by Anhalt University of Applied Sciences and Perm National Research Polytechnic University

Proceedings of the 4th International Conference on Applied Innovations in IT

Volume 4

EDITION
Hochschule Anhalt

Proceedings of the 4th International Conference on Applied Innovations in IT

Volume 4

Koethen, Germany
March 10th 2016

Editors:

Prof. Dr. Eduard Siemens* (editor in chief),
Dr. Bernd Krause*,
Dr. Leonid Mylnikov**

(*Anhalt University of Applied Sciences,
** Perm National Research Polytechnic University)

This volume contains publications of the International Conference on Applied Innova-tions in IT (ICAIIT), which took place in Koethen March 10th 2016. The conference is devoted to problems of applied research in the fields of automation and communications. The research results can be of interest for researchers and development engineers, who deal with theoretical base and the application of the knowledge in the respective areas.

ISBN: 978-3-96057-013-4 (Online)
ISSN: 2199-8876

Additional copies of this publication are available from:
FB6 Anhalt University of Applied Sciences
Postfach 1458
D-06354 Koethen, Germany
Phone: +49 3496 67 2327
Email: e.siemens@emw.hs-anhalt.de
Web: http://icaiit.org

Content

Section 1. Automation and algorithms

Section 2. Communication Technologies

Section 3. Data Processing and Data Analysis

Section 4. Projects Management

An Approach to Assessment of the Technical Condition of Overhead Transmission Lines

Anton Petrochenkov
Perm National Research Polytechnic University – Electrical Engineering Faculty
Komsomolskiy Prospect 29, 614990, Perm, Russia
E-mail: pab@msa.pstu.ru

***Abstract*—The approach of estimating the current state of the overhead transmission lines is considering. The performance functions of the overhead transmission lines has been generated on the basis of experimental data and reports. Results of simulation of the approximation functions for overhead transmission lines are analyzed.**

***Keywords:* overhead transmission lines, technical condition, fractional factorial experiment.**

I. Introduction

This research package is one of the part of the decision-making informational support system of the electrotechnical equipment life cycle management based of energy-information model [1][2].

Given package is aimed at achieving the following goals:

- Assurance of tolerance of power supply systems on the basis of data-analytical decision-making environments (the task should be considered in modern monitoring providing methodology complex, diagnostics and maintenance).
- Simulation and optimization of power systems, assurance of operational reliability of power complexes.
- Organizations of operating control of electrical power networks.
- Management of life cycle of electric equipment networks on the basis of modern methods of CALS-technologies [3].

Control of efficient operation of electrical engineering systems (EESs) involves numerous factors that must be taken into consideration. The factors (parameters) to be considered must include only those that can be really controlled or varied during operation at enterprises of the branch in question [2, 4, 5].

II. Probabilistic Assessment Of Failures

Probabilistic assessment of failures is applicable predominantly to the overhead transmission lines (OLs) [2][6]. Let us consider a OL as a sequence of links, i.e., lines and towers. The operating characteristic of a link is a reduced dimensionless quantity that considers the wire strength, the time factor, operating conditions, etc.

The function of the OL distribution is characterized by the equation [2][7]

$$F(x) = P(l < x),$$

where l is the current value of the OL operating characteristic [7]:

$$l = \min\{l_1, l_2, \ldots l_z\},$$

and z is the number of the links of which the line is comprised.

The exponential law

$$F(x) = \begin{cases} 1 - e^{-\alpha z(x-l)} \\ 0, \ x \le 1 \end{cases} \quad (1)$$

is taken as the probability law where α is the generic parameter equal to the value of the response function of the line the operating characteristic of which is minimal [6][8][9].

The current value of the operating characteristic of every OL link is found by the formula

$$l_i = 1 - e^{-\gamma_i (1-T_i)^2}, \quad i = \overline{1,z}, \quad (2)$$

where γ is a generic parameter that considers different factors (the value of the response function can be used as this parameter), T_i is a parameter that considers the in-service time of the ith unit, $T_i = t_i$ is the current in-service time of the ith unit, and t_i max is the maximum in-service time of the ith unit [7].

III. Assessment Of The Technical Condition Of Overhead Transmission Lines

Assessment of the technical condition of overhead transmission lines is characterized by the following parameters [10]-[12]:

$$X_{\text{OL}} = \{x_1, \ldots, x_4\},$$

where x_1 and x_2 are the deflections of a tower from the vertical line along and across the OL, x_3 is the factor of defectiveness, and x_4 is the excess temperature.

The objects of the assessment of the OL technical condition are

$$O_{\text{OL}} = \{O_1, O_2\}, \quad (3)$$

where O_1 is the maximum security and O_2 is the minimum time consumed to replace or repair the parts [7][11].

An expert arrangement of the objects' ranks for assessment of the OL technical condition is presented in Table 1 (number of the expert $N = 5$).

