to read the logs. Many users do not pay attention to this request, but the danger is that any installed application that requested access to read logs, and at the same time received approval from the user, will get the right to read all the information that the application puts into the logs, if logging not turned off by the user.

## CONCLUSIONS

For each separately created mobile application, the number of applied protection levels will vary greatly. For example, if the application is not at all client-server, does not contain any critical data, and does not operate with valuable internal algorithms, then it makes no sense to attach any protection to it. If the application is oriented, for example, to performing banking operations, or storing user passwords, then the level of its security should be the highest. However, the previously mentioned common vulnerabilities of the mobile sector can easily be excluded from the application, most often it does not add any additional costs if the applying of the required level of protection was started at the early stages of application development. But the implementation of post-factum protection in an already running application may well be associated with significant expenses of the developers' efforts and time. Therefore, the choice and coordination of the level of security, as well as a list of critical data in the application being developed, should be performed at the earliest stages of design.

## REFERENCES:

- [1] K. Anton, J. Bird and J. Manico, *Top-10 Proactive Controls 2016*, The OWASP Foundation, 2016. [E-book] Available: OWASP e-book.
- [2] ENISA, *Smartphone Development Guidelines*, European Union Agency for Network and Information Security 2017. [E-book] Available: ENISA e-book.
- [3] Firebase, "Build apps fast, without managing infrastructure", *Firebase*. [Online]. Available: <https://firebase.google.com> [Accessed: Apr. 10, 2018].
- [4] NowSecure, *Secure Mobile Development Best Practices*, NowSecure Inc., 2016. [E-book] Available: NowSecure e-book.
- [5] Victor de Casto, *Just cryptography*. St. Petersburg: Strata, 2014.



# Web Service for Monitoring the Prices of Online Stores

Borys Kuzikov<sup>1</sup>, Maksym Vynogradov<sup>2</sup>

Sumy State University, Ukraine, <sup>1</sup>b.kuzikov@dl.sumdu.edu.ua, <sup>2</sup>vinogradov.max97@gmail.com

**Abstract** – Research of modern concepts and methodologies for data collection from web resources in the Web-mining plane. Creating a flexible and highly scalable service for monitoring the prices of Ukrainian online stores, using the latest methods and tools for web crawling. Primary data analysis a collection of the relevant data set and its visualisation for further study of the collected data.

**Keywords** – Data-mining, Web-mining, Scraping, Crawling, Data Analysis, Datasets.

## I. INTRODUCTION

During the last decade, information and telecommunication technologies are developing at an enormous rate, and together with them, various fields of human activity, ranging from scientific research to the sphere of services, are improved and integrated with the global Internet network. The vast majority of different companies, companies, stores have created their online resources representing their products, and even more – they integrate with social networks, banking systems, which allows them to sell goods online. In other words, we can choose a product, learn its price, features, reviews of his other customers and bye without leaving home. On the other hand, we have a lot of risks, namely:

- inaccurate specifications, or even lack of information about certain technical aspects of the product;
- the inability to see and check the real product before its purchase and receipt;
- overstatement of prices compared to the cost of goods or other online stores.

However, services of this type continue to gain in popularity, although Internet shop users are not in any way insured against the risks mentioned above.

The purpose of this work is to develop a flexible and extensible web service for monitoring the prices of online stores, which includes the means of collecting and analysing information from these Internet resources, as well as visualising the obtained data. This service will allow comparing prices for various products from various internet-shops and keep statistics of changes in prices for its further processing.

## II. THEORETICAL OVERVIEW

The main source of application data is, of course, websites, online stores. It based on their data to provide the user with the necessary information. There are two similar but fundamentally different in their ideology and implementation, approaches to data analysis: web scraping and web crawling.

**Web crawling** system or search robot is a program created for the sequential defined by some set of rules, bypassing web pages to index the data contained on them.

**Web scraping** is a program that is designed to not only crawl and index data from web pages; it mimics human behaviour on the Internet. Unlike crawlers, scrappers can simulate human behaviour through site registration by executing JavaScript or ActionScript scripts, as well as changing their logic of interaction with the website depending on the data already obtained.

**Parsing** is a process of analysing the input character set according to a given formal grammar to analyse its structure for further use of the obtained data. Parsing is an integral part of both scraping and crash, and is used to select the necessary information from a site – it allows to analyse the structure of an HTML document for further work with these data.

The methods of scraping and crunching in our case are part of the **Data Mining** process. In other words, the data extraction mechanism, also known as in-depth data analysis, used to detect hidden patterns in large data sets and allows us to classify these data as well as carry out on them based simulation and forecasting procedure. Classification, modelling and forecasting are the main goals of Data Mining Processes.

If specified, then this task is solved in the plane of **Web Mining** – is the use of methods of data mining intelligence to automatically detect various resources, documents, services on the Internet and the allocation of information from them to identify patterns in the global information network.

In Web Mining are the following processes:

- the input stage, extracting the original links;
- initial data processing stage;
- stage of information modelling;

- stage of the model analysis – the interpretation of the results.

What required in the mechanisms of Web Mining?

Formally, we can divide the needs into two groups:

- data aggregation and their convenient display for the end user – in the context of the system described in this paper – is the display of the list of goods and their graph of price changes in a user-friendly form;
- data set creation for further analysis by statistical methods or use as the basic data for methods of machine learning.

### III. CONCEPTS AND METHODOLOGY OF WEB-SCRAPER WORK

The main purpose of web scrapers is to gather information. In most cases, this process has two steps: first, downloading the necessary information from the Internet, and secondly, analysing and grouping the data.

There is a sufficiently large number of ways to download the necessary information from web resources. The feasibility of using one or another method directly depends on the size of the target information, in other words, the amount and volume of web pages that need to analyse. The most common of them:

- Manual download using a browser is a suitable option with a minimum amount of data or single use.
- Libraries and implementations such as wget, curl, requests – but they work synchronously, meaning that their use in pure form makes it impossible to send multiple requests at one time.
- Launching utilities or libraries from the previous item at the same time in several threads or processes that solves the problem of simultaneous sending requests.
- The event loop is a system that, although it works in single-threaded mode, but can perform I/O operations in parallel, that is, it downloads pages at a time when the program receives a response from the server execute a callback function. This approach is implemented in one of the most popular scraping frameworks – Scrapy.

The next stage of scrappers is an analysis of the downloaded page. At this stage, we can highlight the following approaches for processing received HTML-pages:

- processing pages as text using, for example, regular expressions or other templates;
- the visual approach is a high-level website analysis method that even ordinary users can use. The main disadvantages are the impossibility to change the algorithm of site analysis, and the price of the service is mainly commercial products;

- handling pages as a DOM (Document Object Model) tree.

The most popular and most effective option is the last one since most pages use HTML markup. It in turn also has several subsets, which differ in the method of analysis of the DOM tree, namely the use of:

- CSS selectors;
- XPath – is more complex than CSS, but more versatile;
- specialized DSL (domain-specific language) for HTML parsing. From the benefits: the ability to use the means of one or another programming language, better support in the IDE;
- it is necessary to highlight separately the use of the ready-made APIs that some sites provide to users to be able to obtain data from a site in a form convenient for processing, such as JSON.

### IV. THE ARCHITECTURE OF THE SOFTWARE SYSTEM

Web service for monitoring the prices of online stores – a system that has a multicomponent structure, each component of which is responsible for certain functionality in the system. The very creation and consolidation of the following components enabled the implementation of a flexible, extensible web-based price monitoring service:

- Scheduler – is one of the simplest components of the system, which is the script to start spiders every 24 hours;
- Scraping Engines – is a set of tools for obtaining, analysing and storing data from online stores;
- MongoDB – is a set of MongoDB collections to store the necessary data;
- Express Application Server – a server that connects the user interface and the database;
- React Single Page Web Application – is a tool for displaying Scraping Engines data from websites in a user-friendly manner.

Figure 1 shows the main component of the system:

1. Scheduler – is a Python script that runs in the background, every 24 hours, runs all existing web scrapers.
2. Scraping Engines – scrapers that collect data after work and data retrieval sent it to the MongoDB collection. If the product exists, then they add a new price to the list, if not the first, create a record (document) of the goods.
3. The Express Web Server sends queries to the MongoDB database to obtain the necessary product search or visualisation of their data prices.
4. React Single Page Web Application – is the login point for system users, which is a website where

the customer can select a category, find a specific product, and see the dynamics of price changes for the target products.

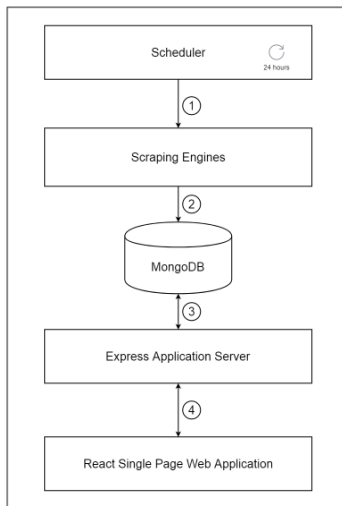


Figure 1. The architecture of the software system

Figure 2 shows relations between categories in the database. It stores the following collections (docs are described in JSON format, which is traditional for MongoDB).

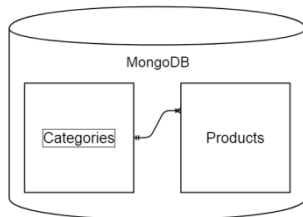


Figure 2. Database collections

In this type of database, there is no such thing as "relations" between collections, but the one-to-many logical link between categories and products can be distinguished since this or that category includes many goods. Products schema describe products and prices. Categories schema – the scheme with the name of the categories of goods. Scraping Engines – this is a component of the software system, which is a set of tools for receiving, initial analysis, processing and storing information from websites.

The component (see figure 3) works as follows:

1. Scrapers read all necessary for start data, from the configuration files. The number of scrapers is equal to the number of stores, information from which to collect information, namely: comfy.ua, rozetka.com.ua, moyo.ua, foxtrot.com.ua. Scrapers operate on the principle of single responsibility. Namely, one scraper works with only one online store, which allows us to add new scrapers or to remove existing ones without changing the functionality of all others.

2. After performing their work, the scrapers send data to the ProductPipeline – Python class that has the process\_item method that records the scrapper data to the intermediate repository – the items list.
3. Another method – spider\_closed, triggers after the scrapper is shut down, and stores the data (items from the items list) into the MongoDB database, namely the Products collection.

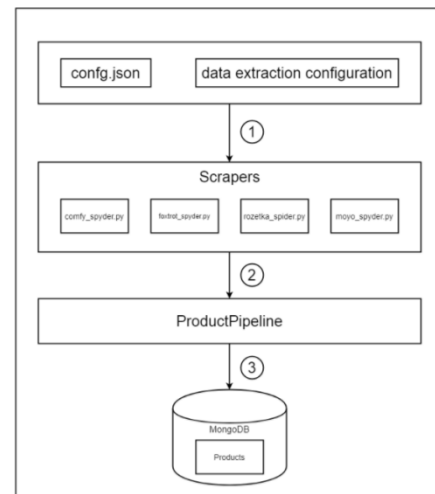


Figure 3. The architecture of «Scraping Engines.»

Regarding configuration files, – config.json has information about other system components, such as the MongoDB entry point (host, port, database). It is useful in a case when we need to change saving location, or storage type.

All scrapers inherit two basic classes: CrawlSpider and Spider Tools.

The CrawlSpider class from the scrapy.spiders package implements the logic of crawling the pages of the site. His key fields:

- name – is the name of the crawler, using which we can run it;
- start\_urls – list of initial pages for analysing;
- allowed\_domains – a list of allowed domains, in which the crawler can analyse the pages;
- rules – is a set of rules according to which the scrapper will look for each subsequent link, that is, the next page for further analysis. There are quite a lot of rules, and customer can also create his own rules for site bypass, but for the current task it was enough restrict\_xpaths rules, that is, the links are through XPath.

A SpiderTools class has been created and has the methods needed for each scrapper, namely the loading and analysis of configuration files, creating rules for page navigation, extracting product information on a page using XPath rules, and creating these products.

«Data extraction configuration» – is a set of configuration files, each of which corresponds to a certain online store, containing information on how to analyse information from the pages of online stores.

The main feature and advantage of this component is the ability to add a new scraper for a new online store easily. It is enough to create a new configuration file and a template class-scraper, with a minimum number of settings, and from the next iteration, we will start receiving data from the new site.

The **Express Web Server** component is an intermediate link between the data received from the Web stores and the Single Page Web Application that receives requests from the browser, processes them, and sends a response to the client.

The server is based on REST architecture principles and has a simple interface for managing information without the use of any additional inner layers. A global identifier, such as a URL, uniquely determines each unit of information, in other words, a model. Each URL in its turn has a strictly specified format. A component of the code, that according to the request URL determines which operation and what model to execute, is called a router.

The web server has models equivalent to the MongoDB schemas described above. Operations over these models are performed using typical input points: write, delete, modify, and read.

React Single Page Web Application – this component is responsible for the user interface. It is with this that the end user of the system can:

1. Find the necessary items on the system.
2. Go to the page of the necessary product of one or another shop.
3. View the graph with the visualisation of the dynamics of changes in prices for goods and compare them.

The interface of the application is shown in figure 4.

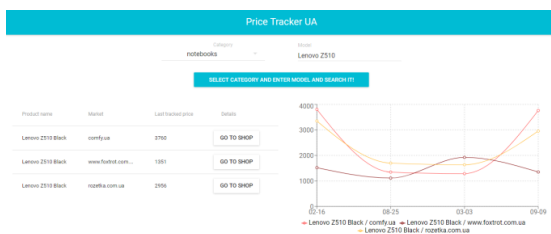


Figure 4. UI of web service

*React.js* – is the foundation and the bulk of the Single Page Application was created with its use. The *Recarchts.js* library was used to construct the graphs, with the help of which the *LineChart* graph was created which automatically rebuilds the graph with repeated requests to the server (as a consequence, the change in the number of goods).

The *axios* – is a library for sending AJAX requests to the server, needed to get information about categories, products of the dynamics of price changes from the server without reloading the page

## CONCLUSIONS

A web-based service for monitoring the prices of Ukrainian online stores has been designed and developed; web scrapers have been implemented, collecting and analysing data on online stores, creating a REST server for access to collected data, and adaptive, user-friendly, reflecting the dynamics of price changes for Internet product stores.

One of the components of the service created is a set of web scrapers that store and perform initial processing of information from the following stores: comfy.ua, rozetka.com.ua, moyo.ua, foxtrot.com.ua. In each of the stores analysed products of the following categories: laptops, digital cameras and cameras, smartphones and mobile phones. With the current configuration, for one iteration of work, scraper allows getting information about more than 5000 goods from all online stores.

The application allows the customer to find products of a certain category by their name in several online stores, displays the found products in the table, and also shows the dynamics of price changes on the graph. The graph allows to interactively find out the price of a product in a particular online store for a certain period.

The result of the work can be applied both in everyday life, to find out which online shop it is more profitable to buy goods, and for further scientific research. For example, to predict the price of a particular product in the future, or to find and analyse a correlation relationship between the prices of goods and between prices in general and other macroeconomic indicators in the country.

## REFERENCES:

- [1] D. Myers and J.W. McGuffee. 2015. Choosing Scrapy. *Journal of Computing Sciences in Colleges*, vol.31, no.1, pp.83-89, 2015
- [2] K. Matsudaira, Capturing and structuring data mined from the web, *Communications of the ACM*, vol.57 no. 3, March 2014. doi:10.1145/2567664
- [3] A.-A. Fatemeh and Ali Selamat, “An architecture for a focused trend parallel Web crawler with the application of clickstream analysis”, *Information Sciences*, vol. 184., no.1, pp. 266-281, 2012.
- [4] J. M. Hsieh, S. D. Gribble, and H. M. Levy, “The Architecture and Implementation of an Extensible Web Crawler”. In NSDI, 2010.
- [5] T. Srivastava, P. Desikan, V. Kumar, “Web Mining – Concepts, Applications and Research Directions”. In *Foundations and Advances in Data Mining. Studies in Fuzziness and Soft Computing*, W. Chu, T. Young Lin, Eds. Berlin: Springer, Heidelberg, 2004, pp. 275-307.
- [6] H. Chen and M. Chau, “Web Mining: Machine Learning for Web Applications”, *Annual Review of Information Science and Technology*, vol.38, pp.289-329, 2004.
- [7] P. Boldi, B. Codenotti, M. Santini, and S. Vigna, “Ubicrawler: A scalable fully distributed web crawler”, *Australian World Wide Web Conference*, vol. 34, no. 8, pp. 711-726, 2002. [Online] Available: citeseer.ist.psu.edu/article/boldi03ubicrawler.html [Accessed Aprl. 23, 2018]

**SESSION 5**  
**INTELLIGENT DATA PROCESSIN**

# Features of Calculating Ratings in Information Systems

Olga Pronina<sup>1</sup>, Elena Piatykop

State Higher Educational Institution "Priazov State Technical University", Ukraine, <sup>1</sup>pronina.lelka@gmail.com

**Abstract** – The research presents the method of calculating the driver's rating in the system. The key parameters which affect equivalently on the overall rating are the number of votes and mark. The formulas according to the key parameters are reduced to a single scale that to use them in the general formula for calculating the rating. The key three situations were identified, on the basis of which the comparison of the methods of rating calculation was made. The results of experiments comparing of the rating according to the description of method and the calculation of the rating as the arithmetic mean are presented. The analyze results of the experiment are analyzed.

**Keywords** – calculation rating, average arithmetic score, threshold value.

## I. INTRODUCTION

Today, the consumer has a huge selection of interesting objects which provide in the market of services and goods. With a huge amount of diversity, the user faces the question of how to choose the right product or service. It is worthwhile to trust a more famous company or to choose a lower price, to pay attention to customer testimonials or to take the chance and to choose an unknown company without feedback [1].

In the sphere of delivery environment, one of the key parameters of the service evaluation is the rating. The most general way to calculate the rating is to calculate the arithmetic mean. This method does not take into account the number of votes, also the positions with an overvalued rating can appear and in a one rating.

Another way to calculate the rating is to introduce a threshold value. Thanks to this method, less popular (advertised) valuation objects approach the average rating of all objects. The less popular the object, the higher the coefficient its rating for a single rating list rises. The formula for calculating the rating taking into account the limit value is calculated by the formula (1).

$$\text{Rating} = \left( \frac{V}{V+M} \right) * R + \left( \frac{M}{V+M} \right) * C \quad (1)$$

where V - number of votes for the object; M - threshold of votes required to participate in the rating "Top" (on the

application: 500); R - the arithmetic mean of all the voices of the object; C - the average rating of all objects (for example: 7.3837 on a ten-point scale).

This technique eliminates less popular objects at the expense of the threshold, and in order to obtain a higher rating, a representative of the services or goods needs to be more thoroughly invested in advertising [2].

## II. MAINE RESEARCH

In the transport sphere, when carrying out traffic, the driver's assessment is one of the key parameters [3]. In automated systems, where the choice of a car for trip is carried out by the user, the calculating of the driver's rating is an actual task.

Today, the driver's assessment is based on the method of calculating the arithmetic mean. But this technique takes into account only one key indicator - an estimate (mark) made by the user.

Each driver after an assessment has two key values: the average score and the number of voters.

If we focus only on the average rating, as the arithmetic mean of all the assessments assigned to it, then the evaluation index will not be objective. Since in some cases the assessment can be high, for example 5.0, but this estimate was put only by 3 users. Or the situation when the score is 4.6, obtained from 30 users. The second estimate is lower, but it was supplied by more users, so the sample was larger [4].

There are three cases when the numbers of votes play a key role:

- "cold start", when who appeared in the system just now or recently will always be certainly worse, because it does not fall into the choice of the optimal trip and can not increase his rating;
- an unreasonably high rating when the number of votes is minimal, but the score is high;
- a few low ratings submitted by users greatly underestimate the driver's rating.

In some cases, there are situations when the client has experienced negativity, to which the driver and the quality of the service rendered by him are not involved. But the client can put the low mark to the driver undeservedly just because succumbed to some negative emotions. Possible a situation when there was an error when evaluating by the user. The driver has a one

negative evaluation of user, which strongly affects the overall rating.

Such situations can greatly underestimate the driver's rating. The lower the driver's rating, the less chances that the driver will be popular when user makes his choice. Therefore, in the future any evaluation greatly affects the driver's work and his earnings. Since at a low rating the driver does not take part in the selection of the optimal individual city trip [3].

TABLE I. EXAMPLE OF A RATING CALCULATED BY FINDING THE ARITHMETIC MEAN FOR DIFFERENT SITUATIONS

Situation	1	2	3
The rating obtained as the arithmetic mean	4	5	3,2
Number of votes	2	3	72

The analyzing the results from the table, we can notice that for the third situation, the driver, who has long provided his services and has a 72-vote index of rating is lower than the first two situations. Therefore, the calculation of the rating by this method does not reflect the actual situation.

It follows that the valuation and the number of votes are approximately equal in importance factors. To equalize these two values - you need to bring them to a single scale. The maximum number of votes is always different, but the rating is from 1 to k, in the system at the moment it is accepted k = 5. Therefore, it is necessary to bring the number of driver's votes to the percentage from the highest possible among all drivers of the system.

The final mark - the rating of one driver is built on a set of sets, has the following form:

$$R = \langle V, C, M, L \rangle, \quad (2)$$

where  $R = \{R_i\}_{i=1}^n$  - set of all existing drivers ratings;

$V = \{v_i^j\}_{i=1}^n$  - a plurality of votes reduced to a rating scale corresponding to a plurality of estimates C, where i - is the driver's sequence number, n is the number of drivers to be assessed;

$C = \{c_i^j\}_{i=1}^n$  - a set of estimates corresponding to the plurality of voices V, where i - is the driver's sequence number, j - is the number of his evaluation, n - is the number of drivers being evaluated;

$M = \max v_i^j$  - the maximum number of votes for the i - the warrior, given to the rating scale;

$L = \{L_q\}_{q=1}^6$  - a set of correction factors,

representing the average rating of all drivers, according to the percentage of votes.

The algorithm for finding of the driver's rating in the system is follow:

1. Initially, the maximum existing number of votes is determined and converted to a single scale.

2. Finding the correction factor, according to the number of votes of the analyzed driver.

3. Reducing the number of votes of the analyzed driver to a single dimensional scale.

4. The calculation of the current rating of the analyzed driver.

The finding of the correction factor. The correction factor is the average rating of all drivers. Its meaning was obtained empirically. Based on the number of votes, their percentage ratio is taken from the value of the coefficient from the table. The pattern is taken according to the percentage of voters for the driver analyzed from the maximum number of votes for one driver. For this to purpose, the percentage of the current number of votes from the maximum number of votes which one driver has and calculated:

$$h = \frac{v_i^j * 100\%}{M} \quad (3)$$

where  $h$  - the percentage of votes from the maximum number of voters for one driver, currently available in the system.

After calculating of percentage of votes from the maximum number of voters for each driver, the value of the correction factor is determined from the table.

This correction factor in the form of a percentage match was introduced that to average the single rating in the system. Empirically, it was found that the introduction of a single common average rating of all drivers, understates the value of the rating of the more active drivers. That is why, the indicators were calculated for which the averaged rating was taken in groups by the number of votes. Table II shows the percentage and the corresponding correction factor.

TABLE II. EXAMPLE OF DISTRIBUTION OF PERCENT OF VOTES AND THE CORRESPONDING CORRECTION FACTOR

The percentage of votes from the maximum number of voters for one driver, currently available in the system, $h$	Value of correction factor $L_q$
95 - 100%	$L_6 = 5$
85 - 94 %	$L_5 = 4,95$
75 - 84 %	$L_4 = 4,9$
51 - 74 %	$L_3 = 4,85$
26 - 50 %	$L_2 = 4,8$
1 - 25 %	$L_1 = 4,75$

Reducing the number of votes for the analyzed driver to a single dimensional scale. In order to bring the number of votes for the analyzed driver to the scale of rating, it is necessary to use the formula (4).

$$v'_i = \log_k v_i \quad (4)$$

where  $v_i$  – number of evaluations exhibited by the analyzed driver;  $k$  – the maximum value of the selected scale,  $k=5$ .

After all the input data has been brought to a single dimensional scale and the correction factor is determined, the current driver's rating is calculated, formula (5).

$$R_i = \left( \frac{v'_i}{v'_i + M} \right) * \sum_j c_i^j + \left( \frac{M}{v'_i + M} \right) * L_q \quad (5)$$

Thus, a clear rating value is brought to a single form, where the significant indicators are the number of votes and the direct assessment itself.

the same three control situations were taken that to compare the calculation of the rating by the proposed method and as the arithmetic mean, Comparison of control situations is given in Table III.

TABLE III. COMPARISON OF THE RATING AND ARITHMETIC MEAN OF THE RATING IN DIFFERENT SITUATIONS

Situation	1	2	3
The rating obtained as the arithmetic mean	4	5	3,2
Number of votes	2	3	72
The rating obtained by the proposed methodology	4,51	4,82	4,07

Using this technique, we can adjust the pricing of the estimates, taking into account the number of votes as an equivalent parameter together with the estimate itself. In addition, this method of calculating the rating allows to improve the result with a "cold start", thereby giving a chance to new drivers to take part in choosing the optimal individual city trip. Also, in the case of a single negative mark, the overall rating is not lowered so drastically.

Based on the developed methodology, a clear rating value is calculated, which in the future can be used in a fuzzy model of determining the optimality of an individual city trip. The rating value is one of the key parameters used to determine the degree of confidence in the optimality of trip with the help of device of the fuzzy logic [3].

To test the adequacy of the above methodology for calculating the driver's rating, an experiment was conducted. This technique was implemented in the Vtaxi.info system [5] and all input values are relevant and were taken from it. For the experiment, the maximum

number of evaluations from one driver was taken, which was 100 votes.

The ratings of drivers were taken in the range from 3 to 5. The rating is lower than 3 is extremely low and assumes that this driver is not able to work with clients or provide service an individual city trip. After that, the main indicators included in the final mark of the rating were calculated.

In total, 260 ratings of drivers were taken for experiment. To demonstrate the results from the entire experiment, a random minimum sample was taken. Below is a table with examples of calculating all the key indicators and the final ranking in the system. Also in the table for each rating is given the arithmetic mean of value to the analyzed driver.

TABLE IV. EXAMPLE OF CALCULATING OF THE RATING IN THE SYSTEM

Number	Number of votes	$v'_i$	M	$L_q$	$\sum_j c_i^j$	$R_i$
1	2	3	4	5	6	7
1	3	1,24	2,51	4,75	5	4,82
2	18	1,78	2,51	4,75	4,6	4,67
3	72	2,35	2,51	4,85	3,2	4,07
4	91	2,46	2,51	4,95	4,9	4,89
5	27	1,93	2,51	4,95	4,75	4,86
6	66	2,31	2,51	4,85	4,1	4,49
7	82	2,41	2,51	4,9	4,55	4,73
8	97	2,5	2,51	5	4,95	4,98
9	46	2,15	2,51	4,8	4,3	4,57
10	58	2,25	2,51	4,85	4,85	4,85
11	21	1,84	2,51	4,75	4,95	4,83
12	61	2,28	2,51	4,85	4,6	4,73
13	74	2,37	2,51	4,85	4,5	4,68
14	33	2,01	2,51	4,8	3,9	4,4
15	11	1,62	2,51	4,75	3	4,06
16	81	2,41	2,51	4,9	4,2	4,56
17	93	2,48	2,51	4,95	4,75	4,85
18	69	2,33	2,51	4,85	4	4,44
19	9	1,55	2,51	4,75	5	4,85
20	2	1,15	2,51	4,75	4	4,51
21	44	2,13	2,51	4,8	4,6	4,71
22	99	2,51	2,51	5	4,95	4,98
23	100	2,51	2,51	5	5	5

Since the sample of data was taken for demonstration the results of the experiment was randomly generated, that is why an equivalent search of all values was not established. But in the sample the main key three situations are presented, on the basis of which the methods for calculating of ratings are compared. The final results of rating results shown in the table are displayed on the chart, and the rating is calculated as the arithmetic mean and the rating calculated taking into account the threshold value. For calculation of the rating taking into account the threshold, the following values



were taken: a threshold value of 100 votes, the average value in the system for all existing ratings is 4.2.

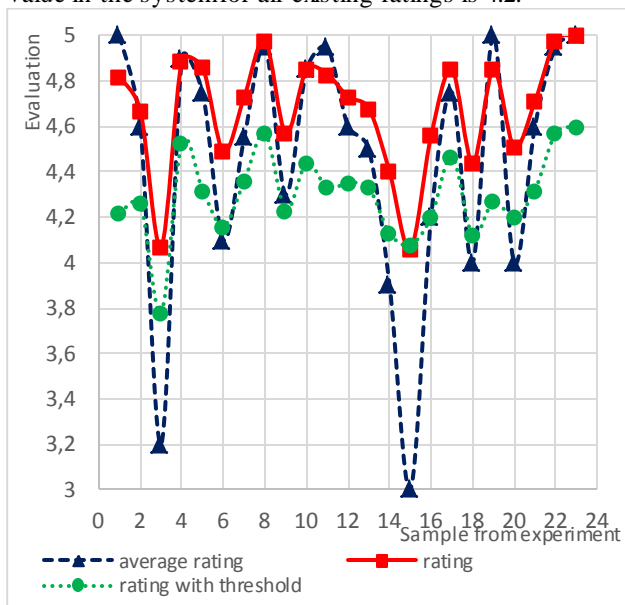


Figure 5. The comparisons of driver and average mark

The results shown in Figure 1 visually demonstrate that with a small number of votes, the final rating of the driver is underestimated, unlike the average arithmetic rating. Because of the small number of votes, we can not judge about adequacy of the rating in a whole. At the same time, with an increase in the number of votes, the final rating increases with respect to the arithmetic average. This is due to the fact that the more often the driver is assessed, the more popular he is in the system. The cases where one or few negative estimates significantly underestimate the value for the average arithmetic rating, according to the result of the developed methodology, the changes occur in an insignificant manner. That is, a very low driver's rating means a negative evaluation of a large number of users.

Analyzing the results for a rating with a threshold value, you can see that the value for all drivers does not exceed the score of 4.6. The maximum value of 4.6 is achieved in those cases when the number of votes is close to the threshold value. Comparing with the other two ratings, the rating with the threshold gives much lower results.

Thus, the rating value, which is reduced to a single kind of key parameters gives a more accurate and less understated result. The key parameters for the developed methodology are the number of votes and the assessment itself, these parameters are equivalent and have the same weight. This ranking methodology allows you to more accurately assess the objects, providing the user with an accurate and objective assessment.

#### CONCLUSIONS

The rating of driver is one of the key parameters in the model of an optimal individual city trip. Because the user chooses a representative of the service the most often on the basis of the rating. To calculate the rating, there are several basic techniques. The calculation of the arithmetic mean, calculation using a threshold value, which is the main criterion for selection in the top and calculation of the rating according to the proposed methodology. The most popular method for calculating the rating is the arithmetic mean, is not accurate because it does not take into account the number of votes the analyzed object has. The calculating the rating with a threshold, works fine for large samples, when the system has a set of estimates for each object. This technique too much underestimates the "newcomers" not allowing them to break into the top of the top, because their rating is certainly lower.

If we take into account the number of votes and the ratings themselves as equivalent criteria in calculating the rating, then the "beginner" has all the possibilities to get into the top. On the other hand, he does not fall immediately to the very top of the top ratings, as when using the method of arithmetic mean.

Since the number of votes and the score itself are two criteria not brought to a single scale, it is not possible to use them in a single formula. The above method is based on bringing two key criteria to a single scale using a series of formulas. Since this technique has already been implemented in the working system Vtaxi.info, a reduction to a single five-point scale was performed. The formulas are universal and can be used to bring to any order of the scale. Empirically, it was revealed that the correction factor and percentage of votes corresponded from the maximum value of existing votes which one driver has.

To verify the adequacy of the developed methodology, an experiment was carried out, part of which is given in the work. Based on a sample of ratings submitted to drivers, the rating is calculated as the arithmetic mean and the rating according to the developed methodology. The results of the comparison showed the advantages of the developed method.

#### REFERENCES:

- [1] "How Star Ratings Influence Customers' Behavior". [Online]. Available: <https://www.yotpo.com/blog/star-ratings-influence-customers>. [Accessed: April. 19, 2018].
- [2] R. A. Britsov "Raking data based on user's marks and behavior", *T-Comm*, vol 10, no.1, pp. 62-65, 2016. (in Russian).
- [3] O. I. Pronina, "Formalized representation of an individual city trip on the basis of linguistic variables", *Bulletin of Kharkiv National University of Air Forces named after. Ivan Kozhedub "Information Processing Systems"*, vol. 1, no.151, pp. 39-47, 2017.
- [4] N. V. Komleva and O. A. Khlopkova, "Processing the information content on the basis of fuzzy neural model of decision making", *Statistics and economics. Applied Informatics*, vol. 5, pp. 188-192, 2013.
- [5] S. V. Parasiuk, "The method of processing and distribution of orders transport services", UA, No. 82013, 2013.

# Mathematical Model of the Managing Strategy Selection in Public Catering Establishment (Based on Markov Decision-Making Problem)

Liliana Danilova

Sumy State University, danilova8085@gmail.com

**Abstract** – The possibility of introducing a mathematical model for choosing the management strategies of a catering institution is explored. The main business processes of conducting economic activity of the enterprise are analyzed. The optimal mathematical model for constructing a short-term strategic plan of the enterprise is determined.

**Keywords** – catering institution, business process, optimization, management strategy, Markov's process.

## I. INTRODUCTION

The main strategic goal of using information technologies and systems in the restaurants business is ensuring its rapid and regulated development, manageability and quality, competitiveness, saving financial and time resources for business processes.

The most popular process is the development of software products for performing diverse operations, but processes that focus on building an optimal business strategy at different types of enterprises are also important.

These enterprises include catering establishments, which direct their activities to sales efficiency, taking into account moderate prices, attracting a large number of regular customers in the face of fierce competition, providing services with up-to-date and high-quality products.

The modern restaurant business requires the using of information systems in order to increase its efficiency. The level of introduction of innovative technologies in the organization and management of restaurant business directly proportional to the increase in the volume of progressive processes in it.

In order to achieve the goal of success and competitiveness of the public catering establishment, it is necessary not only to know what types of information technologies and automation systems to use, but also to thoroughly study the algorithms for their implementation and application.

The optimal combination of the type of mechanism and the practice of its operation directly affects the degree of compliance with the general requirements for business profitability.

In the modern restaurant business, information technology is used not only as a tool for business processes, but also as a factor that allows a performing of managerial functions. These include the management of the effective activities of the complex of interrelated divisions, as well as the management of the institution as a whole.

## II. ANALYSIS OF ECONOMIC ACTIVITY OF THE ENTERPRISE AND MAIN BUSINESS PROCESSES

An example of a catering institution whose leadership is primarily interested in achieving a higher level of profitability by introducing an optimal management strategy is the CoffeeMan coffee shop (Sumy, Ukraine).

To ensure the efficiency of doing business in the CoffeeMan coffee shop, Keeper is used. POS system (POS-Point of Sale) is a software and hardware complex that provides a typical set of cash functions: accounting and product release, reception and delivery money, cancellation of the purchase, functioning on the basis of the fiscal registrar.

The using of the reception's system and processing orders through the introduction of information technology in the restaurant business is aimed at achieving the main goal - improving the efficiency of processing customer orders, due to achievements in the following areas: saving financial, people and time resources.

## III. ANALYSIS OF THE MAIN PROBLEMS OF THE ENTERPRISE'S ACTIVITY

In the course of the analysis of the public catering establishment's activity the main issues were identified, which negatively affects the achievement of the main goal of the restaurant business - obtaining and raising the level of profitability:

- slowed down processes of receiving and processing orders that affect the high level of resource-intensiveness of business processes aimed at working with the client;
- low level of competitiveness;

- high level of resource intensity of managerial processes for monitoring the activities of structural units;
- lack of a well-established system of searching for methods of attracting potential clients;
- low level of managerial activity to attract additional tangible revenues.

The above problems are significant and create the need to automate the work of the institution through the use of information technology, and also determine the need to choose an effective business management strategy.

#### IV. ANALYSIS OF EXISTING METHODS FOR OPTIMIZING ENTERPRISE ACTIVITY

Optimization of the enterprise activity in the global sense is aimed at introducing innovative technologies and improving business processes. In the process of optimizing the activities of the enterprise, the deficiencies of business processes are eliminated, and at the same time, the most effective business processes are improved [1].

Optimization of the catering institution's activity is a complex of interrelated administrative, organizational, information measures, united by a certain technology, aimed at improving the indicators of both individual processes and indicators of the institution as a whole in order to meet the needs of interested parties.

The optimization procedure involves the implementation of six steps:

- Phase 1. Organization of improvement of the business process.
- Phase 2. Documentation. A choosing of approach.
- Phase 3. Analysis. Identify opportunities for optimization.
- Phase 4. Designing a new, improved business processes.
- Phase 5. Implementation. Implementation of the strategy.
- Phase 6. Management of business processes for continuous improvement.

First of all, the optimization of the company's activity is aimed at its improvement, restructuring of the interrelations of the production substructures functional, which allows the raising of the level of competitiveness, profitability of the enterprise [2].

The basis for choosing the optimal strategy for the development of a catering institution is primarily determined by a number of organizational-economic, legal and labor actions aimed at providing the benefits of the business entity to its competitors.

A prerequisite for building an enterprise development strategy is its strategic analysis, for example, by conducting a SWOT analysis (a method aimed at schematically mapping the main objects for analysis: strengths and weaknesses of the enterprise, threats and opportunities of the environment) [3].

#### V. MARKOV'S PROCESSES AS A MATHEMATICAL WAY OF DESCRIBING THE FUNCTIONING OF COMPLEX SYSTEMS

In order to achieve the highest level of accuracy of the results obtained during the implementation of the management strategy, it is necessary to use mathematical devices that enable to calculate exactly the achieved level of profitability. In order to make the right decisions during the implementation of the proposed project, the choice of the management strategy of the catering institution should mathematically justify the selection of the optimal of the proposed alternatives.

The main task - to develop a model that allows, based on available information about the possible state of business profitability of the enterprise and the transition between these states, as well as about the potential profit, choose the optimal strategy of management through the adoption of strategic decisions. The task of choosing a management strategy is reduced to the Markov problem of decision making, on which the model is developed.

In the process of making optimal solutions should take into account the random factor. This means that the considered random phenomena must obey certain static laws. The condition of static stability allows using effective mathematical methods of the theory of random processes in the decision-making process, and in particular one of its sections, the theory of Markov processes.

The Markov process is a random process with continuous time, the determining feature of which is the exponential character of the functions of the distribution of the time of the process's stay in their states prior to the transition to other states. The advantage of the Markov process is that in addition to evolution, one can investigate such characteristics as the time of the Markov process in a certain state.

In accordance with the complication of Markov processes, the level of their adequacy increases, but at the same time the level of complexity of the mathematical apparatus increases, which is connected with the processing of statistical data.

Markov's decision-making process is a specification of the problem of consistent decision-making for a fully observed environment with the Markov model of transition. It serves as a mathematical basis for simulating decision-making in situations where the results are partially random and partially under the control of the decision maker.

## VI. FORMALIZATION OF GENERAL PRINCIPLES OF DESIGN AND REQUIREMENTS

According to the main theoretical principles of project management, the activity on developing models of choosing the strategies for managing a catering institution can be described by the concept of "IT project".

The concept of an IT project was developed, the main aspects of which were the detailed design of the project goal by the SMART method (that presented with the help of TABLE I), the description of the project's product functionality, the formalization of the limitations and assumptions of the project product [4].

TABLE I. DETAILING THE GOAL OF THE PROJECT BY SMART METHOD

Feature	Description
<b>Specific</b>	Develop a model for use in the restaurant business, which will allow you to modify the input data and algorithm for solving problems in accordance with the requirements of the customer.
<b>Measurable</b>	The results of the model implementation will be evaluated both financially (profitability, etc.) and expert (number of visitors, etc.)
<b>Achievable</b>	The results of the implementation of the model will allow achieving the strategic goals of the enterprise, aimed at satisfying the needs of the institution's management in raising the profitability and needs of visitors in obtaining relevant services and products.
<b>Relevant</b>	The project will be implemented by a team of 5 people, and the budget will be financed by the owner of the catering institution (for which the project is being developed)
<b>Time-framed</b>	The project will be executed on time, regulated by the creation of the Gantt chart and PDM network

Product IT project will be a model of choosing of strategies for managing the catering institution, presented in the form of solved Markov tasks for a specified and indefinite period. The results of the project product will be used as follows:

- calculation of input data (will determine the level of profitability of the institution, choose the best management strategies);
- calculation of the Markov problem for a certain period (will allow to create a temporary strategic plan depending on the chosen period);
- calculation of the Markov problem for an indefinite period (will allow to form a long-term strategic management plan);
- calculation of initial data (will allow to analyze and compare the levels of profitability in the periods before and after the project implementation);

The product of the project is also allowed for use in enterprises for wholesale and retail trade. Modifications in the course of the project implementation are for the

phases of the development of the technical task, the formation of input data and model research.

## VII. IMPLEMENTATION OF THE TOOLKIT FOR MAKING DECISIONS ON CHOOSING THE OPTIMAL DEVELOPMENT STRATEGY FOR A CATERING INSTITUTION

The solution was performed in MS Excel with detailed explanations of each of the steps performed.

**Step 1.** To determine the optimal strategic action in the "CoffeeMan" institution, the task set the number of stages ( $m = 3$  months), the level of sales in the institution (excellent, good and satisfactory) and three strategic measures ( $k_1, k_2, k_3$  : 10% discount on orders for the amount of UAH 300, a tenth of a cup of coffee as a gift, not to take measures).

**Step 2.** The method of conducting expert evaluations was: transitional probabilities between the levels of sales levels and the corresponding profit. These data take the form of interconnected matrices (Fig. 1).

Figure 1. Transitional probability and expected profit matrices

The transition probability matrix is a matrix that shows the probability of transitions between states of sales levels. In this problem, the number of such matrices is reduced to 3, and they are designated respectively P1, P2, P3.

With each transition probability matrix, income matrices R1, R2, and R3 are related, whose values represent the income or damage, depending on the states between which the transition is made.

**Step 3.** The calculation of the expected profit is as follows: cells of the matrix of expected profit actually reflect the sum of the products of the values of transitional probabilities for their respective profits.

Expected profit on project implementation will be calculated according to the Equation (1):

$$v_j^i = \sum_{j=1}^m p_{ij}^k \times r_{ij}^k, \quad (1)$$

The results of the calculation of expected profits are shown in Figure 2.

**Step 4.** Fill in the matrix of expected earnings at stage 3 (3rd month) is reduced to the transfer of values from the above matrix. This is due to the lack of the need to search for the target function, since there is no preceding step (Figure 3).

The optimal strategic action for the current month has been found for the selected maximum values from each line.

**Step 5.** Next step we calculate the cells of the matrix of expected income in stage 2 by the Equation (2):

$$f_2(i) = v_i^k + p_{i1}^k \times f_3(1) + p_{i2}^k \times f_3(2) + p_{i3}^k \times f_3(3) \quad (2)$$

At this stage, the visibility of the relationship between the values of the cells of the filled matrix with the values of expected revenues and target functions that were found in the previous step appears.

Figure 2. Tables of the calculation of the value of expected profit

$i$	1	2	3
$v_i^1$	1840	1405,5	697,5
$v_i^2$	1528	1107	722,4
$v_i^3$	1372	1342,8	592

Stage 3:					
$i$	$v_i^k$			An optimal solution	
	$k=1$	$k=2$	$k=3$	$f_3(i)$	$k^*$
1	1840	1528	1372	1840	1
2	1405,5	1107	1342,8	1405,5	1
3	65	722,4	592	722,4	2

Figure 3. Completed matrix of expected profit in stage 3

**Step 6.** We find the value of expected income at stage 1 (month 1), similarly using the rules of the relationship with the calculated values of the cells of the previous stage. The calculation is based on Equation (3):

$$f_1(i) = v_i^k + p_{i1}^k \times f_2(1) + p_{i2}^k \times f_2(2) + p_{i3}^k \times f_2(3) \quad (3)$$

After performed calculations, the best strategic actions should be chosen for each step separately and determine the expected profits for 3 months, taking into account the levels of sales and interconnections between the stages.

Stage 3:					
$i$	$v_i^k$			An optimal solution	
	$k=1$	$k=2$	$k=3$	$f_3(i)$	$k^*$
1	1840	1528	1372	1840	1
2	1405,5	1107	1342,8	1405,5	1
3	65	722,4	592	722,4	2

Stage 2:					
$i$	$v_i^k$			An optimal solution	
	$k=1$	$k=2$	$k=3$	$f_2(i)$	$k^*$
1	3369,6	2815,5	2591,2	3369,58	1
2	2693	2233	2599,2	2692,97	1
3	1911,2	1649,7	1506,9	1911,18	1

Stage 1:					
$i$	$v_i^k$			An optimal solution	
	$k=1$	$k=2$	$k=3$	$f_1(i)$	$k^*$
1	4782,6	4121,8	3887,6	4782,578	1
2	3999,3	3533,9	3874,1	3999,255	1
3	2210,7	2868,1	2732,5	2868,117	2

Figure 4. Completed expected profits of three stages

## CONCLUSIONS

In the process of scientific research, the following theoretical aspects concerning restaurant business were considered: the peculiarities of the economic activity of the catering institution, the main business processes of the coffee shop, the main problems of the enterprise.

The next step was to analyze the existing methods of optimizing the company's activities and review the main principles and methods for choosing a management strategy for a catering institution.

During the analysis of information on the construction of the strategic plan, an urgent need was made for the use of the mathematical apparatus to calculate data regarding the level of profitability and the identification of optimal strategic measures for raising the level of competitiveness of the institution.

The main problems of the institution are caused by the processes of strict competition of surrounding institutions. It directly affects the increase of the loss-making enterprise. The above-mentioned factor was conditioned by the necessity of setting the task for modeling an optimal strategic plan, which would increase the profitability, attraction of potential customers, and competitiveness.

During the analysis of the methods of choosing the best strategies for managing the institution to achieve the goals, the mathematical apparatus of the Markov processes was selected as the main method. The staging of the Markov problem for the selection of optimal control strategies is conditioned by the relative simplicity of the apparatus. The following feature is also important: the Markov process serves as a mathematical basis for simulating a decision in situations where the results are partially random and partly under the control of the decision maker.