TABLE 1
EXPERT ARRANGEMENT OF OBJECTS' RANKS TO ASSESS OL TECHNICAL CONDITION

Expert number	**Objects' ranks**	
	r_1	r_2
1	1	2
2	2	1
3	1	2
4	2	1
5	1	2
R_i	R_1=7	R_2=8

Processing of rank matrix $[r_{id}]$ allows for the weight of each parameter to be assessed as

$$v_i = \frac{v_i'}{\sum_{i=1}^{N} v_i'}, \quad (4)$$

where

$$v_i' = 1 - \frac{R_i}{nN} + \frac{1}{n},$$

and n is the number of the parameters in the list [13].

The values of the weights of each object in the assessment of the OL technical condition calculated by Eq.(4) are summarized in Table 2.

TABLE 2
EXPERT EVALUATION OF WEIGHTS OF OBJECTS TO ASSESS OL TECHNICAL CONDITION

Object	**Object designation**	**Object designation**	**Weight value**
Maximum security	O_1	v_1	0.53
Minimum time expended for replacement and repair	O_2	v_2	0.47

Let us, applying the well-known approach to carrying out a fractional factorial experiment [7], construct for the electrical equipment of various types the response functions in the polynomial form as

$$y = \beta_0 x_0 + \sum_{i=1}^{n} \beta_i x_i,$$

where β_0, and β_i are the polynomial coefficients, x_0 is a dummy parameter (factor), $x_0 = 1$, x_i is the ith parameter in the list, and n is the number of the parameters in the list.

To assess the parameters of a OL, a fractional factorial plan of the type 2^{4-1} is used that is set by the generating relation

$$x_4 = x_1 x_2.$$

A matrix of the plant to assess the OL technical condition is constructed as

$$\overline{\mathbf{D}}_{4-1}^{\circ} = \begin{array}{c} \begin{matrix} x_0 & x_1 & x_2 & x_3 & x_1x_2 \end{matrix} \\ \begin{pmatrix} +1 & -1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 & -1 \\ +1 & -1 & +1 & -1 & -1 \\ +1 & +1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 & +1 \\ +1 & +1 & -1 & +1 & -1 \\ +1 & -1 & +1 & +1 & -1 \\ +1 & +1 & +1 & +1 & +1 \end{pmatrix} \end{array}.$$

The normalized permissible values of the factors that determine the technical condition of an OL are presented in Table 3.

Determination of the factor levels ranges that determine OL technical condition is presented in Table 4.

For objects (3) effective achievement is passed:

1. The indicator for object O_1 is categorization. Factors (parameters) are assigned to the following categories (Table 5).

The values of categories of different variants in a fractional factorial experiment for object O_1 are presented in Table 6.

The effective achievement of object O_1 for each of variants is presented in Table 7.

TABLE 3
PERMISSIBLE VALUES OF FACTORS THAT DETERMINE OL TECHNICAL CONDITION

Designation	Factor	Permissible values	Optimal value
x_1	Deflection of tower from the vertical axis along the OL	1:150 - 1:75	112.5
x_2	Deflection of tower from the vertical axis across the OL	1:150 - 1:75	112.5
x_3	Factor of defectiveness	1.2-1.5	1.35
x_4	Excess temperature	5°-10°	7.5

TABLE 4
DETERMINATION OF THE FACTOR LEVELS RANGES THAT DETERMINE OL TECHNICAL CONDITION

Low level (-1)	Middle level (0)	High level range (+1)		Middle level (0)	Low level (-1)
>150	150-116.25	116.25	108.75	108.75-75	<75
>150	150-116.25	116.25	108.75	108.75-75	<75
<1.2	1.2-1.335	1.335	1.365	1.365-1,5	>1.5
<5	5-7.25	7.25	7.75	7.75-10	>10

TABLE 5
FACTOR'S INDICATORS FOR OBJECT O_1

Factor	Category
x_1	2
x_2	2
x_3	2
x_4	1

TABLE 6
THE VALUES OF CATEGORIES OF DIFFERENT VARIANTS IN A FRACTIONAL FACTORIAL EXPERIMENT FOR OBJECT O_1

Variant number	Category
1	2
2	1
3	1
4	2
5	2
6	1
7	1
8	0

TABLE 7
THE VALUES OF EFFECTIVE ACHIEVEMENT OF OBJECT O_1 FOR DIFFERENT VARIANTS IN A FRACTIONAL FACTORIAL EXPERIMENT

Variant number	Effective achievement
1	0.5
2	0
3	0
4	0.5
5	0.5
6	0
7	0
8	1

2. The indicator for object O_2 is time of troubleshooting or repair of equipment for a given parameter.

The recovery time (in days) for the elimination of invalid values for each factor is shown in Table 8.

TABLE 8
RECOVERY TIME FOR THE ELIMINATION OF INVALID VALUES FOR EACH FACTOR THAT DETERMINE OL TECHNICAL CONDITION

Factor	Recovery time
x_1	5
x_2	5
x_3	20
x_4	10

Then, assuming that the equipment at fault will be corrected at the same time, the recovery time for different variants will be as shown in Table 9 [10, 11].