## REFERENCES:

- [1] H. James Harrington, Erik K. C. Esseling, Harm Van Nimwegen, *Business Process Improvement Workbook: Documentation, Analysis, Design, and Management of Business Process Improvement*. USA: The McGraw-Hill Companies Inc., 1997, pp. 132-144.
- [2] Michael D. Intriligator, *Mathematical Methods of Optimization and Economic Theory*. Moscow, Russia: Airis-Press, 2002, pp. 54-67. (in Russian)
- [3] S. Cook, *Process improvement: a handbook for managers*. Aldershot, Hampshire, England; Brookfield, USA: Gower, 1996, pp. 113-159.
- [4] Project Management Institute, *PMBOK Guide*. Moscow, Russia: Olimp-Business, 2014. (in Russian)

# Object Detection Based On Growing Convolutional Neural Network for Autonomous Systems

Viacheslav Moskalenko, Alona Moskalenko<sup>1</sup>, Artem Korobov, Borys Lypivets  
Sumy State University, Ukraine, <sup>1</sup>a.moskalenko@cs.sumdu.edu.ua

**Abstract** — Model based on convolutional neural network for features extraction and information-extreme classifier for object detection without localization on video frame with new training methods to build decision rules are presented in the paper. Transfer learning for extract low-level features, Growing sparse-coding neural gas algorithm for unsupervised training of the high-level convolution filters and Fish School Search algorithm for training the classifier model are proposed. Unsupervised training of convolutional neural network allows to efficient using unlabeled datasets which is majority of available data. Information-extreme classifier is characterized by low computational complexity and better accuracy in case of small labeled dataset. Simulation results with optimal model on test open datasets confirm the suitability of proposed algorithms for practical usage.

**Keywords** – object detection system, convolutional neural network, growing neural gas, sparse coding, information criterion, information-extreme machine.

## I. INTRODUCTION AND RELATED WORK

Existing autonomous systems for detection object of interest on video frame still do not provide high-reliability decisions, as there are the unsteadiness of the environment and variety of object's modifications and a small number of relevant labeled data [1, 2].

Thus, the use of handcrafted features in object detection systems for the description of observations leads to a decline the informativeness of the features and the training effectiveness of the decision rules [2, 3]. Therefore, the most promising approach to the synthesis of a features extractor is the use of ideas and methods of machine learning for the hierarchical representation of observations for unlabeled data and transfer knowledge [4, 5].

Convolutional multilayer neural networks allow forming an informative hierarchical features presentation of observations [5]. In addition, they have already shown high efficiency in solving problems of machine vision and analysis of time series [6, 7]. However, the unsupervised training of convolutional networks is usually carried out based on using an autoencoder or

Restricted Boltzmann machine, which requires a large amount of training data and long learning time to obtain an acceptable result. In work [8] it is proposed to use cluster-analysis based on the k-means algorithm to train convolution filters. However, k-means is characterized by slow convergence and sub-optimality of the results due to the hard-competitive nature of its learning scheme and the sensitivity to initial cluster initialization.

In work [9] is proposed a combination of the principles of neural gas and sparse coding for the training of convolutional filters for unlabeled data. Given approach is characterized by soft-competitive learning scheme that facilitates robust convergence to close to optimal distributions of the convolutional filters over the training sample. At the same time, embedding of sparse coding methods can increase the immunity against interference and generalization ability of features representation. However, the number of codebook vectors is unknown beforehand and it is determined or optimized by the developer, which leads to an increase in the number of training iterations.

The required amount of convolution filters in each convolutional layer is difficult to predict in advance, so the promising approach to learning convolutional filters is to use the principles of growing neural gas, which automatically determines the required number of neurons [10]. The presence of a mechanism for the adding of new neurons, as well as the removal of excessive old ones, makes the algorithm more flexible compared to the classical neuron gas, but it also has serious disadvantages. The small values of the period between the iterations of the generation of new neurons  $\lambda$  lead to the instability of the learning process and the distortion of the formed structures, as here observed the excessively frequent adding of new neurons. The high value of the period  $\lambda$  provides the expected effect, but at the same time it leads to a significant slowdown in the algorithm. However, in the work [11] it was shown that achieving stability of learning could be done by setting the “radius of reach” of the neurons, which involves the replacement of the parameter  $\lambda$  on the threshold of maximum distance of the neuron from each points of the training set attributed to it. However, the mechanisms for updating neurons and

assessing the remoteness of the points of the input space to the neurons have not yet been reviewed in order to adapt the learning process to the sparse coding of observations.

In addition to the feature extractor an important component of object detection system is the decision rules, which can be represented by classifier. In some cases, decision rules also include a regression model to prediction bounding box for detected object on image. At the same time, the effectiveness of training a classifier is often considered as a measure of the effectiveness of learning the features extractor [5].

The most popular method for classification analysis is the support vector machine, where the training of decision rules takes place within the framework of a geometric approach by constructing in the secondary features space a linear separable hyper surface. However, this algorithm requires a lot of manual adjustments for regularization of the model and its performance depends on the complexity of the kernel functions of the transformation of the features space [12]. In work [13], were proposed the construction of decision rules by adaptive binary quantization of the features space and the construction of radial basis functions in the Hamming binary space. Such a classifier has high operational efficiency, since it uses low computing complexity operations as comparison and “exclusion OR”. In this case, the use of population-based meta-heuristic algorithms for searching optimal parameters which provide global maximum of information criterion of learning effectiveness can increase the reliability of the model and the efficiency of its hyper parameters tuning [13].

The question of choosing the optimal in the informational sense of the number of neurons and their weight at each level of the hierarchical model of object detection in conditions of unsteadiness and limited volume of labeled training dataset are not sufficiently investigated and are still not fully resolved

**The purpose of the paper** is the development of a new method of training a convolutional network with a pre-unknown number of filters, as well as increasing in the informational sense the effectiveness of machine learning for the detection of objects on video frame in condition of the limited computing resources and the small volume of labeled training sample.

## II. MATERIALS AND METHODS

The use of any prior information to synthesis of the architecture of the convolutional network for visual information analysis in the conditions of the limited computing resources of the autonomous system can increase the efficiency of its training, re-training and inference. Transfer Learning is one of the good examples of priori information accumulation about the features of visual patterns and its reuse. Moreover, the closer the domain area where neural network was trained to the new

domain area, in which accumulated knowledge will use, the more layers of the network can be borrowed. High-level layers of the neural network usually need training from scratch to maximize their adaptation to the new domain area. Prediction of the presence of an object of interest in the video frame is performed by information-extreme classifier. Fig. 1 shows the architecture of the proposed system.

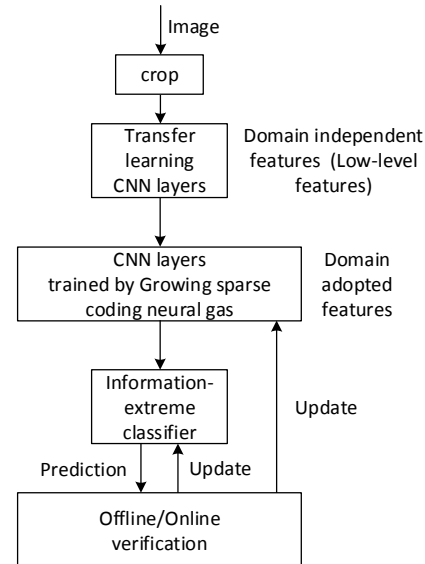


Figure 1. The architecture of the detection system based on low-level transfer learning layers, high-level layers trained by growing sparse coding neural gas and information-extreme decision rules

Around 180 images from Inria Aerial Image Labeling Dataset is used to train the object detection system in the field of view of the video camera [7]. Each image has a resolution of  $5000 \times 5000$  px. Amount of generated images using random  $227 \times 227$  px cropping for unsupervised training of convolutional network is 10,000. For training and evaluation of classifier model, 500 images of  $500 \times 500$  px with vehicles and 500 images of  $500 \times 500$  px without vehicles were selected. Each image with shape  $500 \times 500$  px are rotated and randomly cropped to series of images with shape  $227 \times 227$  in order to avoid the influence of objects localization on classification result.

Transfer Learning layers is copied from trained convolutional neural network Squeezenet [5]. It is proposed to encode incoming images with a feature map from output fire7 module of Squeezenet. In this case, shape of the feature map is  $27 \times 27 \times 384$ . To construct high-level features adapted to new domain, it is suggested to use two additional convolutional layers with sub-sampling operation between them. In convolutional layers filters with  $1 \times 1$  and  $3 \times 3$  kernels are used that training in unsupervised manner. The number of convolution filters is not fixed and is determined during layer-wise learning. The convolutional filters of each level are three-dimensional, the depth of which is equal to the number of

channels of the input activation map. The pixel activation of each channel of features map is offered to calculate based on Orthogonal Matching Pursuit algorithm with the function of ReLU activation [4, 9].

The dataset for training of convolutional filter is formed by decomposition of activation maps to 3D-dimensional patches that match the size of the filter layer. These patches are reshaped to 1D vectors, which put on the input of growing sparse coding neural gas algorithm, main steps of which are given below.

1. Initialization of the counter of training vectors  $t := 0$ .
2. Two initial nodes (neurons)  $w_a$  and  $w_b$  are assigned by random selection from the training set. Nodes  $w_a$  and  $w_b$  are connected by an edge whose age is zero. These nodes are considered non-fixed.
3. Selected from the dataset the following vector  $x$ , which is normalized to a unit length (L2-normalization).
4. Normalizing each base vector  $w_k, k = \overline{1, M}$  to a unit length (L2-normalization).
5. Calculation of the similarity of the input vector  $x$  to the base vectors  $w_{s_k} \in W$  for their sorting

$$-(w_{s_0}^T x)^2 \leq \dots \leq -(w_{s_k}^T x)^2 \leq \dots \leq -(w_{s_{M-1}}^T x)^2.$$

6. The closest node is selected  $w_{s_0}$  and the second closest to the node  $w_{s_1}$ .
7. Increase the age of all incident edges  $w_{s_0}$  by one.
8. If  $w_{s_0}$  is fixed, then should move to step 9, otherwise, to step 10.
9. If  $(w_{s_0}^T x)^2 \geq \nu$ , then proceed to step 12. Otherwise, should be added a new non-fixed neuron  $w_r$  to a point that coincides with the input vector  $w_r = x$ , also is adding a new edge, that connects  $w_r$  and  $w_{s_0}$ , then proceed to step 13.
10. The node  $w_{s_0}$  and its topological neighbors (the nodes connected to it by the edge) are displaced in the direction to the input vector  $x$  according to Oja's rule [9] by next formulas

$$\Delta w_{s_0} = \varepsilon_b \eta_t y_0 (x - y_0 w_{s_0}), \quad y_0 := w_{s_0}^T x,$$

$$\Delta w_{s_n} = \varepsilon_n \eta_t y_n (x - y_n w_{s_n}), \quad y_n := w_{s_n}^T x,$$

$$0 < \varepsilon_b \leq 1, \quad 0 < \varepsilon_n \leq \varepsilon_b,$$

$$\eta_t := \eta_0 (\eta_{final} / \eta_0)^{t/t_{max}},$$

where  $\Delta w_{s_0}, \Delta w_{s_n}$  – vectors of correction of weight of the neuron-winner and its topological neighbors, respectively;  $\varepsilon_b, \varepsilon_n$  – the constants of updated forces of weighting coefficients of the neuron-winner and its

topological neighbors respectively;  $\eta_0, \eta_t, \eta_{final}$  – initial, current and final learning rate respectively.

11. If  $(w_{s_0}^T x)^2 \geq \nu$ , note the neuron  $w_{s_0}$  as fixed.
12. If  $w_{s_0}$  and  $w_{s_1}$  re connected by edge, then its age is reset, otherwise a new edge with a zero age is formed between  $w_{s_0}$  and  $w_{s_1}$ .
13. All edges in the graph with the age more then  $a_{max}$  are removed. In the case that some nodes do not have incident edges (become isolated), they are also removed.
14. If  $t < t_{max}$  then proceed to step 15, otherwise – increment of the counter of steps is  $t := t + 1$  and then proceed to step 3.
15. If all neurons are fixed, the execution of the algorithm stops, otherwise proceed to step 3 and a new epoch of learning begins (repetition of the training set).

The information-extreme classifier that evaluates the relevance of network traffic to one of the  $R$  classes performs quantization of a features description in the training set  $\{x_{r,i}^{(j)} \mid i = \overline{1, N}; j = \overline{1, n_r}; r = \overline{1, R}\}$ . It is done by comparing the value of  $i$ -th features with the corresponding  $l$ -th threshold of  $L$ -dimensional receptive field,  $l = \overline{1, L}$ . That is, the formation of a binary training set  $\{b_{r,i}^{(j)} \mid i = \overline{1, N \cdot L}; j = \overline{1, n_r}; r = \overline{1, R}\}$  is carried out according to the rule

$$b_{r,i}^{(j)} = \begin{cases} 1, & \text{if } x_{i, \max} [1 - \delta_{i,l} / \delta_{\max}] \leq x_{r,i}^{(j)} \leq x_{i, \max}; \\ 0, & \text{otherwise.} \end{cases}$$

The calculation of the coordinate values of the binary support vector  $x_r$ , according to which the construction in the radial basis of class containers is taking place, is carried out by the next rule:

$$b_{r,i}^{(j)} = \begin{cases} 1, & \text{if } \frac{1}{n_r} \sum_{j=1}^{n_r} b_{r,i}^{(j)} > \frac{1}{R} \sum_{c=1}^R \frac{1}{n_c} \sum_{j=1}^{n_c} b_{c,i}^{(j)}; \\ 0, & \text{otherwise.} \end{cases}$$

As a criterion of the effectiveness of classifier's machine learning, the standardized modification of the Kullback's information criterion is considered [13]:

$$E_r = \frac{1 - (\alpha_r + \beta_r)}{\log_2(2 + \zeta) + r \log_2 10} \cdot \log_2 \left[ \frac{2 - (\alpha_r + \beta_r) + \zeta}{(\alpha_r + \beta_r) + \zeta} \right], \quad (1)$$

where  $\alpha_r, \beta_r$  – errors of the first and second kind of classification decisions regarding the membership of the input vectors to  $r$ -th class;  $\zeta$  – any small non-negative number that is entered to avoid uncertainty when dividing by zero.

In order to optimize the parameters vector of the receptive fields of classifier  $\{\delta_{i,l} \mid i = \overline{1, N}; l = \overline{1, L}\}$  is



proposed to use the population-based Fish School Search (FSS) algorithm, which is characterized by simplicity of implementation, interpretation and high convergence rate [14]. In this case, the optimization of each level of receptive fields occurs separately and consistently. In this algorithm, the position of the agent in the N-dimensional space of solutions is represented as a numerical vector  $P^{(N)}$  with a length N and which is corresponding to the vector of optimized parameters. The individual success of each agent (fish) in the process of finding a solution is characterized by its weight, which plays the memory role. Each search iteration performs two groups of operators – feeding and swimming operators.

The feeding operator formalizes the success of the agents research of the “aquarium” areas and calculates the weight of the  $z$ -th agent, which is proportional to the normalized difference of the values of the fitness function in the next and the current iterations. In the FSS algorithm distinguish three types of swimming – individual, instinctively-collective and collective-volitional. These types of swimming are carried out sequentially one after another at separate intervals of time.

During an individual swimming of agents there is their movement has an equally probable random nature. In this case, for one iteration of the FSS algorithm, the step of individual swimming is executed a fixed number of times. In the process of instinctively-collective swimming, each of the agents is influenced by all other agents of the population and its influence is proportional to the individual success of the agents. Collective volitional swimming is considering as a shift of all agents in the direction of the current center of population gravity, in the case of increasing the total weight of the fish shoal as a result of individual and instinctively-collective swimming. If the total weight is reduced then the shift occurs in the opposite direction.

Thus, an assessment of the effectiveness of the learning of convolutional features extractor, the filters of which are trained by proposed growing sparse coding neural gas algorithm, will be carried out by the results of information-extreme training. At the same time, the process of machine learning of the information-extreme classifier relates to the search a global maximum of the modification of the information criterion by S. Kulbak’s according to the FSS population algorithm.

### III. RESULT OF THE SIMULATION AND DISCUSSION

In growing sparse coding neural gas algorithm were chosen the following parameters  $\varepsilon_b = 0.5$ ,  $\varepsilon_b = 0.05$ ,  $\alpha_{\max} = 100$ ,  $\eta_0 = 1$  та  $\eta_{\text{final}} = 0.01$ . The parameter of the threshold of neuron fixation  $\nu$  and the parameter of the number of thresholds per the feature  $L$  of the system of receptive fields of the classifier are adjusted by scrolling through the values. Table 1 shows the dependence of the number of neurons in the first  $M_1$  and second  $M_2$

trainable layers, the criterion of the effectiveness of training averaged over the classes  $\bar{E}$  and accuracy by the validation sampling of the parameters  $\nu$  and  $L$ .

TABLE I. DEPENDENCE OF INFORMATION CRITERIA AND NUMBER OF NEURONS FROM MODEL PARAMETERS

$\nu$	$L$	$M_1$	$M_2$	$\bar{E}$	Validation accuracy
0,5	1	50	41	0.36	90.0
0,6	1	72	52	0.39	91.0
0,7	1	95	300	0.42	92.0
0,8	1	320	1500	0.42	92.0
0,5	2	50	41	0.46	93.0
0,6	2	72	52	0.54	95.0
0,7	2	95	300	0.65	97.0
0,8	2	320	1500	0.65	97.0
0,5	3	50	41	0.46	93.0
0,6	3	72	52	0.55	95.2
0,7	3	95	300	0.81	98.9
0,8	3	320	1500	0.83	99.0
0,5	4	50	41	0.46	93.0
0,6	4	72	52	0.55	95.3
0,7	4	95	300	0.74	98.1
0,8	4	320	1500	0.83	99.0

The analysis of Table 1 shows that increasing the threshold  $\nu$  leads to an increase in the number of convolutional filters in the process of unsupervised training the features extractor. At the same time, increasing the threshold from 0.7 to 0.8 practically does not affect the accuracy of the decision rules. It means, that the value  $\nu^* = 0,7$  is optimal and allows to form a more compact features representation (compression effect), meanwhile  $\nu = 0,8$  allows to form a sparse representation based on overcomplete basis.

The optimal value of the hyper parameter  $L^*$  is equal to 3. Further increase of the parameter  $L$  does not lead to an increase in the accuracy of the decision rules. At optimal parameters of the extractor and the classifier, the accuracy of detection of malware traffic is 98.9%.

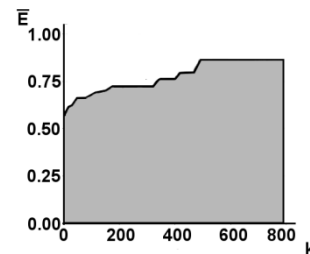


Figure 2. A graph of the change of the average information criterion (2) in dependence from the number of iterations of the optimization swarm search algorithm

The analysis of Fig. 2 shows that the algorithm required only 500 iterations to reach the global maximum. It indicates on the informative nature of the features descriptive of observation. Fig. 2 shows the dependence of the information criterion (2) on the code radius of the container of each class.

The analysis of Fig. 2 shows that the maximum values of information criterion of learning for the first and second classes are equal to  $E_1^* = 0.80$  and  $E_2^* = 0.82$ , respectively, and the optimal values of radii of the corresponding containers of the classes of recognition  $d_1^* = 8$ ,  $d_2^* = 15$ . In this case, the inter-center Hamming distance is 23, indicating compactness of the feature vector distributions.

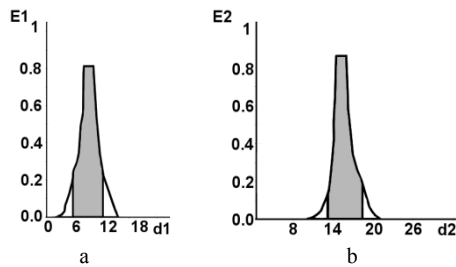


Figure 3. Charts of dependency of the information criterion (1) on the radii of containers of classes: a – class of normal traffic; b – class of malware traffic

Thus, the proposed training algorithm allows determining automatically the optimal number of neurons at each high-level convolution level in the information sense. In this case, the results of simulation on data from Inria Aerial Image Labeling dataset showed the suitability of received models for practical application.

## CONCLUSIONS

1. **The scientific novelty** of the obtained results is as follows:

- for the first time is proposed the growing sparse coding neural gas algorithm that allows to do incremental unsupervised dictionary learning adapted to sparse coding feature representation;

- it was improved the algorithm of the growing sparse coding neuron gas is improved by introducing the radius of the reach of the neurons, which allows instead of setting the upper limit of the number of neurons use the direct regulation of the accuracy of the training set reconstruction;

- it was improved the method of effectiveness evaluation of the features extractor for the task of detecting object of interest on video frame through information-extreme machine learning, which allows to automate the process of configuring hyper parameters of the model.

2. **The practical value** of obtained results obtained for autonomous object detection systems is a developing a new learning method that effectively uses both labeled and unlabeled training sets. In this case, the results of simulation with using Inria Aerial Image Labeling dataset confirm the effectiveness of the obtained decision rules of vehicle detector in test image samples.

## ACKNOWLEDGMENT

The work was performed in the laboratory of intellectual systems of the computer science department at Sumy State University with the financial support of the Ministry of Education and Science of Ukraine in the framework of state budget scientific and research work of DR No. 0117U003934.

## REFERENCES:

- [1] K. Levi and Y. Weiss, "Learning object detection from a small number of examples: the importance of good features," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 2004, vol. 2, pp. II-53-II-60 DOI: 10.1109/CVPR.2004.1315144.
- [2] G. Zhang, J. Kato, Y. Wang and K. Mase, "How to initialize the CNN for small dataset: Extracting discriminative filters from pre-trained model," *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Kuala Lumpur, 2015, pp. 479–483. DOI: 10.1109/ACPR.2015.7486549.
- [3] B. A. G. D. Oliveira, F. M. F. Ferreira and C. A. P. D.S. Martins, "Fast and Lightweight Object Detection Network: Detection and Recognition on Resource Constrained Devices," in *IEEE Access*, vol. 6, pp. 8714–8724, 2018. DOI: 10.1109/ACCESS.2018.2801813.
- [4] X. Ren, K. Chen, X. Yang, Y. Zhou, J. He and J. Sun, "A new unsupervised convolutional neural network model for Chinese scene text detection," *2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, Chengdu, 2015, pp. 428–432. DOI: 10.1109/ChinaSIP.2015.7230438
- [5] G. B. Hacene, V. Gripon, N. Farrugia, M. Arzel and M. Jezequel, "Budget restricted incremental learning with pre-trained convolutional neural networks and binary associative memories," *2017 IEEE International Workshop on Signal Processing Systems (SiPS)*, Lorient, 2017, pp. 1–6. DOI: 10.1109/SiPS.2017.8109978.
- [6] B. Zhao, H. Lu, S. Chen, J. Liu and D. Wu, "Convolutional neural networks for time series classification," in *Journal of Systems Engineering and Electronics*, vol. 28, no. 1, pp. 162–169, Feb. 2017. DOI: 10.21629/JSEE.2017.01.18.
- [7] H. Tayara, K. Gil Soo and K. T. Chong, "Vehicle Detection and Counting in High-Resolution Aerial Images Using Convolutional Regression Neural Network," in *IEEE Access*, vol. 6, pp. 2220–2230, 2018. DOI: 10.1109/ACCESS.2017.2782260
- [8] Qiying Feng, C. L. P. Chen and Long Chen, "Compressed auto-encoder building block for deep learning network," *2016 3rd International Conference on Informative and Cybernetics for Computational Social Systems (ICCS)*, Jinzhou, 2016, pp. 131–136. DOI: 10.1109/ICCS.2016.7586437
- [9] K. Labusch, E. Barth and T. Martinetz, "Sparse coding neural gas: learning of overcomplete data representations," *Neurocomputing*, 2009, vol. 72, i. 7–9, pp. 1547–1555. DOI: 10.1016/j.neucom.2008.11.027.
- [10] I. Mrazova and M. Kukacka, "Image Classification with Growing Neural Networks," *International Journal of Computer Theory and Engineering*, vol. 5, no. 3, 2013, pp. 422–427. DOI: 10.7763/IJCTE.2013.V5.722.
- [11] E. J. Palomo and E. López-Rubio, "The Growing Hierarchical Neural Gas Self-Organizing Neural Network," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 9, pp. 2000–2009, Sept. 2017.
- [12] S. Kim, Z. Yu, R. Man Kil, and M. Lee, "Deep learning of support vector machines with class probability output networks," *Neural Networks*, vol. 64, 2015, pp. 19–28. DOI: 10.1016/j.neunet.2014.09.007
- [13] A. S. Dovbysh, and M. S. Rudenko, "Information-extreme learning algorithm for a system of recognition of morphological images in diagnosing oncological pathologies," *Cybernetics and Systems Analysis*, 2014, vol. 50, i. 1, pp. 157–162. DOI: 10.1007/s10559-014-9603-y.

# On-board Geographic Information System of Images' Identification

Juliy Simonovskiy<sup>1</sup>, Vladislav Piatachenko<sup>1</sup>, Nikita Mironenko<sup>2</sup>

<sup>1</sup> Sumy State University, Ukraine, <sup>2</sup> National Technical University «Kharkiv Politechnic Institute», Ukraine,  
<sup>1</sup>bronseghoste@gmail.com, <sup>2</sup>nikitam1996@ukr.net

**Abstract** – The problem of the identification of the images of the e-map of the region, compiled with aircraft geographic information system, was examined. The method of the images' refinement and the algorithm of the computer-aided teaching of the on-board geographical information system in frames of information-extreme intelligent technology, based on the maximization of the information capacity of the system during its teaching, was suggested. At this, to provide the invariant nature of the final rules the formation of the input teaching matrix was done according to the results of the refinement of the digital images of the region in polar coordinates.

**Keywords** – geographic information system, information-extreme computer-aided teaching, identification, informational criteria, image, polar coordinates.

## I. INTRODUCTION

Widespread usage of the aircraft on-board geoinformation systems (GIS) for the Earth surface observation allows to solve many important tasks for the social-economical sphere of society. One of the important tasks of the on-board GIS is segmenting of the region to localize the object of the observation. This task can be solved, for example, while the search of the vehicle, while the observation of the agricultural crops, while environmental management, etc.

Difficulties in the solving of this task are in the free initial conditions of the images formation of the object that is being defined, based on the different perspectives, aerial-mapping, heights of the aircraft, location and position of the object on the e-map image of the region, etc. This requires high calculational ability of the on-board ECS and imposes on the system strict requirements as to the reliability and rapid decision-making. That is why the task of the distinguishing of the range of interest, in which the object under consideration is expected to be present, acquires important meaning for the functional effectiveness of the on-board GIS. While the search of the vehicle such range of interest, first of all, should be considered highways of the region, and in case of the negative result it is necessary to pass to the recognition of the objects in other ranges of interest.

Thus, the relevant task while the region segmenting is the provision of the invariant nature of the final rules as

to the objects' position and location in the images of the regional e-map.

## II. LITERATURE REVIEW

The most widespread way of solving the problem of automatic segmenting of the region is application of the cluster-analysis methods [1, 2]. Constructed clusters are in future identified directly by the on-board GIS through comparison with the samples, stored in the database. In case of great number of clusters, the necessary speed of the objects' identification is gained through the application of the contemporary powerful on-board computers. In the same case, the accuracy of the clusterization depends much on the initial conditions of segmenting, that take place on practice. That is why one of the future-oriented ways of increasing of the accuracy of the region segmenting is application of the ideas and methods of computer-aided teaching and recognition of images. At this, during the intelligent data analysis the most widespread became the methods, based on the artificial neuron networks [3]. The main fault of the majority of the famous methods of Data Mining Technology, including artificial and immune networks, is connected with their sensitivity to the multidimensionality of the recognition features' vocabulary and recognition classes alphabet. Correction of this fault is especially relevant in the tasks of recognition of the non-standard, according to the brightness, images, that can be seen, for example, while recognition on the ground.

One of the promising approaches to the analysis and synthesis of on-board GIS of object recognition through the pre-segmenting e-map of the region is the usage of ideas and methods of information-extreme intelligent technology (IEI-technology), based on the maximization of the informational capacity of the system in the process of its computer-aided teaching [4, 5]. In the work [6] within IEI-technology the task of segmenting of the regional e-map by the on-board GIS of the unmanned aerial vehicle was studied. At this, built during the process of computer-aided teaching, final rules did not provide the high and full possibility of correct identification of images. The reason of this result is in that, that images were considered as texture and were refined according to the cartesian reference system. This

did not allow to provide the invariant nature of the final rules of the images identification, which contained, for example, free-oriented segments of the highway.

In this article the algorithm of information-extreme computer-aided teaching of on-board GIS according to the teaching matrixes, compiled during the refinement of the images of regional e-map in polar coordinates was studied.

### III. PROBLEM STATEMENT

Let us consider the formalised statement of the information-extreme computer-aided teaching of the on-board GIS, formed according to the results of refinement of images of the regional e-map. Let the regional e-map split into  $K$  images, among which the alphabet  $\{X_m^o | m = \overline{1, M}\}$  of the recognition classes is formed. To each class of recognition belong images that characterize different ranges of interest. For the given alphabet of recognition classes, the teaching matrix  $\|y_{m,i}^{(j)}\|$  of the brightness of the pixels of the images receptive field was formed. In the teaching matrix the line  $\{y_{m,i}^{(j)} | i = \overline{1, N}\}$ , where  $N$  – number of recognition features, is the realization vector of  $m$ -image, and the matrix column – accidental teaching selection  $\{y_{m,k,i}^{(j)} | j = \overline{1, n}\}$  with volume  $n$ . Besides, for each recognition class  $X_m^o$  the structural vector of the teaching settings was defined.

$$g_m = \langle \delta; d_m, x_m \rangle, \quad (1)$$

where  $\delta$  – characteristic of the symmetric field of the acceptable permissible errors for the recognition features;  $d_m$  – radius of the hyperspherical case of the recognition class  $X_m^o$ .

In the formula (1) the characteristic of the computer-aided teaching  $\delta$  is equal to the half of symmetrical field of the acceptable permissible errors for the recognition features

The range of values of the teaching characteristics have such limits:

– range of values of the  $\delta_i$  is defined by the equation

$$\delta_i < \delta_{H,i} / 2,$$

where  $\delta_{H,i}$  – standard field of permissible errors on the recognition  $i$ -characteristic, which defines the range of values of  $\delta_i$  characteristic;

– range of values of brightness of the pixels of the receptive field image is in the interval  $[0; 255]$  of brightness scale;

– range of values of the case radius of the recognition class  $X_m^o$  is defined by an inequation

$$d_m < d(x_m \oplus x_c),$$

where  $d(x_m \oplus x_c)$  – center-to-center distance between the realization-vector  $x_m \in X_m^o$  and averaged realization-vector  $x_c$  of the nearest class  $X_c^o$ .

During the process of computer-aided teaching of GIS it is necessary:

1) optimize the vector coordinates (1), which provide the maximum value of averaged according to the alphabet class of information criteria recognition of optimization of computer-aided teaching characteristics

$$\bar{E}^* = \frac{1}{M} \sum_{m=1}^M \max_{G_E \cap \{k\}} E_m^{(s)}, \quad (2)$$

where  $E_m^{(k)}$  – calculated in the  $k$ -step of teaching value of information criteria of optimization of computer-aided teaching characteristics to recognize the class  $X_m^o$  realizations;  $G_E$  – acceptable region of information criteria function;  $\{k\}$  – ordered set of steps of teaching;

2) to build final rules according to the optimum case geometric settings of recognition classes.

During the optimization of the settings of computer-aided teaching it is necessary to take into account that the search of global information criteria maximum (2) is done in the operating (feasible) range of values of its function. As the information criteria (2) is the composed function of accurate characteristics its operating range of definition of its function for two alternative decisions is characterized by the fidelity of the first and the second kind, which get the value higher than 0,5.

While operating of the on-board GIS in the test mode, to distinguish range of interest on the ground, it is necessary to identify images according to the final rules, settled on the computer-aided teaching stage.

Thus, the task of information synthesis of region segmenting by the able to learn on-board GIS is in the optimization of the settings of its computer-aided teaching by approximation of the global information criteria maximum (2) to its maximum limit value, which is calculated in the operating range of the function criteria definition.

### IV. MATERIALS AND METHODS

Let us consider the main stages of computer-aided teaching of the on-board GIS for identification of images of regional e-map in terms of IEI-technologies.

Input data for the information-extreme algorithm of computer-aided teaching of on-board GIS is the 3-D block of teaching matrix, elements of which characterize the light-intensity of the pixels of the receptive field. Besides the teaching matrix the input mathematical analysis of the able to learn on-board GIS, should include alphabet of recognition classes, which characterize various images of regional e-map and vocabulary of recognition features.

On the stage of formation of the input mathematical descriptions of the on-board GIS three classes of recognition were formed. They characterized the appropriate ranges of interest: class  $X_1^o$  - field, class  $X_2^o$  - forest,  $X_3^o$  - highway. The capacity of the vocabulary of the recognition features was defined by the size of the image receptive field of the regional e-map.

In the fig.1 the images of the region, received by the airborne prospecting, are presented.



Figure 1. Images of the region: a – field; b – forest; c – highway

In case images, presented in fig. 1a and 1b, can be related to the type “texture”, image, presented in fig. 1c, is unsteady in brightness.

To provide invariant nature of final rules to the shift and turn of the objects within the images the formation of the input teaching matrix was done by means of refinement of images, presented in the fig.1 in the polar coordinates. At this, the average brightness of pixels of each level of data reading, built around the image geometric center  $100 \times 100$  pixels in size, was calculated according to the formula [5]

$$\Theta_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \theta_i \quad (3)$$

where  $\Theta_j$  – averaged value of the pixels brightness, which are situated in the circle of reading of  $j$ -radius,  $j = \overline{0, R}$ ;  $\theta_i$  – value of the brightness of RGB-compound in the  $i$ -pixel of the image receptive field;  $N_j$  – total quantity of pixels in the  $j$ -circle of reading;  $R$  – radius of the reading circle.

In the fig.2 the input RGB-curves of the pixels’ brightness on the images of recognition class, received by the formula (3) are presented.  $X_3^o$  - is the part of the highway.

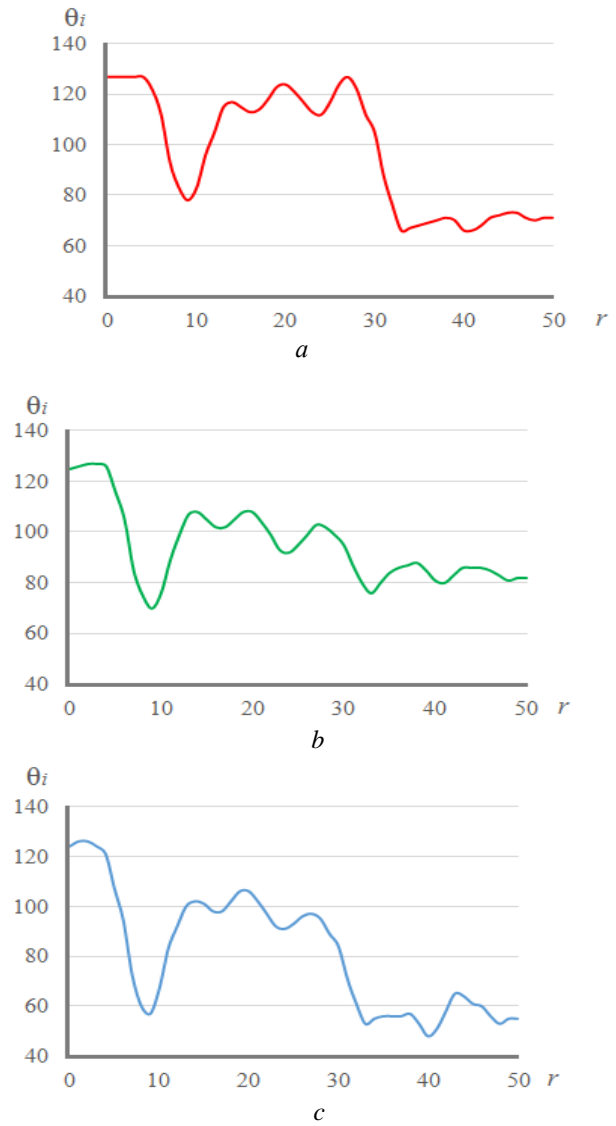


Figure 2. Curves of brightness of  $RGB$  - image compounds of  $X_3^o$  - class: a – red compound; b – green compound; c – blue compound

According to the received during the quantization by the reading radius step-width of the brightness  $RGB$ -curves the structures vectors-realizations of the input teaching matrix for the presented in fig.1 images were formed. Thus, the step-width of the quant  $RGB$ -curves were studied as the features of recognition. At this the structure of the realization vector consisted of consequence of recognition characteristics of all  $RGB$ -compounds of the images.

Algorithm of GIS teaching with the optimization of the system of permissible errors of the recognition features was realized as the dicyclic procedure of the search of global maximum of information criteria (2) in the operating sphere of distinguishing its function.

$$\delta^* = \arg \max_{G_\delta} \left[ \frac{1}{M} \sum_{m=1}^M \max_{G_{Em} \cap \{k\}} E_m^{(k)} \right], \quad (4)$$

where  $E_m^{(k)}$  – information criteria of optimization criteria to teach the system how to recognize the realization vectors of the class  $X_m^o$ ;  $G_\delta$  – allowed value area of  $\delta$  of the permissible errors field of the recognition features;  $G_{Em}$  – operating (allowed) area of distinguishing the criteria  $E_m^{(k)}$ ;  $\{k\}$  – number of teaching steps at which the radiuses of hyperspherical cases of recognition classes were changed.

As the information criteria of computer-aided teaching settings optimization the modified Kullback's measure was considered. For the two-decision problems with equally possible hypothesis it is the following:

$$E_m^{(k)} = [D_{1,m}^{(k)} - \beta_m^{(k)}] \times \log_2 \left[ \frac{1 + [D_{1,m}^{(k)} - \beta_m^{(k)}] + 10^{-r}}{1 - [D_{1,m}^{(k)} - \beta_m^{(k)}] + 10^{-r}} \right], \quad (5)$$

where  $D_{1,m}^{(k)}$  – the first authenticity, which characterizes the possibility of the right classification of the realization vector of the class  $X_m^o$ ;  $\beta_m^{(k)}$  – error of the second type, which characterize the false relation to the class  $X_m^o$ ; of the realization vector of the nearest class;  $10^{-r}$  – enough small number which is entered to avoid zero divide.

In the formula (5) the value  $r$  is equal to the number of signs of mantissa of the information criteria and in practice it is chosen in the interval

$$1 < r \leq 3.$$

According to the received during the computer-aided teaching optimal geometric settings of cases of recognition classes the final rules for the region images' identification during the GIS operating in the test mode were created. For the hypersphere cases of the recognition classes the final rules are as follows:

$$(\forall X_m^o \in \mathfrak{R}^{[M]})(x^{(j)} \in \mathfrak{R}^{[M]}) \{ \text{if } (\mu_m > 0) \& \\ \& \mu_m = \max(\mu_1, \dots, \mu_M) \} \text{ then } x^{(j)} \in X_m^o \}, \quad (6)$$

where  $x^{(j)}$  – realization-vector under recognition;  $\mu_m, \mu_c$  – functions of independence of realization vector under recognition from the neighboring (closest) cases of recognition classes  $X_m^o$  and  $X_c^o$  correspondently.

In case of hyperspherical cases of recognition classes functions of independence in formula (6) may consequently be the following

$$\mu_m = 1 - \frac{d(x^{(j)} \oplus x_m)}{d_m^*}; \quad \mu_c = 1 - \frac{d(x^{(j)} \oplus x_c)}{d_c^*},$$

where  $x_c$  – averaged realization vector of the closest neighboring recognition class  $X_c^o$ ;  $d_m^+, d_c^+$  – received during the computer-aided teaching optimal radiuses of the cases of the recognition classes  $X_m^o$  and  $X_c^o$  consequently.

Received in the test-mode results were compared to the results of the identification of the similar images, presented in the paper [6]. For example, according to the test results the full possibility of identification of the image of the  $X_3^o$ -class (part of the highway) was equal  $P_t = 0,96$  compared to  $P_t = 0,83$ , received in paper [6]. At this, the authenticity of the images of the other recognition classes almost did not differ, because their images refer to the type "texture".

## CONCLUSIONS

In paper it was offered the algorithm of information-extreme computer-aided teaching of the on-board GIS based on the teaching matrix, which were formed during the images refinement in the polar coordinates, which, in its turn, allows to provide the invariant nature of the final rules during the identification of images with free-oriented objects.

## REFERENCES:

- [1] A. S. Khashandarag, M. Mimia, and A. Sakhavati, "A New method for medical image clustering using genetic algorithm," *IJCSI International journal of computer science issues*, vol. 10, no. 1, pp. 381-393, 2013.
- [2] A. S. Silva, F. M. Q Severgnini, M. P Oliveira, V. M. S. Mendes, and Z. M. A. Peixoto, "Object Tracking by Color and Active Models Segmentation", *IEEE Latin America Transactions*, no 14(3), pp. 1488 – 1493, 2016.
- [3] S. Subbotin, "The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition", *Optical Memory and Neural Networks (Information Optics)*, vol. 22, no 2, pp. 97–103, 2013. DOI: 10.3103/s1060992x13020082
- [4] A. S. Dovbysh, V. V. Moskalenko, and A. S. Rizhova, "Information-Extreme Method for Classification of Observations with Categorical Attributes", *Cybernetics and Systems Analysis*, vol. 52 (2), pp. 224-231, 2916. DOI: 10.1007/s10559-016-9818-1
- [5] A. S. Dovbysh, S. S. Martynenko, A. S. Kovalenko, and N. N. Budnyk, "Information-extreme algorithm for recognizing current distribution maps in magnetocardiography", *Journal of Automation and Information Sciences*, vol. 43, issue 2, pp. 63-70. DOI: 10.1615/JAutomatInfScien.v43.i2.60
- [6] O. S. Kuzenko, L. B. Kascheev, M. I. Mironenko, "Geo-informational system of identification of the frames for area reconstruction", *Bulletin of NTU "KhPI". Series: System analysis, control and information technology.* – Kharkiv : NTU "KhPI", no. 46, pp. 2017.

**SESSION 6**  
**COMPUTER NETWORKING AND**  
**TELECOMMUNICATIONS**

# Detection and Preventing Leaks of Sensitive Data on Computer Systems

Mihail Babiy

Sumy State University, Ukraine, m.babiy@cs.sumdu.edu.ua

**Abstract – The peculiarity of transmitted sensitive data is their transformation preserving the order of elements. The approach based on the sequence alignment and taking into consideration ordering is presented. Recurrence relation, complexity, and recommendation for using algorithm are described.**

**Keywords – data leak, sequence alignment, dynamic programming.**

## I. INTRODUCTION

The leak of sensitive data on computer systems presents a serious threat to the security of the corporate IT infrastructure. Statistics show that the lack of proper encryption on the files, folders, and communications due to human errors is one of the causes of data loss.

Kaspersky Lab and B2B International conducted the regular surveys of IT professionals from different companies around the world. According to the results of the study, 30% respondents note accidental leaks or sharing of data by staff [1].

First group of sensitive information is the personal identifying information. It refers to documents, containing passport data, credit cards numbers, PIN codes, e-mail address, Internet accounts and passwords, biometric data, fingerprints. Second group is protected health information of individual. It refers to documents with names, geographical addresses, dates, phone and fax numbers, medical records, vehicle identifiers and numbers, Web URLs, IP-addresses, biometric identifiers. Third group is employed data, including financial information. Nonpublic information such as contracts, patents is also the sensitive information.

Existing approaches to data leak detection are based on finding the intersection of two sets. Sets consist of substrings of length  $n$  ( $n$ -grams). One set is selected from the sequence of transmitted data, the other is taken from the sensitive data samples. The intersection of sets gives the number of matching  $n$ -grams, which allows you to determine the degree of similarity of the two data sequences.

A variant of this method uses the Rabin fingerprints. With the sliding window, the sequence is converted to a set of  $n$ -grams, which are also called shingles. The set of all shingles of the size  $w$  is called  $w$ -shingling. Instead of working directly with shingles, it's more convenient to match each shingle to a unique numeric identifier, which

is called a fingerprint. From a computational point of view, it is convenient to use Rabin fingerprints based on polynomial arithmetic. The applications of this approach are described in [2-5].

The peculiarity of the transmitted sensitive data is their transformation. User can do insertion, deletion, shortening of text. Applications can add formatting tags, metadata. Under these conditions, the accuracy of data leak detection is much higher if we can detect the ordered data. Set intersection does not allow this.

## II. DATA LEAK DETECTION MODEL

The method proposed in this research is able to work with ordered sequences. We work with two types of sequences: network traffic sequence and sensitive data sequence.

*Network traffic sequence* on the way from source to destination passes through access devices: a router (hardware, or PC-based), a switch, a VPN server, or a hub. Packet capture can be performed directly on the access device. After this, packets are transmitted to a separate server. In particular, TAP (Test Access Point) can be used. This is a hardware device that connects directly to the computer network cable and creates a copy of the network traffic for transmission to another device. Many enterprise switches copy the activity of ports through a Switch Port Analyzer (SPAN) port, called the mirror port. To access network traffic, an analysis device must be attached to the mirror port.

In the simplest case, if we have a PC-based router with OS Linux, we can use two methods of capturing packets passing through a network card. First method is capturing packets with the *libpcap* library. A copy of the package, passing through the interface, can be requested by a client program written using this library. There is a *libpcap* implementation under Windows. Second method of capturing packets is the passing through the built-in *iptables* firewall. Administrator installs a rule in the firewall that sends traffic to the inside of the handler. For each passing packet, the handler receives its contents as a memory buffer.

*Sensitive data sequence* contains the privileged or proprietary information that only certain people are allowed to see. This information needs to be protected and cannot be exposed to outer world. The sensitive data sequences must be known to the analysis system.



In this work, we are dealing with detecting unintentional data leaks. Data leak can occur when a user publishes sensitive data on the Internet or transfers them to an FTP server. In addition, many malicious program and sniffers can receive sensitive data.

We assume the content in network traffic is available to the analyzing system. A network channel is unencrypted. If channel is encrypted, the content in it can be extracted for analysis.