TABLE 9
VALUES OF RECOVERY TIME FOR DIFFERENT VARIANTS IN A FRACTIONAL FACTORIAL EXPERIMENT

Variant number	Recovery time, days
1	20
2	20
3	20
4	20
5	5
6	10
7	10
8	0

The effective achievement of object O_2 for each of variants is presented in Table 10.

TABLE 10
THE VALUES OF EFFECTIVE ACHIEVEMENT OF OBJECT O_2 FOR DIFFERENT VARIANTS IN A FRACTIONAL FACTORIAL EXPERIMENT

Variant number	Effective achievement
1	0
2	0
3	0
4	0
5	0.25
6	0.5
7	0.5
8	1

The integrated assessment of the efficiency is calculated by the arithmetic mean form as

$$E_{a,j} = \sum_{i=1}^{s} v_i e_{ji} ,$$

where j is the variant number and s is the number of objects, estimate e_{ji} reflects the degree of achieving object O_i when implementing variant j; it is given in the range from 0 to 1 [13].

The results of calculating are presented in Table 11.

TABLE 11
SUMMARY TABLE OF EFFECTIVE ACHIEVEMENT OF OBJECTS O_1 AND O_2 IN A FRACTIONAL FACTORIAL EXPERIMENT

Variant number	v_1 (0.53)	v_2 (0.47)	E_a
1	0.5	0	0.265
2	0	0	0
3	0	0	0
4	0.5	0	0.265
5	0.5	0.25	0.3825
6	0	0.5	0.235
7	0	0.5	0.235
8	1	1	1

The coefficients of the response function $\{\beta i\}$ are defined by the least-squares method as

$$\hat{\beta}_0 = \frac{1}{k}\sum_{u=1}^{k} y_u \ , \ \hat{\beta}_j = \frac{1}{k}\sum_{u=1}^{k} x_{ju} y_u \ ,$$

where u is the number of the observation and

$$\{\beta_{OL}\} = \begin{pmatrix} 0.2978 \\ 0.0772 \\ 0.0772 \\ 0.1653 \\ 0.2391 \end{pmatrix}.$$

Thus, the response function for an OL has the form

$$Y_{OL}(x) = 0.2978+0.0772x_1+0.0772x_2+0.1653x_3+0.2391x_4. \quad (5)$$

Current parameters' values are stored in the database of EES' energy-information model [1].

The values of operating characteristic l for every OL assembly are found by substituting the obtained response function values into Eq.(2). The least of the obtained operating characteristic values of the assemblies is introduced into Eq.(1). Setting the required operating characteristic level, the operator calculates the probability of the failure of the OL in different time periods.

E.g., for OL "AS-95" (Voltage 35 kV, section 95 mm) with length of 11,5 km for one of EES the values of parameters (factors) that determine OL technical condition in different time periods are presented in Table 12 [10][11].

According to Table 4 these values correspond to the levels shown in Table 13.

Calculation of values of the response function $Y_{OL}(x)$ for OL "AS-95" at the time periods t_1 … t_{12} is produced by the Eq.(5) on the basis of the data in Table 12.

Further, according to the Eq.(2) operating characteristic of OL l is calculated. The results are summarized in Table 14.

TABLE 12
VALUES OF FACTORS THAT DETERMINE OL "AS-95" TECHNICAL CONDITION IN DIFFERENT TIME PERIODS

Factor	Values of factors in different time periods						
	t_0	t_1	t_2	$t3$	$t4$	$t5$	t_{12}
x_1	112.5	115	120	125	130	135	145
x_2	112.5	112	125	140	150	160	160
x_3	1.35	1.4	1.48	1.6	1.6	1.6	1.65
x_4	7.5	8.3	8.5	8.5	9	10	10

TABLE 13
ASSIGNMENT OF THE LEVELS TO THE FACTOR'S VALUES THAT DETERMINE OL "AS-95" TECHNICAL CONDITION

Factor	Values of factors in different time periods						
	t_0	t_1	t_2	t_3	t_4	t_5	t_{12}
x_1	1	1	0	0	0	0	0
x_2	1	1	0	0	0	-1	-1
x_3	1	0	0	-1	-1	-1	-1
x_4	1	0	0	0	0	0	0

TABLE 14
VALUES OF THE RESPONSE FUNCTION AND OPERATING CHARACTERISTIC OF OL "AS-95" IN DIFFERENT TIME PERIODS

Parameter	Values of parameters in different time periods						
	t_0	t_1	t_2	$t3$	$t4$	$t5$	t_{12}
Y_{OL}	1	0.4522	0.2978	0.1325	0.1325	0.0553	0.0553
l	1	0.335095	0.21433	0.091292	0.081304	0.030627	0.008809

We assume that all the OL's nodes are overriding and secondary, i.e., ranks for them are shown in Table 15.