### III. SOLUTION OF ALIGNMENT TASK

We first give the following definitions. String is continuous finite sequence of symbols (bytes) taken from file or data flow. The symbol set is not limited to alphanumeric characters. The empty string is the string with zero occurrences of symbols. Length of a string is a number of symbols in the string. Standard notation for the length of the string  $W$  is  $|W|$ . A substring is a consecutive segment of the initial string. A prefix of string is any string obtained from initial string by removing symbols from the end of string. Respectively, a suffix of string is any string obtained from initial string by removing symbols from the beginning of string. Subsequence of a string is a string obtained by deleting zero or more symbols, which need not be contiguous.

Sequence alignment is most commonly used in bioinformatics. The method is based on placing two or more sequences of DNA or protein monomers under each other in such a way that it is easy to see similar sections in these sequences. The similarity of the primary structures of two molecules can reflect their functional or structural interrelationships. The aligned sequences are generally represented as matrix rows. Gaps between the sections are added in such a way that the same or similar elements are located in columns of the matrix.

Sample:

Sequence 1      **TMYE**CN**HECH**P**T**  
 Sequence 2      ---**YET**Y---**CH**S**K**

Pair-wise alignment is used to find similar sections of two sequences. There are global and local alignments. Global alignment assumes that the sequences have the same structure along the entire length. Local alignment is used if the sequences contain both related and unrelated regions. The result of the local alignment is the selection of the region in each of the sequences and the alignment between these regions.

To solve the alignment problem, we will use the methods of dynamic programming.

Our task is to find the degree of similarity between the samples of sensitive information and fragments of the transmitted information. The presence of spaces between regions, containing sensitive information, in the transmitted stream is natural and does not eliminate the fact that there is a leak of this information. Therefore, the lengths of gaps in estimating the degree of similarity can be ignored.

In this case, the similarity degree is best determined by the length of the longest sequence, common to the two initial sequences.

Suppose that there are two sequences:  $X=(x_1, x_2...x_m)$  and  $Y=(y_1, y_2...y_n)$ . Here  $x_1$  and  $y_1$  are symbols. We denote the prefixes of  $X$  by  $X_1, X_2 \dots X_m$ , and prefixes of  $Y$  by  $Y_1, Y_2 \dots Y_n$ . Let  $\mathbf{S}(X_i, Y_j)$  represent the set of longest coincidental subsequences  $S$  of prefixes  $X_i$  and  $Y_j$ .

The finding of  $\mathbf{S}$  relies on two properties.

*Property 1.* Suppose that there are two subsequences  $X_i$  and  $Y_j$  that end in the same symbol. To find  $\mathbf{S}(X_i, Y_j)$ , we need to remove the last character from each sequence  $S$  entering  $\mathbf{S}$ , find  $\mathbf{S}(X_{i-1}, Y_{j-1})$  for these truncated sequences, and add a remote character to each  $S(X_{i-1}, Y_{j-1})$ .

*Property 2.* Suppose that there are two subsequences  $X_i$  and  $Y_j$  that do not end in the same symbol. Then  $\mathbf{S}(X_i, Y_j)$  is the set of the longest subsequences, included in union of  $\mathbf{S}(X_i, Y_{j-1})$  and  $\mathbf{S}(X_{i-1}, Y_j)$ .

Corresponding recurrence relations in dynamic programming are following:

$$\begin{aligned} \text{if } i=0 \text{ or } j=0 \text{ then } \mathbf{S}(X_i, Y_j) &= \emptyset; \\ \text{if } x_i = y_i \text{ then } \mathbf{S}(X_i, Y_j) &= \{S \mid S = \\ &= \text{Concat}(S(X_{i-1}, Y_{j-1}), x_i)\}; \\ \text{if } x_i \neq y_i \text{ then } \mathbf{S}(X_i, Y_j) &= \{S \mid S = \\ &= \text{Longest}(\mathbf{S}(X_i, Y_{j-1}) \cup \mathbf{S}(X_{i-1}, Y_j))\}. \end{aligned}$$

Here, function *Concat* performs concatenation of two strings, function *Longest* finds the longest string in the set. There can be several such lines.

For example, we will find the longest coincidental subsequence  $S$  for sequences  $X=abcd$  and  $Y=bac$ . Because the recurrence relations use the empty set, it is convenient to define zero prefixes for our sequences:  $X_0 = \emptyset$ ; and  $Y_0 = \emptyset$ .

Process of the solution is shown in Table 1.

TABLE I. PROCESS OF THE SOLUTION

	$\emptyset$	a	b	c	a	d
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
b	$\emptyset$	$\emptyset$	(b)	(b)	(b)	(b)
a	$\emptyset$	(a)	(a) (b)	(a) (b)	(ba)	(ba)
c	$\emptyset$	(a)	(a) (b)	(ac) (bc)	(ac) (bc) (ba)	(ac) (bc) (ba)

The second row and second column in the table are filled with  $\emptyset$ . Other cells contain the longest subsequences, obtained for each step of recurrence calculation. The right-bottom cell contains the final result:  $S=\{ac, bc, ba\}$ .

The main objective of our work is not so much to find actual coincidental subsequence, as the length of this subsequence. In this case, we can simplify our recurrence relations. For this purpose, we define function  $f$  that returns the length of longest coincidental subsequence.

Corresponding recurrence relations are following:

if  $i=0$  or  $j=0$  then  $f(i, j) = 0$ ;

if  $x_i = y_j$  then  $f(i, j) = f(i-1, j-1) + 1$ ;

if  $x_i \neq y_j$  then  $f(i, j) = \max(f(i, j-1), f(i-1, j))$

Simplified process of the solution is shown in Table 2.

TABLE II. SIMPLIFIED PROCESS OF THE SOLUTION

	$\emptyset$	a	b	c	a	d
$\emptyset$	0	0	0	0	0	0
b	0	0	1	1	1	1
a	0	1	1	1	2	2
c	0	1	1	2	2	2

The actual subsequence can be restored in backward procedure. For this, in the calculation process, we need to save not only the current length of the subsequence, but also the directions, moving along which we can obtain the desired subsequence. Process with the storing of the directions is shown in the table 3.

TABLE III. BACKWARD PROCEDURE

	$\emptyset$	a	b	c	a	d
$\emptyset$	0	0	0	0	0	0
b	0	0	1	-1	-1	-1
a	0	1	-1	-1	2	-2
c	0	1	-1	2	-2	-2

Several paths are possible when two lines are presented in the cell. To restore the longest coincidental

subsequences  $S$ , we start from right-bottom cell in the table. The movement is possible to the left, up and left-up diagonally. Possible paths are marked in the table III with the numbers in bold style.

#### IV. IMPLEMENTATION OF THE ALGORITHM

The searching sensitive data in network traffic sequences is performed as follows. Sensitive data are divided into sequences of identical length. These sequences are compared with sequences from network traffic.

The complexity of the suggested algorithm is  $O(|A| \times |B|)$ , where  $|A|$  and  $|B|$  are lengths of compared sequences. If the sequences will be divided to segments of length  $|A|/m$  and  $|B|/n$ , then number of operation is  $m \times n \times (|A|/m) \times (|B|/n) = |A| \times |B|$ . Therefore, complexity is independent from size of compared sequences. But with short sequences, the main advantage of the algorithm is lost: the ordering of sensitive fragments. Therefore, it is desirable to choose a string size not less than the standard 80 characters.

Increasing productivity can be achieved by placing a sequence comparison in individual threads. More significant performance improvements can be achieved by running parallel comparisons on individual processor cores.

For better security, source data with sensitive information should be provided in an encrypted form. Decryption should be performed already in the program itself.

#### CONCLUSIONS

Detection and preventing leaks of sensitive information is an important problem. The peculiarity of transmitted sensitive data is their transformation preserving the order of elements. The existing methods of sequences comparison are based on the set intersection, and do not analyze ordering. We presented algorithm based on the sequence alignment, and taking into consideration ordering. Recurrence relation, complexity, and recommendation for using algorithm are described.

#### REFERENCES:

- [1] Kaspersky Lab. (2014). *GlobalCorporate IT Security Risk*. [Online]. Available: [http://media.kaspersky.com/en/business-security/Kaspersky\\_Global\\_IT\\_Security\\_Risks\\_Survey\\_Eng\\_final.pdf](http://media.kaspersky.com/en/business-security/Kaspersky_Global_IT_Security_Risks_Survey_Eng_final.pdf). [Accessed March. 10, 2018].
- [2] A. Z. Broder, "Identifying and filtering near-duplicate documents," in *Proc. 11th Annu. Symp. Combinat. Pattern Matching*, 2000, pp.1–10
- [3] K. Li, Z. Zhong, and L. Ramaswamy, "Privacy-aware collaborative spam filtering," *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no.5, pp. 725–739, May 2009.
- [4] X. Shu, D. Yao, and E. Bertino. Privacy-Preserving Detection of Sensitive Data Exposure. *IEEE Transactions on Information Forensics and Security*, Vol. 10, No. 5, May 2015
- [5] L. Ulahannan and R. Jose. Fuzzy Fingerprint Method for Detection of Sensitive Data Exposure *International Journal on Cybernetics & Informatics (IJCI)*, vol. 5, no. 2, April 2016.

# Crowdsourced Measurement: Realisation, Privacy and Security in the Frame of Sumy State University Wireless Network Quality Analyse

Borys Kuzikov<sup>1</sup>, Sergey Panchenko<sup>2</sup>

Sumy State University, Ukraine, <sup>1</sup>b.kuzikov@dl.sumdu.edu.ua, <sup>2</sup>serzhik.panchenko@gmail.com

**Abstract** – Article describes implementing of crowd-based measurement for quality assurance of wireless network in Sumy State University. Described structure and algorithm of work of client-side android application and web-based backend, discussed problems of privacy and security. Building on the on analyse of achieved results proposed a way of enactment wireless network of the university.

**Keywords** – crowdsourced measurement, wireless network, quality assurance, security, privacy, Android, support of education.

## I. INTRODUCTION

Ensuring high-quality education in the IT specialties requires the involvement of a modern material and technical base. However, the constant renewal of the hardware of the university is complicated due to financial reasons. The experience of other domestic and foreign universities indicates the possibility of solving this problem through the involvement of students' own devices in the educational process. Undoubtedly, this approach should be directed by the policy of the university, because it requires the preparation of the appropriate infrastructure, i.e. additional sockets, high-quality network connection.

Also widespread among the students of portable computers and multifunctional communicative electronic devices makes it urgent to create a flexible working environment on the campus. The main feature of such an environment is to provide access at any time and from any place to the information resources of the corporate network of the university and global information networks [1]. Such integration requires qualitative and safe access to the network, which in most cases is implemented through wireless networks. The practice of their application in Sumy State University faced numerous difficulties. For example, a weak network signal in a classroom, the inability to connect to an open wireless access point (“wi-fi point”, WAP) with standard settings, the problem with internet connection.

The problem with the quality of wireless network at Sumy State University (SSU) can be solved in several ways. One of a solution – increase the count of WAP. World Health Organization experts recommend refraining

from using the wireless networks at educational institutions, as electromagnetic radiation creates an additional burden on the body [6]. The other way – optimise usage of existing points, to maintain existing ones in the proper working condition since Sumy State University. In this case, there is a possibility to decrease count of points and reduce the negative impact on students' health and teachers of the educational institution. Such approach requires a tool for monitoring and analysing of wireless network quality and WAP usages.

There three possibilities to measure the quality of network connection in such situation:

1. *Centralized measure network bandwidth usage by WAP on university internet gate or billing system.* This approach is easy to implement. However, it is able to show only a few kinds of problem, such as all the user unable to connect to some WAP.
2. *Measure number of clients directly on WAP.* This approach can show real usage of each point. The lack of this approach: in first we are unable to detect situation, when clients are unable to connect to WAP. On the other hand, there are many types of equipment. So, we need to implement firmware level monitoring tool for different of equipment or scrap information from web-based admin interface (or telnet-based command tool). Also, this possibility stands for security problem: we need admin-level access to all WAP. So, this approach has many disadvantages and is hard to implement.
3. *The crowdsourced measuring with many wi-fi clients.* Such types of measurement help to solve a wide range of the problem, for, i.e., detect potholes in road-system using Android smartphones with accelerometers [2]. Such type possibility especially useful in the case of geographically distributed objects of measurement.

The last approach can help us to solve the initial problem – detect state and quality of wi-fi-based network connection in the frame of SSU campus. In the article, we should describe existed software, and show our solution to this problem. At the end of the article, we should

discuss the problem of privacy and security of this approach.

## II. EXISTED SOFTWARE

There are many different ready-made software solutions for monitoring Wi-Fi networks. However, these are general-purpose, mostly not free of charge software. Some are inconvenient to use or have specific conditions of use. For example, the need for a building map and semi-automatic monitoring in several points of the map each time. Let's introduce some program from this class.

*NetSpot*. The primary function of this application is an analysis of wireless Wi-Fi networks inside and outside buildings and bind results to map. To analyse network, customer need to walk through the building and mark a place of measurement on a map. There is an only windows version of the software. The application can show a level of wi-fi signal on a building plan, but unable to check a quality of the connection, possibility to connect to WAP and internet connection via WAP [3].

*WiFiAnalyzer* is opensource software for Android. [4] It helps to optimise wireless network by examining surrounding Wi-Fi networks, measuring their signal strength as well as identifying crowded channels. Basically, very useful for collecting information about networks at some point. Unfortunately, this software unable to share information between several devices or store it externally. So, it very hard to keep up-to-date an analyse information if we have several widely distributed WAPs.

PRTG - professional WiFi analyser [5]. This software analyses every aspect of the wireless network: devices, load, traffic, availability, signal strengths. PRTG comes all-in, no extra costs, no add-ons necessary. The main disadvantages of that software are the high cost of a license - 1200 euros per year.

The list of software for wireless network analysis can be expanded. The main conclusion of this investigation is: there is lack of software, which able to:

- investigate different aspects of WAP (such as signal level in classrooms and ability to establish a connection in a different period);
- gather information from several devices;
- present information in a form applicable to analyse and forecasting;
- has a reasonable price.

Using the positive features of the studied software, we have built information system, uses the client-side android application to perform measurement and web-based backend to collect and visualise information. Let's describe both parts of the system.

## III. CLIENT-SIDE SOFTWARE

SSU has numerous building with multiple WAP in each. In this case, we need multiple devices to gather

statistics. Most easily way – involve users mobile devices in this process. Such periodic independent mass scanning will allow the staff of the university's Internet service to have the latest information about the Wi-Fi points of the institution.

Therefore, for the problem monitoring, the android application WifiChecker was developed. The purpose of the application is the semi-automatic collection of information about available Wi-Fi networks and the ability to check access the Internet via visible open access points. The application is available for devices running Android 7.0 (API level 24) and above.

The algorithm of application, including interaction with the central server, is described as UML interaction diagram on figure 1.

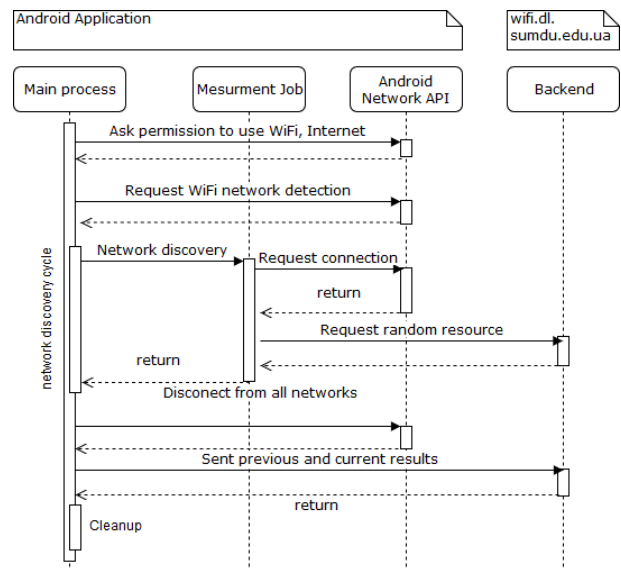


Figure 1. Interaction diagram of the main cycle of application

Application use state-machine pattern with act and rollback action between states. Main states are:

- Initialization – ask all necessary permission from the user. By restriction of Android, some of the permission can be obtained only dynamically.
- Network scanning – application subscribe to network scanning action and collect a list of networks till scanning process bring new networks.
- Network discovery – for each open (accessible without key) network application put network discovery job into single treading queue.
- Sending results – application disconnect from all networks and try to send old (if anyone) and current results via default network (in most cases it is cellular network). Each result is submitted within own job. Application removes a date from internal storage only if is successful scented to the server. After processing all pieces of information

application show button “rollback”, which act moving to the initialisation state.

Figure 2 show application in the state “Network discovery”. The main screen of allocation shows the name of WAP and level of the signal, i.e. WAP “wifi.sumdu.edu.ua” moderate level of the signal, “wifi-60” – low level. The level of the signal of secured WAP does not matter – we unable to connect to them. The exact value of the signal level visible on results analyses on the server. Also, the application shows a state of checking: “connected” (WAP accessible and provide access to the internet, i.e. “wifi.sumdu.edu.ua”), “connection failure” (i.e. “wifi.sumdu.edu.ua-5”), “connection [on this WAP is in] testing” and “pending [for the testing]”.

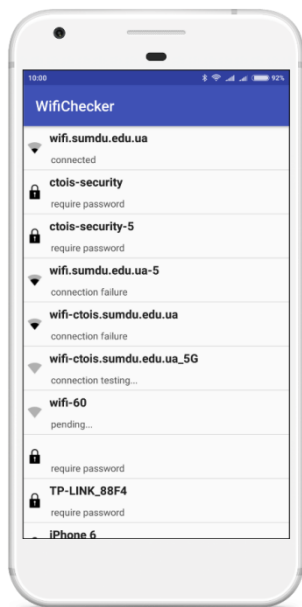


Figure 2. Android application “WiFi Checker.”

Each network discovery job performs next steps:

1. Disconnect from the previous network.
  2. Try to find a job-related network in android network registry or create a new one.
  3. Try to connect to this network.
  4. Wait for “connecting” and then “connected” state of Android Network subsystem. In most cases, for lousy WAP it takes much time. We limit this step to 120 seconds.
- Try to detect if current network equal to desired network (Android network subsystem able choose another network dynamically). Unfortunately, on this step, we able detect current connection network only by SSID, not by BSSID. SSID name of different WAP can be the same for example if they work in roaming mode.
5. Bind connection to the current network (by default Android able to change default network as a fall-back for connection without internet access).

6. Try to fetch the random resource from <https://wifi.dl.sumdu.edu.ua/r/<resource name>>. On this step, we wait for an HTTP-response with code 200 and content equal to the resource name. By this step, we prevent caching resource and wrong-true internet accessibility detection in case of the smart screen (like 404-page on LifeCell, or billing-authentication page of some providers).

When the application finishes scanning all the points, the resulting scan results are mapped to the JSON object and send them to <https://wifi.dl.sumdu.edu.ua>. From that website, information is accessible for monitoring and further analysis. If the current data cannot be sent, application store them at own persistent storage. This information will be resent, after next network discovery, prior sending a new collection of data.

#### IV. BACKEND

The web-based part of the product is created with Ruby-based framework Sinatra. For today, we did not use any database and store user result directly to file system. In future, this approach will be replaced with SQLite RDBMS. The application provides next pages and endpoints:

- GET '/' – show general information about service and common statistic by all public WAP. All evidence a grouped by status: success – “connected” or negative – “connection failure”, “[user] device error”, “internet failure”, “require password” (somewhen WAP was open, but not it requires access key). In this view, all WAP showed by its SSID, but group by BSSID which is more stable and allow differentiate points in roaming-mode (see figure 3).

The screenshot shows the main page of the web-application, titled 'SSU WiFi stats'. It displays two large numbers: 107 (наблюдений) and 6 (устройств). Below these is a table with the following data:

#	Name	require password	connection failure	connected	device error	internet failure
1	wifi.sumdu.edu.ua-5	-	10	-	-	-
2	wifi.sumdu.edu.ua	-	34	5	-	1
3	wifi-60	-	32	-	7	-
4	wifi-142	48	1	-	-	-
5	wifi-ctoiss.sumdu.edu.ua	-	19	33	2	-

Figure 3. Web-application main page example

- GET '/wifi/:bssid' display detailed information on all evidence with some WAP (time, signal level, status). Also, it shows information grouped by day of the week and fraction of a day (see section “Results”). the

- POST '/wifi' – endpoint for gathering information provided on state “Sending results” of the client application;
- POST '/t/random\_seed' – endpoint for checking internet connection, provide cache-free echo for client application (need on step 7 of network discovery job);
- GET '/archive' – provide a Zip archive with raw information, used for detail analyse and practice “Data migration” of course “Database and Information System”;
- POST '/deploy/secret' – web-hook endpoint for deployment (see section “deployment”).

## V. DEPLOYMENT

An information system has two parts: client Android-based application and web-based application. In general, the structure of system and deployment process are shown in figure 4.

The source code of the Android application is storied within public GitHub repository [7]. The repository is linked with Circle CI to automatically build an application on code update event and publish build results on GitHub Releases page of a project. Anybody able to make push-request, and after review got the new version of the application as soon as possible.

Source code on web-application is storied within private repository at Bitbucket. In the same way, after each commit, Bitbucket make the web-hook call of deploy endpoint of application. Application reloads own source code from Bitbucket and restarts itself.

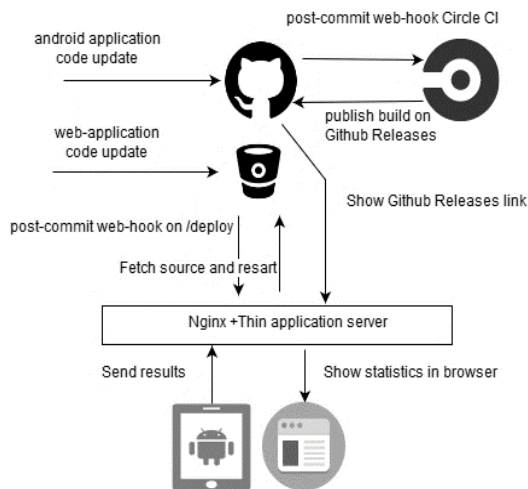


Figure 4. Common project structure and deployment process

## VI. CROWN-BASED MEASURING AND PRIVACY

After the publishing European union regalement EC2016/679 on General Data Protection Regulation (GDPR) [8] all public service which able contact with European Union citizen or residents should provide in a clear way all information about the usage of user’s private

data. Our application story only application-generated unique identification (InstanceID) for each evidence. InstanceID used to count of the involved device and detect devices, which sends suspicious results. In conjunction with time on measurement and list of visible WAP it able to de-classify location of the client but hide any other information about the client. The value Instance ID is anomy’s, and automatically changed on application reinstall or upgrade [9]. According to google recommendation it provides the necessary balance between user-identification and user-privacy.

As for Android application, we need permission for ACCESS\_FINE\_LOCATION, ACCESS\_WIFI\_STATE, CHANGE\_WIFI\_STATE, ACCESS\_NETWORK\_STATE, INTERNET, OVERRIDE\_WIFI\_CONFIG. Such wide list of permission caused fact, which we can de-classify user location upon WAPs signal level. For example, is approach is used for LBS, and similar for A-GPS navigation.

In our situation, the best way will hide information of recent measurement from all statistics. It will be implemented in the feature.

## VII. MEASURING AND SECURITY

Let’s analyse possible security risk. As any other distributed application our system has a lot points for invasion. Most serious are:

- GitHub and Bitbucket accounts. Any changes at this point automatically rebuild part of a system. So, we must use strong passwords and two-phase authentication. The risk on this point a partially smooth out by version control system – all changes are storied and able to be reviewed.
- Circle CI is used for automatically build the Android application. So, we need strong password there too. The risk on this point is partially smooth out by nature of building process – all files and scripts necessary to build the application are uploading each time from GitHub repository. Circle CI has read-only access to this source.
- Application and web server can be perpetrated directly or of from other services on the same machine. At this point, we should use same security policy as for any other web-services. The better way is move service to separate virtual machine or run it into the container for better security.

*User-generated data.* Far-famed security rule says that all user-generated date must be sanitised. In our case there is a possible way to attack service:

- Non-JSON or corrupted JSON workload is checked by an application on store new results.
- The extra-large user-generated response is checked by Nginx web-server with client\_max\_body\_size directive (few kilobytes in our case).

- Fake measurement. We did not provide any protection for this case. Provide such type of protection is on our roadmap. We plan to implement InstanceID-based results signing.

## VIII. INFORMATION ANALYSES

According to server statistics and results of measurement, most of analysed WAP is located at building “C” of Sumy State University. Web-application able show generalised statistic by all points and detailed statistic by each of them. The generalised statistic shows each measurement status (“Connected [to the internet via WAP]” or type of problem). Detailed statistic shows time and level of signal for each measurement and group measurement status by day of week and day fraction (before 8.15 – early morning, 8.15-12.45 – first part of day, 12:45–13:25 – the «big break» between pairs, 13:25–15:00 – 4th pair, 15:00–18:00 – evening, 18:00–24:00 – later evening).

Figure 5 show example of WAP details-page. As we can see on figure WAP “wifi-ctois.sumdu.edu.ua” accessible only in half attempts in the morning and almost all-time in a second part of the day. The symbol “-” in the picture means lack of measurement.

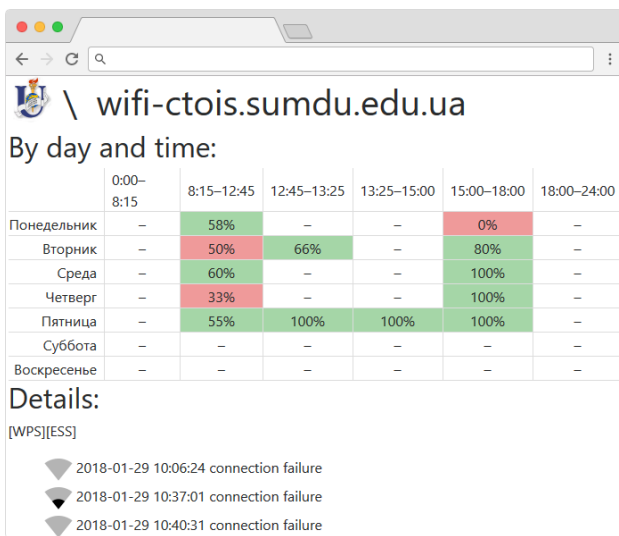


Figure 5. WAP details

The results of measurement show the disappointing state of the network – there is much WAP inside the campus, some of them are open, but most of them are inaccessible most of the time.

Average count of WAP on each network scanning is 5.6 (on the result of 109 evidence), the maximum count – 12. So, increasing amount of WAP will not solve a problem but decrease the quality of the signal.

## CONCLUSIONS

Modern education requires access to different resources via local network or internet. Partially this

requirement is covered by wireless network. The wireless network provides quality service, which means, at least, be accessible and provide access to the internet. Unfortunately, the wireless network at SSU does not meet this requirement. Before any optimisation and reconfiguration, we need a tool to analyse network quality. Exploring of ready-made software show lack of variants, that fit all our requirements. To solve a problem was developed an information system, which consists of a client-side android application to perform measurement and web-based application for gathering, visualisation and analyse information about the state of WAPs.

Analyse show unsatisfactory result – there is a lot of WAPs at the campus, some of them are open, but most of them are inaccessible. The different ratio of success/fail tests on same points during a day means the insufficient capacity of the wireless network or problem with WAPs configuration.

There are several ways to improve the situation without buying additional equipment. There are lot of secured WAP. Almost all of them use internet connection provided by university network, so they should be opened (or has open wlan with limited bandwidth). The second possibility – organise single university-level wlan with wifi-roaming [12]. With is approach we able increase network capacity by client balancing between points.

## REFERENCES:

- [1] P.V.Kuchernyuk, E.M.Melnyk “Wireless access to WI-FI as part of the university's telecommunication network” [Online], Available: <http://kpi.ua/809-7> [Accessed Apr. 23, 2018]
- [2] Artis Mednis, Girts Strazdins, Reinholds Zviedris, Georgijs Kanonirs and Leo Selavo “Real time pothole detection using Android smartphones with accelerometers” In Proc. 2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011, pp. 1-6.
- [3] Monitor Wi-Fi Signal and Map It With NetSpot [Software] Available: <https://www.netspotapp.com> [Accessed Apr. 23, 2018]
- [4] VREM Software Development WiFiAnalyzer [Mobile application software] Available: <https://play.google.com/store/apps/details?id=com.vrem.wifianalyzer&hl=uk> [Accessed Apr. 23, 2018]
- [5] PRTG: Professional WIFI analyzer [Software] Available: <https://www.paessler.com/wifi-monitoring> [Accessed Apr. 23, 2018]
- [6] Vladimir Fediev “Influence of electromagnetic fields (mobile phones, Wi-Fi networks) on human health”, *bsmu.edu.ua* [Online] Available: <https://www.bsmu.edu.ua/uk/news/digest/1930-vplyv-electromagnitnyh-poliv> [Accessed Apr. 23, 2018]
- [7] Kuzikov Borys WifiChecker Android application [Software] Available: <https://github.com/potapuff/WifiChecker> [Accessed Apr. 23, 2018]
- [8] EU GDPR Portal. [Online] Available: <https://www.eugdpr.org> [Accessed Apr. 23, 2018]
- [9] Google developers portal “What is Instance ID?”, *Google*. [Online] Available: <https://developers.google.com/instance-id/> [Accessed Apr. 23, 2018]
- [10] Anton Vinokurov “Seamless Roaming Wi-Fi”, *habrahabr.ru*, 2013. [Online] Available: <https://habrahabr.ru/post/185138/> [Accessed Apr. 23, 2018]

**SESSION 7**  
**MODERN METHODS AND INFORMATION**  
**TECHNOLOGIES OF SUSTAINABLE**  
**DEVELOPMENT**



# Investigation of the Influence of Information Management on the Development of the Country

Vitaliya Koibichuk

Sumy State University, Ukraine, v.koibichuk@uabs.sumdu.edu.ua

**Abstract** –The article deals with the aspects of information management in Ukraine and Poland. Every year, besides of consolidating its positions in the economy, the country needs to develop information management, which is essentially a business card of the state on the international scene as well as the primary source for information in various life's spheres.

**Keywords** – *information fund; information management, information technologies.*

## I. INTRODUCTION

One of the main driving forces that make an influence on the revolutionary innovation changes in the methods of doing business is the rapid development of information technology, the quality and effectiveness of their application, the development of information management.

The profit of any socio-economic object (enterprise, firm, bank, financial institution) is directly proportional to: the ability to organize and carry out information activities; organization of information service of the authorities of any level; development of the main contours of the organization's policy and strategy in the field of information technology; evaluation of the level of information of an object; project management for the creation and implementation of information and technological solutions of any level of complexity; management of innovations and innovative activity in the field of information technologies; Identification of reserves for increasing the efficiency of information activities of the enterprise.

We highlight a number of the most important tasks of information management: 1) study the peculiarities of the use of information resources for the adoption of effective managerial decisions; 2) the use of new information technologies as a system of holistic interconnected techniques of methods and tools for the processing of information and the implementation of communications in organizational management in order to increase the efficiency of decision-making processes; 3) studying the peculiarities of communicative relations in the organization and improvement of document circulation; 4) study of the peculiarities of the use of Internet technologies in management activities; 5) management of the existing information resources of the institution, enterprise, organization. The importance

of information management (MI) for the state will be considered in a systematic approach that addresses strategic, administrative and operational objectives [1]. Among them, one can identify the main areas of MI, namely: the development of information infrastructure, technology management, data management, development of the system concept of the organization.

Depending on the type of task, these areas have different levels of detail and scope of information. At the level of public administration, a successful solution to these problems is possible only with the development and implementation of the systemic concept of management organization.

## II. RESEARCH RESULTS

Information management, in the context of this study, is the management of the processes of implementation and use of information technology in the activities of the government.

As an example, let's look at the government sites of Ukraine and Poland. Comparing macroeconomic indicators (GDP per capita, GDP and GDP share in different spheres), Poland and Ukraine in the period after the 2008 global crisis, one can see a healthy economy (for example, Poland), and state the Ukrainian illness.

Before the crisis, the Ukrainian and Polish economies had similar dynamics. However, recent years have been years of success for Poland and have been lost for Ukraine. The driving force of any country is its attractiveness for foreign investors. After joining the EU, the conditions for doing business in Poland have become more favorable than in Ukraine. For example, Poland in 2016 ranked 25th among 189 analyzed countries, while Ukraine is 83 [2].

The process of transition from a planned to a market economy was difficult for some states, which arose after the collapse of the socialist bloc of countries, because not all of them, even for a quarter century after the collapse of the planned economy, were able to enter the market. Poland have joined after successful market reforms and the privatization of state-owned enterprises in countries with market economies.

On the end of 2016 Ukraine was slowly continuing to make market reforms and privatization, but this did not contribute to improving its macroeconomic

performance. In 2016 Ukraine was on the 63 position in the world in terms of the gross domestic product (GDP) and Poland was on the 24th positions [3,4].

By the volume of the gross domestic product (GDP) per capita in 2016 Ukraine ranked 107th in the world, while Poland was in the 45 position. GDP per capita in Ukraine was 7.91 thousand dollars and in Poland 26.13 thousand dollars [3,4].

GDP per capita in 1991 in these countries was almost at the same level: in Ukraine – 5,9 thousand dollars, 6,00 thousand dollars – in Poland.

According to the World Bank, at the time of the collapse of the "socialist bloc of countries" Poland and Ukraine had similar economy: a similar volume of GDP per capita, as well as similar shares of the agrarian sector, industry and services in GDP.

In 2004, at the time of Poland's accession to the European Union EU, it exceeded the level of Ukraine's GDP [3, 4]:

3.9 times Total GDP;

GDP per capita is 9.12 times.

for the period from 2004 to 2016, the GDP was as follows: the total GDP of Poland exceeded the level of Ukraine by 5.2 times; and GDP per capita by 5.92 times.

At the same time, the needs of the state are growing in the development of information management, which needs constant renewal and support.

Consider the criteria for evaluating information resources (Law of Ukraine "The Procedure for the Functioning of Web-sites of Executive Bodies" [5]):

information structure of the website;

site map;

posting of the news on the first page;

uniform format of website addresses;

availability of references to structural subdivisions;

the presence of controls that help to change the color contrast of the site, the size of the font;

Availability of several language versions of the interface and content of the site;

organization of access to electronic documentation.

In the process of studying compliance with these criteria on the example of the website of the Verkhovna Rada of Ukraine [6] and the Sejm of Poland [7], the following results were obtained.

1. The information structure of the website of the executive body should be adequate to the relevant subject area.

Information content should be, on the one hand, complete, and on the other hand, it should not contain any unnecessary information, inappropriate repetitions,

etc. The first and second sites meet the information content requirement.

2. Regarding sitemap, the site of the Verkhovna Rada of Ukraine has it and the site of a Polish Sejm also has sitemap. The site of the Verkhovna Rada has a sitemap at the top of the site, on the site of the Sejm it's at the bottom.

3. News should be on the first page of the site in accordance with the legislation of Ukraine. Depending on the design, it is possible to place in the center of the main page or in the right field. Placing news is based on the principle of issuing a news headline with the following link to the full text of the document. On the website of the Verkhovna Rada of Ukraine, we can see that the news feed is present, but not in accordance with the law, but it is modernized and has a user-friendly look. But the site of the Sejm of Poland meets the requirements presented in the law of Ukraine.

4. Web addresses of the sites should be reduced to the uniform format, have to reduce the names of ministries, state committees, regional state administrations, other executive bodies, bringing to the second format of domain names of the second level (name.gov.xx).

In addition, you have to indicate links to international web resources that are the main or priority partners of the authority. It is not recommended to post web-addresses and links to commercial web resources that are not on the authority's profile, or if the body's specificity is not a priority for the company's web resource.

The conducted author's research showed that the addresses of both sites meet the requirements of the law, references to commercial non-profile organizations were not revealed.

5. The websites should include contact information of the structural units and / or officials of the executive authorities responsible for the software, technical and information support of the websites. The sites under investigation meet this requirement and contain all the necessary links.

6. The site must contain controls that help to change the color contrast of the site, the size of the font. Unfortunately, these tools were not found on the Verkhovna Rada website. However, on the site of the Sejm of Poland, these elements are present. This indicates that the developers have taken care that the site will be viewed by different people, and for example, people with poor eyesight will be able to enlarge the font.

7. With the regard to the criterion which concerning the multilingual site, both the site of the Verkhovna Rada of Ukraine and the site of the Sejm of Poland adhere to it, however, versions in languages other than the state may differ in the amount of information available.

A generalized study on compliance with these criteria on the example of the website of the Verkhovna Rada of Ukraine [6] and the Sejm of Poland [7] is given in table 1.

It is proposed to use the evaluation: 0 – the site does not meet the requirements; 0.5 – partially satisfies the requirement; 1 – fully compliant. In order to get a correct assessment of the site, we will introduce weighting factors depending on the importance of the criteria set forth in the Law of Ukraine in terms of information management.

TABLE I. RESULTS OF ANALYSIS OF COMPLIANCE OF INFORMATION RESOURCES OF THE STATES WITH THE LEGISLATION OF UKRAINE

Criterion	Indicator's Weight	Website of the Verkhovna Rada of Ukraine		Website of the Sejm of Poland	
		points	weighted points	points	weighted points
Information structure of the website	1	1	1	1	1
The presence of a website map	1	1	1	1	1
Location of the news on the first page	0,25	1	0,25	1	0,25
Design Website addresses in the same format	0,50	1	0,5	1	0,5
References to structural subdivisions	0,75	1	0,75	1	0,75
The presence of controls that help to change the color contrast of the website, the size of the font	0,50	0	0	1	0,5
The presence of several language versions of the interface and content of the website	1	1	1	1	1
Organization of access to electronic documentation	1	1	1	1	1
Total points		7	5,5	8	6

So, with a small margin, the site of the Sejm of Poland is the best. The country is concerned not only

with the development of the economy and public welfare, but also about information management. This is an indication that Poland is paying attention to whether the site will be accessible to people with certain disabilities, whether the site is understandable to the average user, or whether it can quickly orientate on it. The site of the Verkhovna Rada of Ukraine as a whole is in accordance with the law of Ukraine, however, the critical review of the requirements for sites submitted in the law of Ukraine "On the functioning of websites of executive authorities" [5], showed that they are not fully complied with.

A detailed analysis showed that for seven out of eight requirements the government site of Ukraine is responsible. The requirement for the presence of controls that help to change the color contrast of the site, the size of the font does not meet. Functions that help change the font size or color contrast of the site are not as commonly used by ordinary citizens, but they should be. Apparently because of the complexity of implementation at one time, the developers abandoned them.

In order to improve the site, it is proposed to add a clock that will be located on each site's site and show the exact date and time zone of the person who comes to the site and how to write whether or not the government is currently working, so that people who I would like to come to him with certain goals to take into account this information, before going to the government.

In addition, the authors recommend that the visual design be improved, as the government sites of developed countries look more modern, while the design of the Ukrainian government portal is not updated in accordance with modern trends in the field of information technology.

The development of user interfaces should be structured in such a way as to make it as attractive and convenient as possible to optimize its user interface. The main requirement is convenience, practicality and intuition.

Therefore, during the developing of high-quality websites for any purpose, first of all, it is necessary to observe the provisions of the Law of Ukraine "On the procedure for the functioning of executive bodies' websites" [5]; secondly, secondly, to observe the general requirements of User Experience Design (UX) , relating to information architecture, interconnection design, graphic design and content (figure.1).

Consider the basic requirements and principles of User Experience Design.

1. Principle «Keep it short and simple» (KISS). The interface should be simple and understandable, tasks must be solved with a minimum number of actions, and everything should be clear and obvious.

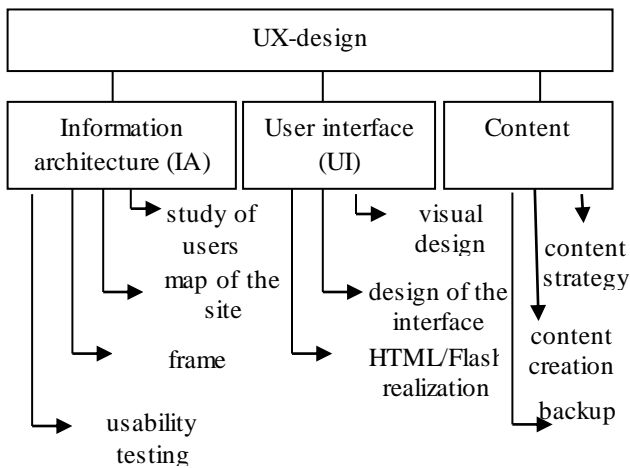


Figure 1. The components of User Experience Design

2. Don't make you think. You need to avoid complicated actions that make users think.

3. Don't display the obvious. You should not show the obvious elements of the interface, you need to focus only on the really necessary things.

4. Signal to noise ratio. In each interface there are important elements (signals) and insignificant for a certain part of the system (noise), of course, you need to focus on signals and avoid noise.

5. Proven development is the best fashion. It is not necessary to cling to the fashion and do it only because others do it, it is better to give preference to the tested elements of the interface.

6. The usual controls. In any modern interface there are many controls, it will be better to use the usual elements and visual images.

7. People don't read, people are browsing. Users don't like, don't always have the time and ability to read large arrays of texts, don't force them to do so.

8. The principle of smart borrowing. Do not reinvent a bike for standard things.

9. Miller's purse. In one functional block there can be no more than 5-7 items, otherwise the user will not be able to hold information in memory.

10. The principle of grouping. It is desirable to place the information on a page with several logical blocks (groups) for the purpose of more convenient orientation.

11. Intuitive clarity. Understanding is better memorization.

12. Everything is useful in mind. All important elements of the interface should be in sight and appropriately highlighted.

13. Principle of 3 clicks. There should be no more than 3 clicks to move from one section to another. The same rule applies to the main page: any important information should be available in no more than 3 clicks.

14. Homogeneity. In large projects, there is often a homogeneous functional in different parts of the site

(for example, comments), it should not be different. The same applies to the style.

15. Ways to solve the problem. Users need to offer ways to solve their tasks through the interface, these methods should be obvious.

16. The principle of bridge railings. This principle implies the protection of users from accidental actions.

17. Correct copy writing. Designing the interface is a lot of copywriting, every letter, especially headings, is important.

18. The principle of unity. Settings and controls need to be tried not to hide in separate sections, but to allow you to manage from one place, where appropriate.

19. Trends It is necessary to take into account the current trends, so that the interface is not obsolete before the project, but to approach it thoughtfully.

## CONCLUSIONS

The article deals with the main economic indicators and aspects of information management in Ukraine and Poland. Every year, besides consolidating its positions in the economy, the country needs to develop information management, which is essentially a business card of the state on the international scene as well as the primary source for information in various life's spheres. As part of the research, the existence of a relationship between information management and the development of the country's economy was confirmed. The faster the country develops its economy, the more and more new trends in the development of its information management, it introduces, thus, improves the understanding of what is happening in the state, develops new methods for communicating for all people including the people with hearing impaired and vision-impaired people. State information management is a powerful tool that should carry out strategic, operational and administrative tasks.

## REFERENCES:

- [1] Melnik V. V. Formation of the concept of information management: the essence, tasks, main directions of development, *Humanitarian Herald*, no5, pp. 12–15, 2012. [Online serial]. Available : [http://www.zgia.zp.ua/gazeta/VISNIK\\_49\\_122.pdf](http://www.zgia.zp.ua/gazeta/VISNIK_49_122.pdf).
- [2] The World bank: doing business [Online]. Available: <http://www.doingbusiness.org/rankings>. [Accessed: April. 30, 2018].
- [3] The World bank: Ukraine [Online]. Available: <https://data.worldbank.org/country/ukraine>. [Accessed: April. 30, 2018].
- [4] The World bank: Poland [Online]. Available: <https://data.worldbank.org/country/poland>. [Accessed: April. 30, 2018].
- [5] The Law of Ukraine on the Procedure for the Functioning of Web-sites of Executive Bodies [online document]. Available: <http://zakon5.rada.gov.ua/laws/show/z1022-02>. [Accessed: April. 30, 2018].
- [6] Verkhovna Rada of Ukraine: official web portal [Online]. Available: <http://portal.rada.gov.ua/en>. [Accessed: April. 30, 2018].
- [7] Sejm Rzeczypospolitej Polskiej [Online]. Available : <http://www.sejm.gov.pl/>. [Accessed: April. 30, 2018].

# Modeling the Probable Losses of Banks from their Involvement in the Process of Legalization (Laundering) of Inflammable Funds

Anton Boiko<sup>1</sup>, Tetiana Dotsenko<sup>2</sup>

<sup>1,2</sup> Sumy State University, Ukraine, <sup>2</sup>Sumy regional branch Joint-stock company “Oschadbank”, Ukraine  
<sup>1</sup>a.boiko@uabs.sumdu.edu.ua, <sup>2</sup>tvdocenko85@gmail.com

**Abstract** – The article is stressed on the essential characteristic and mathematical formalization of determination the probable losses of banks from their involvement in the process of legalization (laundering) of criminal incomes. It involves significance of the main indicators of losses, mathematical models for assessing the relevant risk factors, determining the comparability of factors of bank risks with the possible advantages obtained from overcoming them is carried out. It is implemented the construction of consumable matrices and trees for making possible alternatives to the leveling of banking risks.

**Keywords** – *Losses of banks, legalization of funds, banking risks, expense matrices, tree of solutions.*

## I. INTRODUCTION

In the current conditions of financial instability in Ukraine and the reduction of demand for banking services, the process of optimizing costs with all available methods is becoming increasingly important for banks. Understanding of the management of the bank, which operations lead to significant losses is possible only if they are quantified. Therefore, it is advisable to develop a scientific and methodological approach to assessing the probable losses of banks from their involvement in the process of legalization (laundering) of funds obtained illegally. Of course, the process of legalization of criminal incomes should be neutralized in the bank immediately upon its detection, but the responsibility of banking workers who carry out internal financial monitoring will be significantly higher, provided knowledge of the amount of probable costs in the event of the occurrence of relevant adverse events.

The aim of the work is to develop a scientific and methodological approach to determining the probable losses of banks from their involvement in the process of legalization (laundering) of funds obtained illegally.

It should be noted that scholars are differently considering financial monitoring in the context of combating the legalization of proceeds from crime, terrorist financing and the proliferation of weapons of mass destruction. The contemporary researchers N.V.

Moskalenko, Yu.O. Romanchenko and others reveal directly the general aspects of financial monitoring. [9,12]. The specific consequences of legalization (laundering) of proceeds from crime are described by a number of authors O.Alexeychenko, I.M.Tohtarov [1,13]. To study the process of counteraction to legalization (laundering) of money received illegally, a number of models were proposed: model Dmitrov S.O. [6], structurally logical model O.V. Kuzmenko [7], the model of risk assessment VPBratiuk [3] et al.