TABLE 15
RANKS ARRANGEMENT FOR THE NODES OF THE OL "AS-95"

Number of node	1-10	11-20
Rank's value *r*	1	2

Then for specified operating characteristics we obtain the following values of the probability of failure of the OL (Table 16).

Dynamics of the values' changes of failure probability $F(x)$ of the OL "AS-95" for different operating characteristic's level x levels of reliability x is shown in Figure 1 (when considering the area with the number of nodes $z = 20$).

The dependences of the probability of failures when considering a lower number nodes ($z = 5$) and a large number ($z = 100$) nodes, calculated in a similar manner are shown in Figures 2 and 3.

TABLE 16
VALUES OF THE PROBABILITY OF FAILURE OF OL "AS-95" IN DIFFERENT TIME PERIODS

Operating characteristic's level, *x*	**Probability of failure of OL in different time periods**						
	t_0	t_1	t_2	t_3	t_4	t_5	t_{12}
0.1	0	0	0	0.008099	0.017308	0.065219	0.084839
0.2	0	0	0	0.096536	0.104923	0.15182	0.169622
0.3	0	0	0.070548	0.177087	0.184727	0.230397	0.24655
0.4	0	0.04951	0.146625	0.250456	0.257415	0.301695	0.316352
0.5	0	0.121036	0.216476	0.317285	0.323623	0.366388	0.379687
0.6	0	0.187179	0.280609	0.378154	0.383927	0.425088	0.437154
0.7	0	0.248345	0.339493	0.433597	0.438855	0.478349	0.489297
0.8	0	0.304908	0.393556	0.484097	0.488886	0.526676	0.53661
0.9	0	0.357214	0.443195	0.530094	0.534456	0.570526	0.57954
1	0	0.405585	0.488771	0.57199	0.575963	0.610314	0.618492

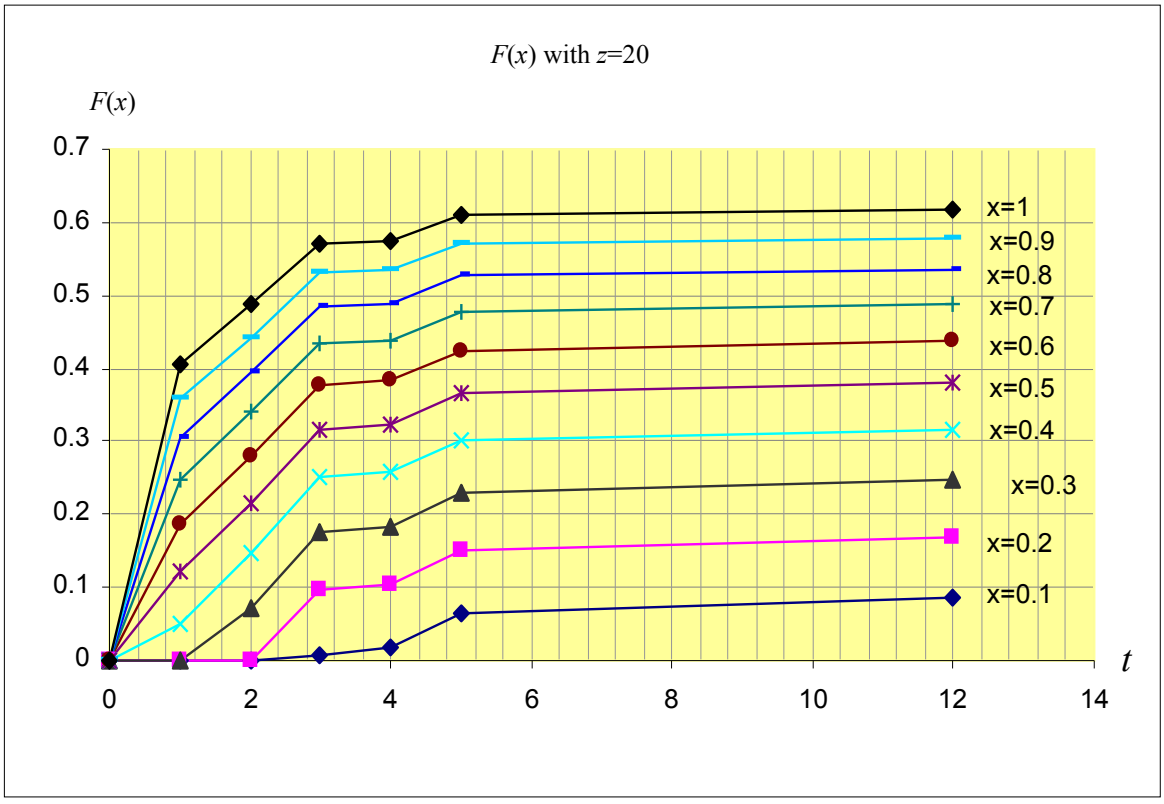


Fig. 1. Values of failure probability $F(x)$ of the OL "AS-95" when considering the area with the number of nodes $z = 20$.