Despite the significant contribution of these researchers to the disclosure of this problem, the conclusions and measures that have been formed up to now need to be supplemented with features of determination and taking into account losses from attracting banks to the process of legalization (laundering) of money received illegally.

## II. RESEARCH RESULTS

Thus, we will carry out a phased implementation of the scientific and methodical approach to determining the probable losses of the bank from their involvement in the process of legalization (laundering) of funds received illegally:

Istage. Formation of the indicative space of the main indicators of the bank's losses from their involvement in the process of legalization (laundering) of funds obtained illegally, taking into account both external and internal changes in the environment of the bank's operations. As part of this phase, it is necessary to determine how relevant the risk factors inherent in banking activities: general (the level of corruption in the country (Z1), the level of stability of the economic situation (Z2), the level of political stability in the country (Z3), the level of well-being of the population (Z4) the level of information technology development (Z5)); Specific (S1), Market Risk (S2), Credit Risk (S3), Strategic Risk (S4), Shareholder Risk (S5), Management Risk (S6), Reputational Risk (S7), Legal Risk (S8) ), personnel risk (S9), technological risk (S10), environmental risk (S11)); country risk (environmental risk (RC1), reputational risk (RC2), legal risk (RC3));

client risk (credit risk (RK1), strategic risk (RK2), shareholder risk (RK3), management risk (RK4), reputational risk (RK5), legal risk (RK6)); service risk (liquidity risk (RP1), market risk (RP2), credit risk (RP3), strategic risk (RP4), shareholder risk (RP5), management risk (RP6), reputational risk (RP7), legal risk ( RP8), personnel risk (RP9), technological risk (RP10), environmental risk (RP11)) - and benefits: increasing financial flows; expansion of the client base of the bank; intensification of demand for banking services; maintaining a license for banking services; stable functioning of a financial institution; cooperation with

international partners - which the bank receives in case of avoiding or overcoming the consequences of the impact of these risks.

The research and identification of relevant risk factors inherent in banking activities, as well as the benefits derived from their avoidance and overcoming, are the basis for the next stage of implementation of the methodological approach to determining the probable losses of banks from their involvement in the process of legalization (laundering) of money received illegally and, accordingly, construct a table of correspondence (see Table 1).

TABLE I. ESTABLISHING THE CORRESPONDENCE OF THE ACHIEVED ADVANTAGES OF BANKS DUE TO OVERCOMING THE RISKS INHERENT IN ITS ACTIVITIES TO THE RELEVANT FACTORS THAT DETERMINE THE RECEIPT OF THESE BENEFITS.

Relevant risk factors inherent in banking activities	The benefits that a bank gets when avoiding or overcoming the effects of risks					
	Increasing the volume of financial flows (P1)	Expansion of the client base of the bank (P2)	Intensification of demand for banking services (P3)	Saving a license for banking services (P4)	Stable functioning of a financial institution (P5)	Cooperation with international partners (P6)
General						
Z1	$z_{11}$	$z_{12}$	$z_{13}$	$z_{14}$	$z_{15}$	$z_{16}$
Z2	$z_{21}$	$z_{22}$	$z_{23}$	$z_{24}$	$z_{25}$	$z_{26}$
Z3	$z_{31}$	$z_{32}$	$z_{33}$	$z_{34}$	$z_{35}$	$z_{36}$
Z4	$z_{41}$	$z_{42}$	$z_{43}$	$z_{44}$	$z_{45}$	$z_{46}$
Z5	$z_{51}$	$z_{52}$	$z_{53}$	$z_{54}$	$z_{55}$	$z_{56}$
Specific						
S1	$s_{11}$	$s_{12}$	$s_{13}$	$s_{14}$	$s_{15}$	$s_{11}$
S2	$s_{21}$	$s_{22}$	$s_{23}$	$s_{24}$	$s_{25}$	$s_{21}$
...	...	...	...	...	...	...
S11	$s_{111}$	$s_{112}$	$s_{113}$	$s_{114}$	$s_{115}$	$s_{111}$
Country risk						
RC1	$c_{11}$	$c_{12}$	$c_{13}$	$c_{14}$	$c_{15}$	$c_{16}$
RC2	$c_{21}$	$c_{22}$	$c_{23}$	$c_{24}$	$c_{25}$	$c_{26}$
RC3	$c_{31}$	$c_{32}$	$c_{33}$	$c_{34}$	$c_{35}$	$c_{36}$
Customer risk						
RK1	$k_{11}$	$k_{12}$	$k_{13}$	$k_{14}$	$k_{15}$	$k_{16}$
RK2	$k_{21}$	$k_{22}$	$k_{23}$	$k_{24}$	$k_{25}$	$k_{26}$
...	...	...	...	...	...	...
RK6	$k_{61}$	$k_{62}$	$k_{63}$	$k_{64}$	$k_{65}$	$k_{66}$
The risk of services						
RP1	$p_{11}$	$p_{12}$	$p_{13}$	$p_{14}$	$p_{15}$	$p_{16}$
RP2	$p_{21}$	$p_{22}$	$p_{23}$	$p_{24}$	$p_{25}$	$p_{26}$
...	...	...	...	...	...	...
RP11	$p_{111}$	$p_{112}$	$p_{113}$	$p_{114}$	$p_{115}$	$p_{116}$

Stage 2. Selection or development of mathematical models to provide a quantitative description of each of the selected relevant risk factors. At this stage, it becomes necessary to take into account the fact that the risk factors acquire both qualitative and quantitative values.

Stage 3. Determining the comparability of the factors of bank risks and benefits that the bank receives when avoiding or overcoming the effects of risks, as well as the formalization of the identified compliance in tabular form. In addition, in this phase, it is necessary to

analyze the sensitivity of the relevant risk factors inherent to banks, taking into account the sum of the binary indicators of the tables of comparability of relevant risk factors and related benefits.

Stage 4. Implementing a cost-effective approach for relevant risk factors that does not provide an opportunity to obtain appropriate benefits for banks by constructing cost matrices (formula 1) and determining the probabilities of obtaining them in each particular situation (formula 2).

$$L = \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \left( \begin{array}{l} \left( \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} + \right) \left( \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} + \right) \\ + \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \left( + \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right) \\ \left( \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} + \right) \left( \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} + \right) \\ + \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \left( + \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right) \end{array} \right) \quad (1)$$

$$P = \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \left( \begin{array}{l} \left( \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right) \left( \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right) \\ \times \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \left( \times \min P \left[ \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \right) \\ \left( \min P \left[ \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right) \left( \min P \left[ \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right) \\ \times \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \left( \times \min P \left[ \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \right) \end{array} \right) \quad (2)$$

Stage 5. Formation of a decision tree for possible alternatives to overcoming banking risks (formula 3-4):

$$R = \left( \begin{array}{l} \left( \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} + \right) \left( \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} + \right) \left( \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} + \right) \\ + \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \left( + \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right) \left( + \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right) \end{array} \right) \left( \begin{array}{l} \left( \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right) \left( \min P \left[ \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right) \left( \min P \left[ \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right) \\ \times \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \left( \times \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \right) \left( \times \min P \left[ \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \right) \end{array} \right) \quad (3)$$

$$\left\{ \begin{matrix} L \\ P(R \leq L) \end{matrix} \right\} = \left\{ \begin{matrix} \left( \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} + \right. \\ \left. + \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right) \\ \left( \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right. \\ \left. \times \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \right) \\ \left( \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] + \right. \\ \left. + \min P \left[ \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right. \\ \left. \times \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \right) \\ \left( \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} + \right. \\ \left. + \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right) \\ \left( \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right. \\ \left. \times \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] + \right. \\ \left. + \min P \left[ \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \times \right. \\ \left. \times \max P \left[ \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right] \right) \\ \left( \max \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} + \right. \\ \left. + \min \left\{ L_{lj} \mid 1-r_{lj}=1 \right\} \right) \\ 1 \end{matrix} \right\} \quad (4)$$

where  $L$  - matrix of bank expenses at various combinations of occurrence of negative consequences of risk;

$P$  - probability of occurrence of expenses of the bank in each concrete situation;

$$r_{lj} = \begin{cases} 1, & \text{if } l \text{ relevant factor of risk gives } j \text{ advantage} \\ 0, & \text{in the other case} \end{cases} \quad (5)$$

where  $r_{lj} = z_{lj}$  - in the context of a group of common risks;

$r_{lj} = s_{lj}$  - in the context of a group of specific risks;

$r_{lj} = c_{lj}$  - in the context of the country's risk group;

$r_{lj} = k_{lj}$  - in the context of a client's risk group;

$r_{lj} = p_{lj}$  - in the context of the product risk group;

$L_{lj}$  - the amount of expenses borne by the banking institution in the event of non-compliance with the established requirements.

#### CONCLUSIONS

Summarizing the results of the conducted research, it should be noted that the use in the practical work of scientific and methodological approaches to determining the probable losses of banks from their involvement in the process of legalization (laundering) of funds obtained illegally.

On the basis of mathematical formalization of the above calculations, using a cost approach, the construction of costly matrices, the formation of a tree of solutions for possible alternatives to overcoming the risks of banking activity.

In parallel with the increase of the internal system of bank monitoring is getting a number of these benefits: increasing the volume of financial flows; expansion of the client base; intensification of demand for banking services; maintaining a license for banking services; stable functioning of a financial institution; cooperation with international partners.

#### REFERENCES:

- [1] O. Alekseychenko, "Forensic characteristic of legalization (laundering) of proceeds from crime". [Online]. Available: <http://apnl.dnu.in.ua/arkhiv?id=41>.
- [2] V.P. Bratyuk, "Application of risk assessment model in the process of financial monitoring", *Young scientis*, vol. 1, no. 16, pp. 56-59, 2105.
- [3] S.O. Dmitrov, "Ways to assess the effectiveness of financial monitoring in a bank", *Economy and the state*, vol. 2, pp. 22-26, 2011.
- [4] O.V. Kuzmenko, "Development of Methodological Principles for Modeling the Risk of a Banking Institution for the Use of Its Services for the Legalization of Criminal Income or Terrorist Financing in the course of Inspection", *Economy and the State*, vol. 2, pp. 46-55, 2017.
- [5] N.V. Moskalenko and N.Ya. Klymchuk, "Management of the risks of legalization of proceeds from crime", *Collection of scientific works of the University of the State fiscal service of Ukraine*, vol. 1, pp. 183-194, 2017.
- [6] Y. O. Romanchenko, "Financial monitoring as an instrument for preventing the legalization of criminal incomes", *Scientific view: economics and management*, vol. 1, pp. 69-75, 2017.
- [7] I. M. Tokhtarova, "Legalization (laundering) of "dirty" money in Ukraine as a corrupt component", *Bulletin of the Academy of the Customs Service of Ukraine Series: Public Administration*, vol. 1, pp. 126-131, 2015.



# Regional Sustainability Assessment Through Multivariate Statistical Analysis

Tetiana Marynych<sup>1</sup>, Stanislav Smolenko  
Sumy State University, Ukraine, <sup>1</sup>t.marynych@ssu.edu.ua

**Abstract** – The work assesses regional differentiation based on the extended and classical variables' selection using multivariate statistical techniques: cluster and principal component analysis.

**Keywords** – Cluster Analysis, Principal Component Analysis, Sustainable Regional Development.

## I. INTRODUCTION

The concept of sustainable and inclusive development (SID) has been widely discussed in recent decades and has been recognized as the priority policy areas both at the regional and international levels [1]-[2]. Their successful realization requires proper identification of the main indicators, their assessment, and monitoring. Modern researchers extend classical approach that focuses on the issues of poverty, inequality, unemployment, and growth by the variables that evaluate education, institutions, health, environmental sustainability and informality [3]-[4]. Methodology varies from aggregate indexing and causality modeling to regional, international structuring using multivariate statistical techniques, e. g. clustering, factor and principal component analysis [3]-[5].

The main objective of this paper is to make regional differentiation based on the extended and classical variables' selection using cluster analysis (CA) and principal component analysis (PCA).

## II. DATA AND METHODOLOGY

Our research explores 22 economic, financial, social and environmental indicators for 24 Ukrainian regions. Regional codes used in the analysis are given in Table I.

TABLE I. DESCRIPTION OF REGIONAL CODES

Region	Code	Region	Code
Vinnytsia	VN	Mykolaiv	MK
Volyn	VO	Odessa	OD
Dnipropetrovsk	DP	Poltava	PT
Donetsk	DN	Rivne	RV
Zhytomyr	ZH	Sumy	SM
Zakarpattia	ZA	Ternopil	TP
Zaporizhia	ZP	Kharkiv	KH
Ivano-Frankivsk	IF	Kherson	KS
Kiev	KV	Khmelnytskyi	KM
Kirovohrad	KG	Cherkasy	CH
Luhansk	LG	Chernivtsi	CS
Lviv	LV	Chernihiv	CN

We calculated variables as relative indicators using average values for 2013-2016 years, based on official data of State and regions' statistics offices [6]. It should be noted that due to the data unavailability the values of GRP for 2016 were estimated on the basis of the 2015 values, multiplied by the index of industrial production in 2016.

Table II provides a description of the indicators used in this multivariate analysis.

TABLE II. DESCRIPTION OF VARIABLES

Indicators	Code
<b>1. Social and demographic indicators</b>	
The ratio of the regional population to the total population of Ukraine	ppl
The ratio of the population over the age of 60 to the economically active population of working age	old
Total fertility rate (per woman)	br
The share of urban population	ur
The ratio of the average life expectancy at birth to the average in Ukraine	ls
The number of offences in the region per 1000 population (crime rate)	off
The ratio of average monthly wages by region to the average monthly wage in Ukraine	sal
<b>2. Economic and financial indicators</b>	
Terms of trade (the ratio of exports to imports)	tr
The ratio of direct foreign investments to GRP	di
The ratio of capital investments to GRP	ci
The ratio of gross regional product (GRP) to gross domestic product (GDP) of Ukraine	grp
The ratio of local budget expenditures to GRP	exp
The ratio of local budgets' incomes to their expenditures	inc
Spread between the credit interest rates of banking institutions and the National Bank discount rate	ir
The ratio of loans to GRP	loa
The ratio of deposits to loans	dep
Consumer price index (compared to December of the previous year)	pi
<b>3. Environmental indicators</b>	
CO2 emissions per GRP	poll
The volume of wood harvesting per unit of area	wd
The extraction of water bioresources per 1000 population	wt
The amount of waste per capita	wst
The amount of recycled waste per capita	ut

PCA can substitute a data set with a smaller number of representative variables that explain most of the variability in the original data [8]. PCA as a dimension reduction tool is a common strategy to deal with the problem of correlation dependence between factors, as well as the problem degrees of freedom (correspondence of the number of variables to the number of observations) in small samples. Principal components represent linear combinations of the original variables weighted by their contribution to explaining the variance of the variables. A matrix of the factor loadings represents a system of the following equations [8]:

$$PC_i = W_{1i} \cdot X_1 + W_{2i} \cdot X_2 + \dots + W_{3i} \cdot X_3, \quad (1)$$

where  $PC_i$  –  $i$ -principal component;  $X_k$  – a  $k$ -feature,  $W_{ik}$  – the corresponding loading (weight).

Principal components' calculation applies a singular value decomposition of the centered and scaled data matrix. Centering and scaling are performed using the following formula:

$$x_{ij} = \frac{x_{ij} - M(X_i)}{\sigma(X_i)}. \quad (2)$$

Here  $X_i$  – is a vector of corresponding features;  $x_{ij}$  – a particular value of  $i$ -feature of  $j$ -element;  $M(X_i)$  – is a mean of  $X_i$ ;  $\sigma(X_i)$  – a standard deviation of  $X_i$ .

Generated principal components and a matrix of the estimated factor loadings are used further in the regional structuring based on the clustering techniques. Our research employs the  $K$ -means algorithm based on the recommendations given in a "Typology of the regions of Ukraine", prepared within the framework of the EU project on supporting regional development policy in Ukraine [2]. This method implies partitioning observations into  $k$  groups by minimizing the sum of squares from the points to the assigned cluster centers [8]:

$$\sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \rightarrow \min. \quad (3)$$

Here  $S_i$  represents  $i$ -cluster;  $x_j$  –  $j$ -element coordinates;  $\mu_i$  – coordinates of the  $i$ -cluster's centroid (mean of the coordinates of the  $i$ -cluster's elements).

### III. ANALYSIS

In this section we present the main findings of our research. Fig. 1 shows numerous correlation dependences between variables.

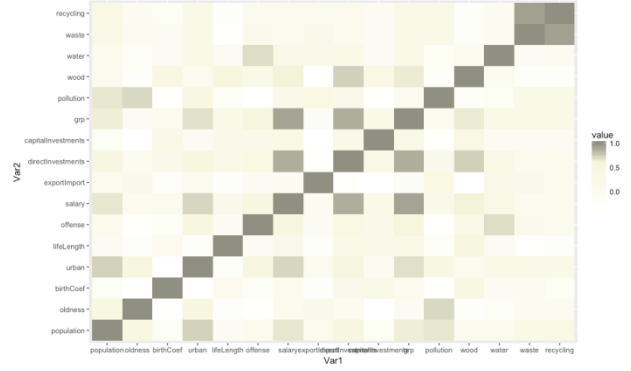


Figure 1. Correlation matrix  
Source: Own processing based on official statistics [6]

Table III summarizes the importance of the principal components and demonstrates six most disperse components, which proportion of variance is greater than 0.05, total cumulative proportion of variance 0.885.

TABLE III. PRINCIPAL COMPONENT ANALYSIS SUMMARY

Region	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	2.197	1.934	1.412	1.326	0.961	0.952
Proportion of variance	0.302	0.234	0.125	0.11	0.058	0.057
Cumulative proportion	0.302	0.536	0.66	0.77	0.828	0.885

Fig. 2 exhibits associations between the diversity of the regions of Ukraine and the first two principal components displayed as red vectors' projections according to the weight and direction of the factor loadings. This graph does not show the factors "ppl", "ut", "off", as these vectors overlap and blind the visual perception.

The first principal component reflects economic and financial development of the regions. It is mostly affected by the following economic indicators: the ratio of local budgets' revenues to their expenditures ("inc"); the average salary in the region ("sal"); the share of GRP in GDP of Ukraine ("grp"); the ratio of loans to GRP ("loa"); the ratio of local budget expenditures to GRP ("exp") (negatively).

The second principal component characterizes human, natural and capital resources of the regions. Main contributions are made by the following factors: the ratio of capital investments to GRP ("ci"); the ratio of people aged 60+ to the economically active population ("old") (negatively); the birth rate ("br"); the difference between the credit interest rates of banking institutions and the National Bank discount rate ("ir") (negatively); the volume of wood harvesting per unit of area ("wd"). The absolute weights of these loadings are higher than 0.3.

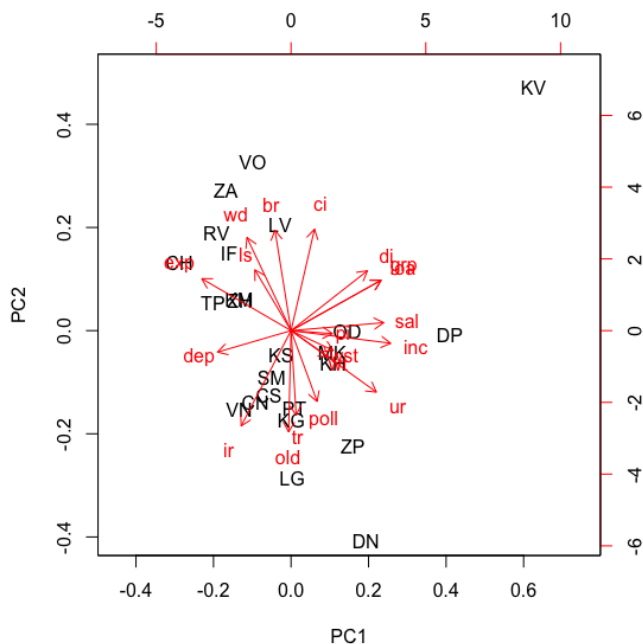


Figure 2. PCA visualization  
Source: Own processing based on official statistics [6]

The third principal component accounts for ecological, monetary stability and the rule of law. This component is influenced primarily by the ratio of CO2 emissions to GRP (“*poll*”) (negatively); the crime rate (“*off*”); the consumer price index (“*pi*”); the water usage (“*wr*”). Notably, the fourth principal component also applies to the sustainable development of the regions. It is mostly affected by two variables – the amount of waste per person (“*wst*”) and the amount of recycled waste per person (“*ur*”). Both factors have a near 0.5-weight impact with a negative sign that declares a reverse relationship between waste-related factors and the level of region’s sustainable development.

The fifth principal component is mainly determined by the human capital indicators, i.e. by the birth rate (“*br*”), the ratio of people aged 60+ to the economically active population (“*old*”) (negatively), the ratio of the average life expectancy at birth to the average in Ukraine (“*ls*”) (negatively). Besides, the extraction of water bioresources per 1000 population (“*wr*”), terms of trade (“*tr*”), also have absolute loading values greater than 0.35.

The sixth principal component duplicates the loadings of the fifth component (“*ls*”, “*wr*”, “*tr*”), adding the price index (“*pi*”) and the ratio of capital investments to GRP (“*ci*”), which have positive loading values of 0.3.

The findings of the principal component analysis show that variables of human capital, natural resources usage and monetary and financial stability prevail in the choice of principal components. At the same time such factors as regional population (“*ppl*”), urban population (“*ur*”), direct foreign investments (“*di*”), and the ratio of

deposits to loans (“*dep*”) do not particularly affect any of the first six principal components.

For cluster analysis, we take the first six main components that summarize the main features of the original variables. As mentioned above, we apply the *K*-means clustering algorithm, setting  $k=5$  based on the typology of the regions of Ukraine [2], findings of Ukrainian economists [5] and our own empirical assessments. Fig. 3 exhibits the results of the regional clustering within the estimated coordinates of the first two principal components that represent economic, financial, human capital and sustainable development of the regions. Here we use basic geometric structures [9] for simplicity.

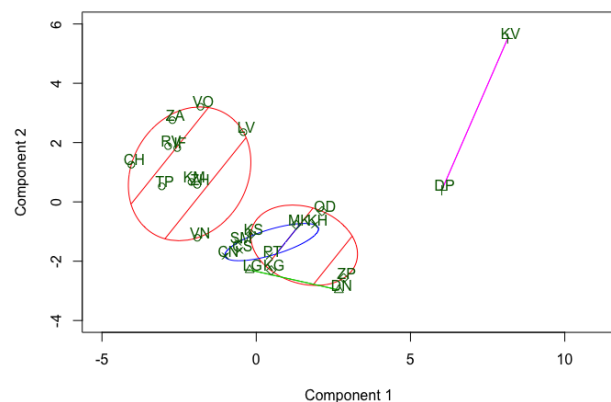


Figure 3. *K*-means clustering  
Source: Own processing based on official statistics [6]

Fig 4 highlights the groups (clusters) of regions of Ukraine based on the results of the cluster analysis:

Group 1 (blue): Vinnytsia, Volyn, Zhytomyr, Zakarpattia, Ivano-Frankivsk, Lviv, Rivne, Ternopil, Khmelnytskyi and Chernivtsi regions.

Group 2 (violet): Donetsk and Luhansk regions.

Group 3 (red): Dnipropetrovsk region, Kyiv region including Kyiv.

Group 4 (orange): Chernihiv, Sumy, Cherkasy, Poltava and Kharkiv regions.

Group 5 (green): Kirovohrad, Kherson, Mykolaiv, Zaporizhia and Odessa regions.

Obviously, the regions with the most developed cities (Kyiv and Dnipro) turn up to be in the same group. Donetsk and Luhansk as occupied territories have significant reasons to differ from other regions a lot. Surprisingly, all the other regions of Ukraine are grouped based on the first two principal components in conformity with their geographical location, creating Western, Northern, Eastern and Southern clusters.

Our results partly correspond with the typology of the regions of Ukraine [2] constructed in 2012 on the basis of geographical, socio-demographic, economic, ecological indicators, that focuses on the sustainability issues as well (Fig. 5).

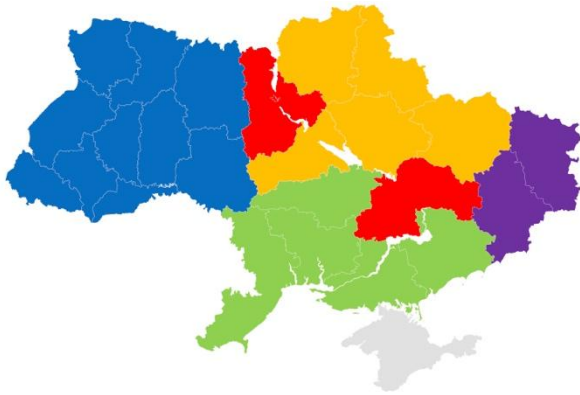


Figure 4. Mapping the clustering results  
Source: Own processing based on official statistics [6]



Figure 5. Mapping the clustering results  
Source: [6]

The differences are explained by the choice of the original variables and the time period used in the analysis. We can see that recent changes in political and economic development of Ukraine provoked deepening of the differentiation between the western border regions and Northeastern regions, and deteriorating of the sustainability of the majority of the regions.

#### CONCLUSIONS

In this work, we have investigated the factors that determined regional sustainability and inclusiveness. Following up the comprehensive modern approach, we extended the traditional set of variables with the factors that influence financial stability, human capital development, and sustainability goals' achievement.

We estimated relative standardized indicators to avoid possible differences in the size of regions and scales of measurements, removing abnormal observations. Multivariate statistical tools were applied (PCA and clustering) to reveal the simplified patterns of possible relationships and data structuring.

It was proved that economic factors explain a dominant part of sustainable regional development, explaining 30% of the cumulative variance of the data. At the same time, our findings showed that human capital

and ecological factors turned to be the most impactful factors that account for the rest 58% of the cumulative variance of the first six principal components of the observed data set.

We used the principal components, which describe the most important features of the data to make comparative cluster analysis of the regions of Ukraine. It was found that clusters were grouped by geographic position of the regions. In particular, empirical results revealed that Western regions had better indicators of the human capital (fewer people of the retirement age, higher life expectancy, lower urbanization and crime rates) and sustainable development (better use of water resources; lower waste volumes, but the worse situation with the deforestation). Central and Southern regions on average had better economic and financial position, while Eastern regions showed the biggest problems with human capital and environmental protection.

To conclude, this work contributes to solving the problem of the assessment and monitoring of the regional development, emphasizing on the importance of the elaboration of the multidimensional frameworks that account for a broad group of indicators that depict new challenges in terms of sustainability, inclusiveness and economic freedom. Short-term and long-term aspects of sustainable regional development and possible associations between variables can be further explored using methods of linear regression, linear discriminant analysis or artificial networks.

#### REFERENCES:

- [1] "Draft Outcome Document of the United Nations Summit for the Adoption of the Post-2015 Development Agenda", A/69/L.85, 2015.
- [2] "Regional Development and State Regional Policy in Ukraine: Analytical Report". [Online]. Available: <http://surdpu.eu/Analytical-Report>. [Accessed April 12, 2018]. (in Ukrainian)
- [3] R. M. Aguilar, "Assessment of socio-economic development through country classifications: a cluster analysis of the Latin America and the Caribbean (LAC) and the European Union (EU)", *World Economy Journal*, No. 47, pp. 43-64, 2017.
- [4] T. O. Marynych, "Empirical assessment of the long-term aspects of sustainable regional development", *Economic Annals XXI*, No. 166(7-8), pp. 86-90, 2017.
- [5] O. A. Ryadno and Berkut O. V. "Study of structure and dynamics of differentiation of regional socio-economic development on the basis of cluster analysis", *Donbass Economic Bulletin*, No. 1(43), pp. 60-67, 2016. (in Ukrainian)
- [6] "State Statistics Service of Ukraine (2013-2016)". *Statistical data*. [Online]. Available: <http://ukrstat.gov.ua/>. [Accessed April 12, 2018]. (in Ukrainian)
- [7] "The comprehensive R Archive Network". [Online]. Available: <https://cran.r-project.org/>. [Accessed April 12, 2018].
- [8] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning with Applications in R*. NY: Springer, 2013.
- [9] Z. Gniadzowski. "New Interpretation of Principal Components Analysis", *Zeszyty Naukowe WWSI*, No 16, Vol. 11, pp. 43-65, 2017.

# Simulation of Scoring of the Bank's Borrowers Creditworthiness

Konstantin Gritsenko

Sumy State University, Ukraine, k.hrytsenko@uabs.sumdu.edu.ua

**Abstract – The advantages of credit scoring are considered. The characteristics of bank borrowers used in the construction of logit model, decision tree and neural network are described. The choice of the best model is made.**

**Keywords – scoring model, credit risk, borrower creditworthiness, intellectual data analysis, logit model, neural network, decision tree.**

## I. INTRODUCTION

The current economic situation in Ukraine has led to an increase in credit risks associated with non-repayment of loans. One of the ways to reduce credit risks is the use of scoring technologies that allow you to rapidly assess the creditworthiness of potential borrowers based on questionnaire data. The use of credit scoring facilitates the speed of decision-making on granting loans and conducting an express analysis of creditworthiness in the presence of the borrower. Most banking professionals consider that credit scoring is the most suitable technology for consumer lending. For example, see [1]. The introduction of credit scoring system in the practice of banks is necessary both for the banks themselves to assure the return of the loan as well as for borrowers because the scoring system significantly reduces the time taken by the bank to decide on a loan.

It is important to note that for a scoring model it is typical to use a certain set of variables (characteristics of the borrower) that reflects the credit risk associated with the borrower. The construction of scoring models is based on statistical methods in which the qualitative and quantitative characteristics of a potential borrower are compared with the level of credit risk, which is determined on the basis of retrospective credit histories.

For each variable of the scoring model cut-offs are determined, according to which the scoring model divides borrowers into “bad” and “good”. For each scoring model the own cut-off determined, which reflects the boundary of vulnerability in relation to the bank's credit policy and external factors. The purpose of credit scoring is to calculate the level of credit risk inherent in one or another borrower, in other words, his credit rating. Comparison of the obtained results with the limit value allows providing borrowed funds to one or another borrower.

The result of the credit scoring is, as a rule, a certain integral indicator, which is proportional to the borrower's creditworthiness. Based on the received credit scoring estimates, the bank has the opportunity to classify borrowers by their level of creditworthiness. For example, see [2]. Unfortunately, there is currently no qualitative statistical database on borrowers in Ukraine, and credit bureaus are not yet operational. Ukrainian banks have to rely on their own methods of assessing credit risk and take the full burden of credit risk.

## II. RESEARCH RESULTS

The source of information used for the practical implementation of scoring models is the characteristics of borrowers of one of the Ukrainian banks: borrower's age, gender, work experience, annual income, annual household income, the existence of borrower's own real estate, the purpose of the loan agreement (consumer credit, mortgage, credit for study fees, business development loan, car loan, credit card, building real estate lending), the average amount of payment for a loan, the amount of the last payment for a loan, the amount of the next payment for a loan, the frequency of payments on the loan, the discipline of payments, the amount of past due payments in the last year, the total amount of past due payments in credit history, the result of the loan agreement (loan was not repaid, loan was repaid). The SAS Enterprise Miner software product was selected for scoring models development. It is an integrated component of the SAS system of intellectual data analysis and designed to detect information in large amounts of data which is necessary for management decisions to be made. For example, see [3].

To construct scoring models, we used the SAS Enterprise Miner tools of regression analysis, decision trees and neural networks. The entire incoming data set (*TRAIN*) of 14559 borrowers was randomly divided into two parts (80% – training data and 20% – validation data) with the preservation of the distribution of the positive response (loan was repaid) and the negative response (loan was not repaid) of the target variable (the result of the loan agreement). Thus, all models were constructed and tested on equivalent data sets (Fig.1).

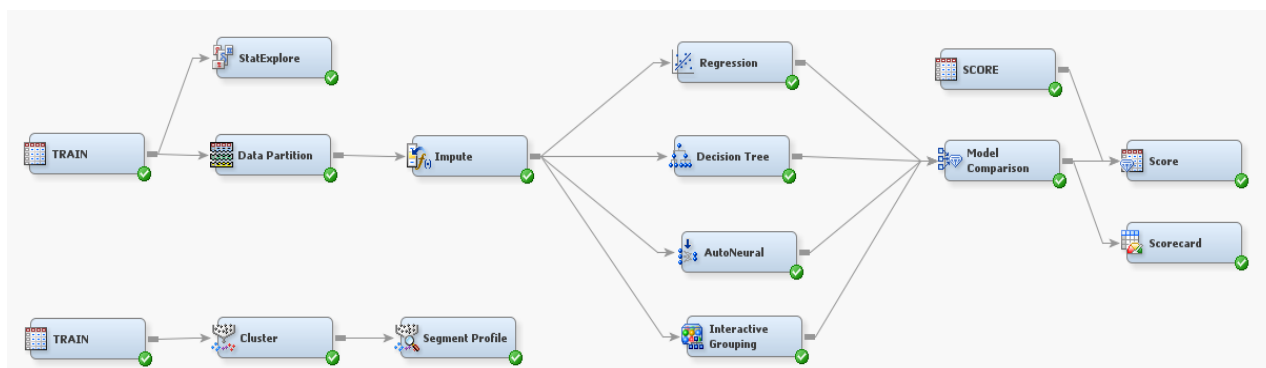


Figure 1. ETL-diagram of the simulation process of scoring of the bank's borrowers creditworthiness

Incoming data was divided into equal parts in percentage terms (*Data Partition* tool). They are given in table 1.

TABLE I. BREAKDOWN OF INCOMING DATA SET

	Number of credit order with positive response		Number of credit order with negative response	
	units	%	units	%
Incoming data	9 716	66,74 %	4 843	33,26 %
Training data	6 800	66,74 %	3 389	33,26 %
Validation data	2 916	66,73 %	1 454	33,27 %

As a result of the primary analysis (*StatExplore* tool), the basic statistical characteristics of the input data were obtained, the role of variables in the simulation was determined, and it was found that the input data array has missing values in the interval variables that were filled (*Impute* tool). Missed data was filled in this way: for interval variables all missed values were replaced by the average for all available values, for categorical input variables, all missed values were replaced by the most frequently encountered category

As a result of the cluster analysis carried out by the *k*-means method (*Cluster* tool), four clusters were obtained. The first cluster (19,7%) has been formed by senior citizens with significant (higher than the average) work experience (*Segment Profile* tool). The second cluster (34,9%) has been formed by borrowers without own real estate, with an average work experience, middle age and average income. The third cluster (39,1%) has been formed by borrowers with own real estate, with an average work experience, middle age and middle income. The fourth cluster (6,3%) has been formed mainly by borrowers with an income higher than the average.

To optimize the complexity of the logit model, the method of stepwise exclusion of non-significant

variables was selected (*Regression* tool). The significance of the latter were determined by us according to the statistical Wald chi-squared test. Result is given in figure 2.

Type 3 Analysis of Effects

Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
DelayAmountLastYear	1	913.0412	<.0001
DelayAmountTotal	1	32.0728	<.0001
Discipline	9	277.0781	<.0001
IMP_CreditorIncomeYear	1	4.4249	0.0354
IMP_FamilyIncomeYear	1	3.9292	0.0475
IMP_NextPaymentAmount	1	91.1914	<.0001
M_Age	1	7.1531	0.0075
M_AveragePaymentAmount	1	4.4490	0.0349
OfficeNumber	16	50.0821	<.0001
PaymentFrequency	5	31.5202	<.0001
Type	7	40.8102	<.0001

Figure 2. Significance of the logit model variables

It turned out that the following variables have high statistical significance: the purpose of the loan agreement (*Type*), the frequency of payments on the loan (*PaymentFrequency*), the amount of the next payment for a loan (*IMP\_NextPaymentAmount*), the discipline of payments (*Discipline*), the amount of past due payments in the last year (*DelayAmountLastYear*), the total amount of past due payments in credit history (*DelayAmountTotal*).

By evaluating the odds ratio, we have investigated how selected variables of the logit model affect the target variable. According to the results of a constructed logit model, a borrower with a minimum amount of past due payments in the last year, a senior age and with a maximum average amount of payment for a loan, who has paid payments every two months and received a credit card, is most likely a creditworthy borrower.

The decision tree was built in an automatic mode (*Decision Tree* tool). Optimization of the complexity of the decision tree was carried out using average squared error. Result is given in figure 3.

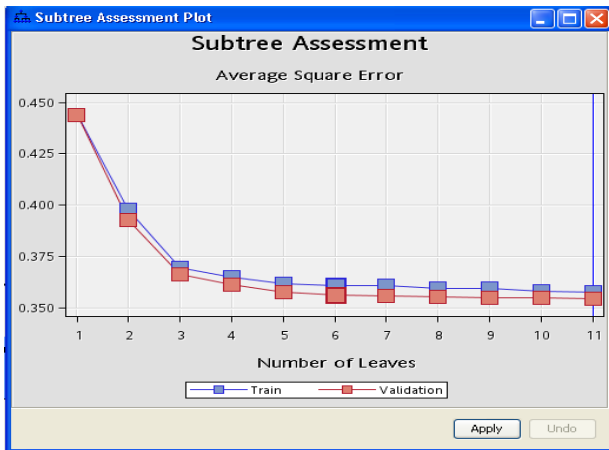


Figure 3. Graph of changes of average squared error

For the training and validation data sets, the graph has a declining trend. At the 6th step, the value of the average squared error starts to decrease less rapidly, and therefore further increase in the number of branches is not needed. Thus, as the optimal option, a tree with 6 branches of branching was selected. Classification rules for this model are given in table 2.

According to the results of the decision tree, the most likely creditworthy borrower is a client who had overdue payment to four months, paid without past due in the last year or repay the loan using a pledge.

The neural network was built in an automatic mode (*AutoNeural* tool), the hyperbolic tangent was chosen as an activation function. Optimization of the complexity of the neural network has been made on the basis of minimizing the proportion of incorrectly classified borrowers. The result is a neural network, which consists of one hidden layer with two neurons. The constructed neural network is characterized by low values of the average squared error (0,189 and 0,191 for training and validation data respectively). The misclassification rate of the model data also has low values, namely: 0,296 and 0,295, respectively. Thus, we can conclude that the built neural network is qualitative and adequate.

The choice of the best model was performed (*Model Comparison* tool) on the basis of the misclassification rate (*MISC*), the average squared error (*ASE*) and the Gini coefficient (*G*). Result is given in table 3.

The lowest values of the misclassification rate, the average squared error and the highest values of the Gini coefficient has a decision tree. In the second place – the neural network, the last place occupies the logit model.

Classification capabilities of the built models were also studied using the *Classification Table of Model Comparison* tool. Result is given in figure 4.

TABLE II. RULES FOR CLASSIFYING A DECISION TREE

Classification rule	Training data	Validation data
<b>The amount of past due payments in the last year (<i>DelayAmountLastYear</i>):</b>		
more than 5	100% borrowers not repayed loan	100% borrowers not repayed loan
less or equals 5	70 % borrowers repayed loan, 30% – no	70 % borrowers repayed loan, 30% – no
more than 1	64% borrowers repayed loan, 36% – no	64% borrowers repayed loan, 36% – no
less or equals 1	94% borrowers repayed loan, a 6% – no	93% borrowers repayed loan, 7% – no
<b>The discipline of payments (<i>Discipline</i>):</b>		
overdue payment more than 120 days, deferment of loan payment	64% borrowers repayed loan, 36% – no	66% borrowers repayed loan, 34% – no
overdue payment to 119 days, payment without past due last year, repayment on a loan using a pledge	98% borrowers repayed loan, 2% – no	98% borrowers repayed loan, 2% – no
<b>The amount of the last payment for a loan (<i>LastPaymentAmount</i>):</b>		
more than 1 403 hm	70% borrowers repayed loan, 30% – no	71% borrowers repayed loan, 29% – no
less or equals 1 403 hm	61% borrowers repayed loan, 39% – no	61% borrowers repayed loan, 39% – no
<b>The total amount of past due payments in credit history (<i>DelayAmountTotal</i>):</b>		
more than 11	59% borrowers repayed loan, 41% – no	59% borrowers repayed loan, 41% – no
less or equals 11	66% borrowers repayed loan, 34% – no	69% borrowers repayed loan, 31% – no

Model Node	Model Description	Data Role	Target Label	False Negative	True Negative	False Positive	True Positive
Tree	Decision Tree	TRAIN	Result	125	648	2741	6675
Tree	Decision Tree	VALIDATE	Result	58	292	1162	2858
Reg	Regression	TRAIN	Result	637	965	2424	6163
Reg	Regression	VALIDATE	Result	278	389	1065	2638
AutoNeural	AutoNeural	TRAIN	Result	21	524	2865	6779
AutoNeural	AutoNeural	VALIDATE	Result	8	240	1214	2908

Figure 4. Classification table

TABLE III. COMPARATIVE CHARACTERISTICS OF THE QUALITY OF THE LOGIT MODEL, DECISION TREE AND NEURAL NETWORK

Data	Coefficient		
	MISC	ASE	G
Logit model			
Training	0,30	0,19	0,41
Validation	0,31	0,20	0,37
Decision tree			
Training	0,28	0,18	0,46
Validation	0,28	0,18	0,46
Neural network			
Training	0,28	0,18	0,44
Validation	0,28	0,18	0,43

At the last stage of the study we attach the output of the *Model Comparison* tool and a set of score data (*SCORE*) into the *Score* tool (Fig.1). This tool generates forecasts using the model that was selected as the best by *Model Comparison* tool. In our case, this is a decision tree.

The SAS Enterprise Miner package includes specialized *Credit Scoring for Banking* solution that address specific credit scoring tasks. For example, see [4]. The component *Scorecard* automatically calculates scorecards based on the results of a model built on the training data. The *Interactive Grouping* component automatically selects the most significant input variables and interactively generates groups of values of input variables with continuous values (Fig.1).

The *Gini coefficient* and *Information Value criteria* were used to automatically select the most significant input variables. For the automatic formation of groups of values as criteria for the breakdown of the range of values into groups, the *Weight of Evidence* was used.

As a result of automatic implementation of the *Interactive Grouping* tool on the basis of the *Gini coefficient* and *Information Value criteria* from the input data set, it was suggested that 13 variables be used to form a scorecard. Each variable based on the weight of the *Weight of Evidence* has been broken down to the levels. Binary variables were automatically rejected (the role of *Rejected* is set) due to the fact that the *Information Value criteria* equals 0. Result is given in figure 5.

The final stage is the development of a scorecard using the *Scorecard* tool based on the results obtained at *Interactive Grouping* and *Model Comparison* tools (Fig.1). The main coefficients characterizing the quality of the built scorecard are given in the table 4.

Variable	Gini Statistic	Information Value	Level for Interactive	Calculated Role	New Role	Pre-Defined Grouping	Level	Label	Information Value Ordering
DelayAmountLastYear	37.938	0.883	NOMINAL	Input	Default			NOMINAL	1
Discipline	15.914	0.084	NOMINAL	Input	Default			NOMINAL	2
NextPaymentAmount	15.431	0.083	INTERVAL	Input	Default			INTERVAL	3
LastPaymentAmount	15.413	0.083	INTERVAL	Input	Default			INTERVAL	4
AveragePaymentAmount	15.365	0.082	INTERVAL	Input	Default			INTERVAL	5
DelayAmountTotal	10.82	0.046	INTERVAL	Input	Default			INTERVAL	6
PaymentFrequency	9.964	0.042	NOMINAL	Input	Default			NOMINAL	7
OfficeNumber	6.307	0.021	NOMINAL	Input	Default			NOMINAL	8
Type	4.999	0.012	NOMINAL	Input	Default			NOMINAL	9
CreditorIncomeYear	3.495	0.005	INTERVAL	Input	Default			INTERVAL	10
FamilyIncomeYear	3.384	0.005	INTERVAL	Input	Default			INTERVAL	11
WorkExperience	2.2	0.002	INTERVAL	Input	Default			INTERVAL	12
Age	1.277	0.001	INTERVAL	Input	Default			INTERVAL	13
Immovable	0.717		BINARY	Rejected	Default			BINARY	14
Gender	0.012		BINARY	Rejected	Default			BINARY	15

Figure 5. Selection of the most significant input variables

TABLE IV. COEFFICIENTS OF SCORECARD QUALITY

Coefficient	Sample	
	Training	Validation
Misclassification Rate	0,299	0,289
Average Squared Error	0,192	0,190
Kolmogorov-Smirnov criterion	0,269	0,289

According to the results of built scorecard, the most likely creditworthy borrower is a middle-aged borrower who has taken a consumer loan and has a maximum amount of the next payment for a loan, made a payment delay to the month during the past year or paid off loan using a pledge.

## CONCLUSIONS

The developed scoring models are primarily aimed at reducing the number of irreversible and “problem” loans, that is, the fulfillment by borrowers of the terms of the loan agreement. Using a scoring model for assessing a bank borrower’s creditworthiness based on a decision tree will provide an opportunity to achieve a range of effects: reducing the role of the subjective component when deciding on lending, significant acceleration of the decision-making process, improving the quality of the loan portfolio as a result of minimizing the share of problem loans. Thus, the simulation of credit scoring is one of the most effective ways to increase the efficiency of lending to bank customers.

## REFERENCES:

- [1] A. Kaminsky, “The expert model of credit scoring of the bank borrower”, *Banking*, no.1, pp.75-81, 2006.
- [2] G. Velikoivanenko, L. Trocoz, “Modeling of internal credit ratings of commercial bank borrowers”, *Economic analysis*, vol.11, no.1, pp.313-319, 2012.
- [3] Georges J., *Applied Analytics Using SAS Enterprise Miner Course Notes*. Cary, NC: SAS Institute Inc., 2010.
- [4] SAS Institute Inc. *Developing Credit Scorecards Using Credit Scoring for SAS Enterprise Miner 12.1*, Cary, NC: SAS Institute Inc., 2012.





**Наукове видання**

**СУЧАСНІ ІНФОРМАЦІЙНІ СИСТЕМИ І ТЕХНОЛОГІЇ**

**Матеріали  
Шостої міжнародної науково-практичної конференції  
(Суми, 16 – 18 травня 2018 року)**

Відповідальний за випуск  
Комп'ютерне верстання:

В. В. Шендрик  
О. В. Бойко

Формат 60×84/16. Ум. друк. арк. . Обл.-вид. арк. .

Видавець і виготовлювач  
Сумський державний університет  
вул. Римського-Корсакова, 2, м. Суми, 4007  
Свідоцтво суб'єкта видавничої справи ДК № 3062 від 17.12.2007