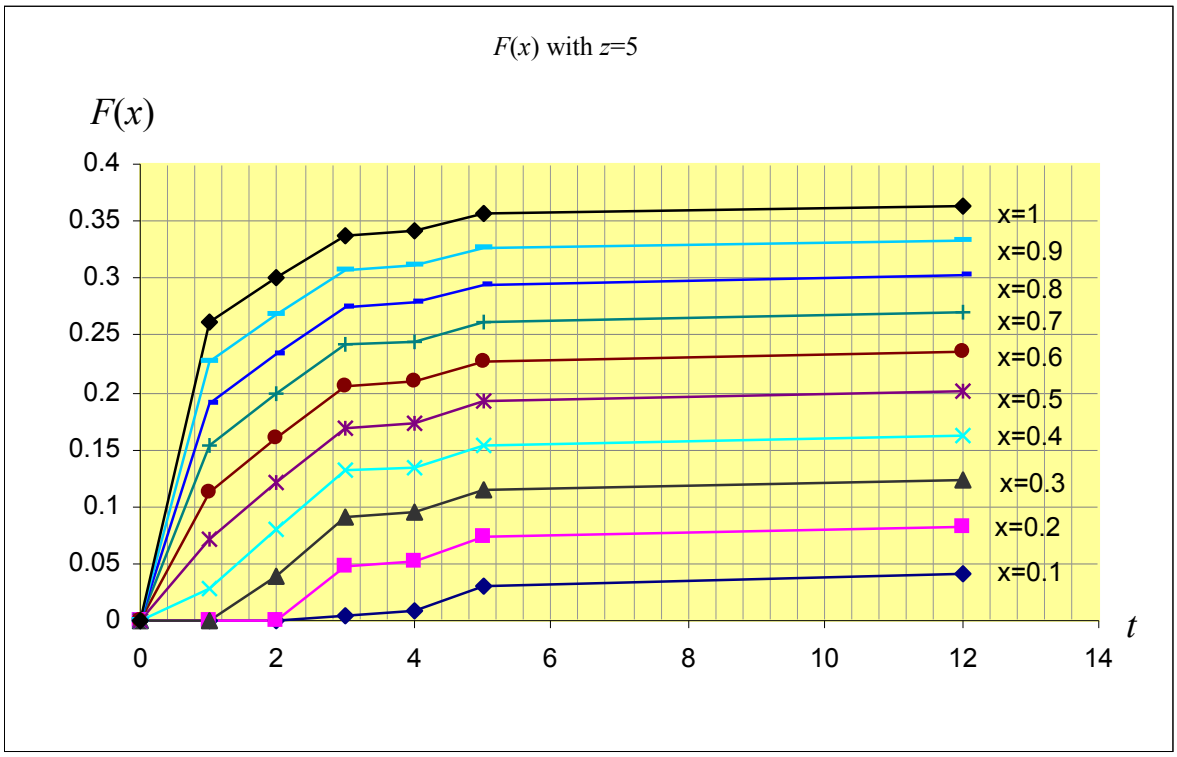


Fig. 2. Values of failure probability $F(x)$ of the OL "AS-95" when considering the area with the number of nodes $z = 5$.

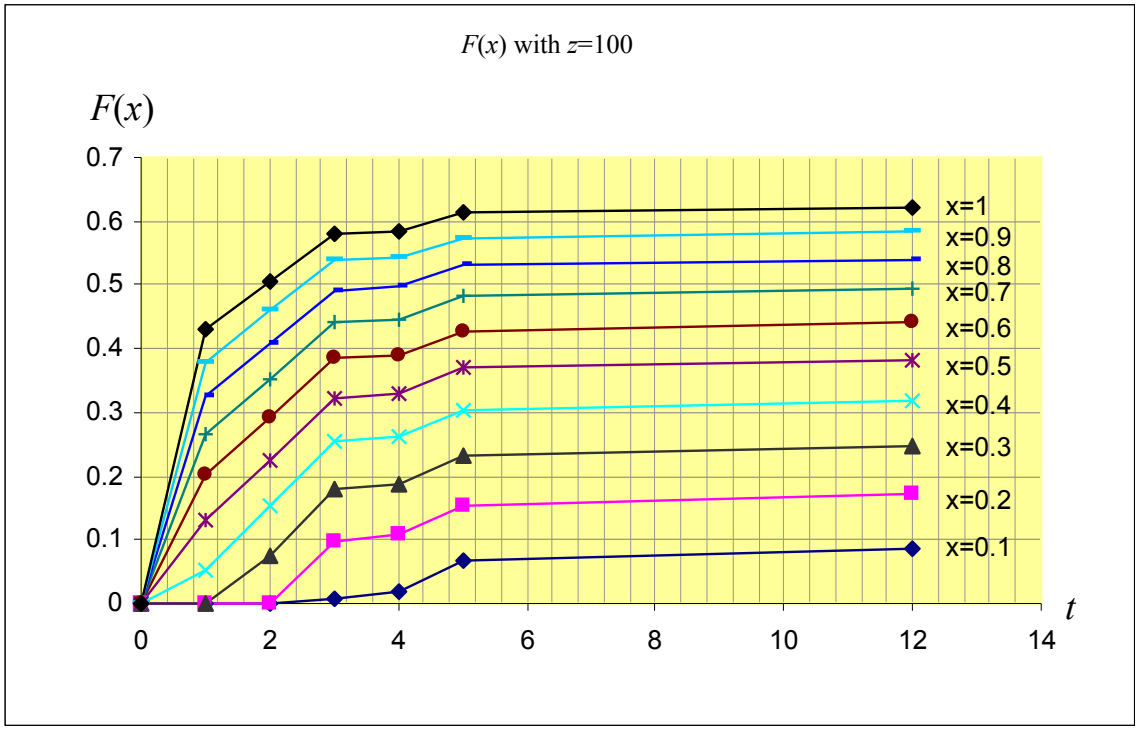


Fig. 3. Values of failure probability $F(x)$ of the OL "AS-95" when considering the area with the number of nodes $z = 100$.

IV. Analysis Of Simulation Of Approximation Functions For Overhead Transmission Lines

The problem of finding analytical dependences of the above described performance functions of the electrical equipment corresponds to the definition of the interpolation problem. On the basis of the resulting curve, the form of the approximation function is determined from a number of analytical functions with simple forms [7][14].

The mathematically best approximation entails a choice of the goodness measure, which is the residual function of nodal points and the values of the approximation function as:

$$J = \sum_{i=1}^{n} (F(x_i) - y_i)^2 \to \min$$

where y_i is the tabulated value of the assumed function at point x_i, $F(x_i)$ is the approximation function value at point x_i, and i is the number of the points, $i = \overline{1,n}$.

Analysis of the simulated approximation functions has shown that the least approximation error is obtained by the hyperbolic function $F(x) = a / x + b$ for the overhead transmission lines [7].

V. CONCLUSION

Considering the problem of usage of the information about electrotechnical equipment technical state for further repair work, it is necessary to select following levels of adequacy of estimations.

The first level – defining operating conditions according to reliability index, i.e. parameter of failure flow or reconstruction intensity.

The second level – defining technical state of the product according to probable defect characteristics and the damages revealed in a certain time.

The third level – state defining according to continuously controlled process variable, defining operating conditions of equipment elements.

According to given scheme, the basic task adds up to taking into account and dataflow management, providing data system operation [15][16].

Works on this direction are conducted within the Russian Foundation for Basic Research Grant of Russia No 14-07-96000 “Development of an intellectual decision support system to ensure of energy facilities trouble-free operation”.

REFERENCES

[1] A. B. Petrochenkov, “An Energy-Information Model of Industrial Electrotechnical Complexes”, Russian Electrical Engineering, 2014, Vol. 85, No. 11, pp. 692–696, doi: 10.3103/ S1068371214110108.

[2] A. B. Petrochenkov, “Management of effective maintenance of the electrotechnical complexes of mineral resource industry's enterprises based on energy-information model”, IEEE Conference Publications / Soft Computing and Measurements (SCM), 2015. P.122-124, doi: 10.1109/SCM.2015.7190430.

[3] A. Petrochenkov, “Methodical Bases of the Integrated Electrotechnical Complexes Life Cycle Logistic Support”, Proc. of the First International Conference on Applied Innovations in IT, E. Siemens (editor in chief) et al. Dessau, Anhalt University of Applied Sciences, 2013. – P.7-11. doi: 10.13142/kt10001.02.

[4] E. V. Cota, L. Gullo, R. Mujal, “Applying Design for Reliability to increase reliability confidence”, Proc. of Annual Reliability and Maintainability Symposium, 2014, art. no. 6798454, doi: 10.1109/RAMS.2014.6798454.

[5] N.A.J. Hastings, “Physical Asset Management”, 2015. doi: 10.1007/978-3-319-14777-2_3.

[6] R. Reinertsen, “Residual life of technical systems; diagnosis, prediction and life extension”, Reliability Engineering and System Safety, 1996, 54 (1), pp. 23-34. doi: 10.1016/S0951-8320(96)00092-0.

[7] A.B. Petrochenkov, “Performance functions for basic electrical equipment of the mineral resources industry”, Russian Electrical Engineering, 2015, Vol. 86, No. 11, pp. 634–639, doi: 10.3103/ S1068371215110103.

[8] S. V. Gurov, L. V. Utkin, “A continuous extension of a load-share reliability model based on a condition of the residual lifetime conservation”, European Journal of Industrial Engineering (EJIE), 2014, Vol. 8, No. 3, doi: 10.1504/EJIE.2014.060995.

[9] L. V. Utkin, I.O. Kozine, “Stress-strength reliability models under incomplete information”, International Journal of General Systems, 2002, 31:6, pp.549-568, doi: 10.1080/0308107021000061885.

[10] Standard OAO “LUKOIL” (Lukoil Standard) no. STP-01-026-2003: System for Servicing, State Control and Technical State Repairing of Oil Industry Equipment of JSC “Lukoil” Subcompanies, Perm, 2003. (rus)

[11] RD (Guiding Document) no. 34.45-51.300-97: Regulations for Electric Equipment Tests, Moscow: Atomizdat, 2001. (rus)

[12] B. Stott, O. Alsac, and A.J. Monticelli, “Security analysis and optimization”, Proc. IEEE, 1987, Vol. 75, No. 12, pp.83-111.

[13] L. V. Utkin, Yu.A. Zhuk, “Combining of judgments in imprecise voting multi-criteria decision problems”, International Journal of Applied Decision Sciences (IJADS), 2012, Vol. 5, No. 3, doi: 10.1504/IJADS.2012.047667.

[14] IEC 60300-3-1:2003 "Dependability management - Part 3-1: Application guide - Analysis techniques for dependability - Guide on methodology".

[15] A. B. Petrochenkov, “Regarding Life-Cycle Management of Electrotechnical Complexes in Oil Production”, Russian Electrical Engineering, 2012, vol. 83, No.11., pp.621-627. doi: 10.3103/S1068371212110090.

[16] A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance”, Mechanical Systems and Signal Processing, 2006, 20 (7), pp. 1483-1510. doi: 10.1016/j.ymssp.2005.09.012.

Improving the Energy Efficiency of Outdoor Lighting Systems Using the SmartLight Technology

Andrei Bachurin
Perm National Research Polytechnic University – Electrotechnical Department
Komsomolsky Ave, 29, 614990, Perm, Russia
E-mail: bachurin@ellips.ru

***Abstract*—This article describes the main approaches to the control of outdoor lighting using the modern means which give opportunities to improve the quality of lighting and improve its energy efficiency. It is about the lighting system of million city control.**

***Keywords*: control system, lighting system, energy efficiency, quality of lighting, control in lighting systems.**

I. Object of Study

a. Public outdoor lighting

The objects of study is outdoor lighting of Perm city. Its length is over 400 kilometers and it includes more than 35000 light points.

The biggest part of all light points are used in street lighting (Fig.1) [2].

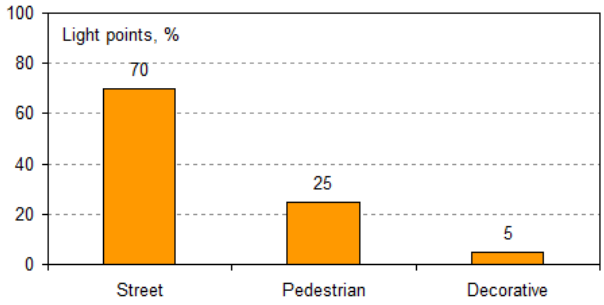


Fig. 1. The ratio of the urban lighting categories.

The greatest effect can be achieved by optimization of the street lighting, which can be divided into 2 groups by operating modes, lighting standards and illuminated objects (Table. 1) [2].

TABLE 1
TYPICAL LIGHTING MODES OF OBJECTS.

Objects	Lighting mode
Highways, Streets	The lights are switched on and off according to the schedule with adjustment for the current value of illumination, at night by low traffic density, the luminous flux can be reduced by 30% while decreasing traffic to 33% and by 50% while decreasing traffic to 20%. Depending on the object, there are different rates of light and illumination should not be reduced at least 4 lux.
Pedestrian crossings, Crossroads	Lighting should switch on and off simultaneously with the appropriate lighting of streets, but the change of the luminous flux at night is not allowed.

Currently, more t han 90% of lamps used in the street lighting comprise sodium lamp (HPS) with control ballast. Lamps with LED and electronic ballasts are present as a "pilot" elements, that demonstrate the development of the lighting system, but not a significant effect in improving the energy efficiency of the lighting system as a whole.

b. Modes of outdoor lighting

Outdoor lighting in the city of Perm operates according to a set of rules 52.13330 [1] and approved schedule on and off outdoor lighting (Table 2).

TABLE 2
SCHEDULE OUTDOOR LIGHTING IN THE CITY OF PERM

Month	January		February		March		April		May		June	
Day	On	Off	On	Off	On	Off	On	Off	On	Off	On	Off
1	18:08	10:22	19:08	9:42	20:15	8:37	21:27	7:06	22:36	5:43	23:56	4:24
2	18:10	10:21	19:10	9:39	20:18	8:34	21:29	7:03	22:39	5:40	23:58	4:22
3	18:12	10:20	19:12	9:36	20:21	8:31	21:31	7:00	22:42	5:37	0:00	4:20
4	18:14	10:19	19:14	9:33	20:24	8:28	21:33	6:57	22:45	5:34	0:02	4:18

The schedule includes two time periods during the day - when the light is switched off and when the light is switched on. Schedule time adjustment allows activating and deactivating of illumination and that's depending on the actual illumination, but no more than 15 minutes. The period when the light is on, is divided into 3 zones: "evening," "night", "morning", according to the intensity of traffic. Morning and evening traffic intensity is the maximum, at night - minimum. This separation makes it possible to control the lighting - to reduce the luminous flux "night" lamps, compared with the maximum luminous flux in the "morning" and "evening". Time of transition zone "evening" in the zone "night" (from the "night" in the "morning") is determined by the municipal administration, it is fixed and does not depend on actual traffic. At the moment, the time zone "night" is valid from 00:00 to 06:00. The lighting system uses electromagnetic and electronic ballasts. Control lamps with electromagnetic ballasts perform per phase

(lights powered by a single power point are divided into three approximately equal groups which are connected to different phases of the AC voltage network). Control lamps with electronic ballasts are done in groups (lights powered by a single power point form one group, from which the arbitrary subgroups can be distinguished).

TABLE 3
METHODS OF CONTROL DIFFERENT BALLASTS IN DIFFERENT AREAS (LIGHTING MODES)

Mode	"Per phase" control	Electronic ballast control
Day	Three phases is off	0% light
Morning (evening)	Three phases is on	100% light
Night	One phase is on	70% light

As it can be seen from Table 1, for some objects in the zone "night" (0:00 to 6:00) 100% luminous flux must be saved (could not be switched off to two phases or limit to 70% of the power consumption). To fulfill this requirement, this light points are connected to the "night" phase (in the case of equipment e.m.ballasts) or combined into a single group (if they are equipped with electronic ballasts).

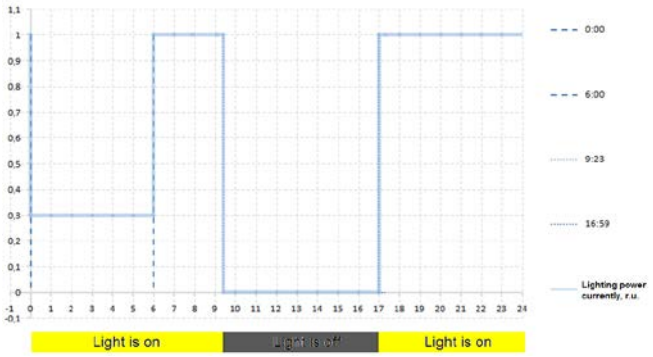


Fig. 2. Schedule outdoor lighting December 22.

Summary. There are three fixed time zones and in each of them the control of outdoor lighting gets a fixed scenario. Light towers are divided into groups based on their electrical connection (power point, phase) and destination (lighting street or crossroads). There are no individual control lamps. Also there are no ambient light sensors, sensors showing the presence of pedestrians and vehicles.

c. Advantages and disadvantages of the existing modes of outdoor lighting

The advantages of the existing modes of outdoor lighting include:

- ✓ Saving up to 50% of electricity in the "night" mode (0:00 to 6:00);
- ✓ Visual selection of crossroads and pedestrian crossings in the "night" mode (0:00 to 6:00);
- ✓ Saving resource of lamps and ballasts in two phases of 6 hours a day.

But there are serious disadvantages in these modes:

- Reduction in excess street illumination in "night" mode (from 0:00 to 6:00);
- Reduction of uniformity excess of in the "night" mode of distribution of pavement brightness (0:00 to 6:00);
- Increase of wear equipment in "night" phase (lamp, ballast, BFC);
- Reducing of the lighting reliability in the "night" mode (0:00 to 6:00) in mind, depending on one phase.

At the moment the merits of used modes (their financial side) significantly outweigh the disadvantages (light quality), so large-scale activities on the modernization of the lighting system and change of lighting modes, that do not provide the same level of budget savings (as at present), are doomed.

d. Implemented measures that improve the energy efficiency of outdoor lighting

There are many well-known ways to improve the energy efficiency of outdoor lighting [2].

The majority of the most common ones have been tried in the city of Perm:

1. Replacement of energy-unefficient lamps with the efficient lamps (90% of the lamps have been replaced on the HPS, LED replacement is not made due to their high cost);
2. Replacement e.m.ballasts with three-mode electronic ballast (replacement is slow, replaced by a small amount in mind the high initial costs and long payback period);
3. Installation of automatic control system into power points (100% of power points are already equipped with automated control systems);
4. On and off illumination in case of the actual illuminance (illumination control sensor installed, but no automatic impact on the system).

These activities are not allowed to abandon the existing algorithm of the lighting control, which has not set lighting parameters with the set of rules 52.13330, although it has significantly reduced power consumption compared to the previous level.

The main reasons for the lack of effectiveness of implemented activities are:

- High cost of equipment (LED, electronic ballasts);
- Failure to revise regulations governing the work of lighting;
- Failure to use 100% of new equipment functions (smooth control, automatic control).

The consequence of these factors is very large payback period of the new equipment, making these events economically ineffective and transforming them into the category of social.

II. SELECTING THE METHOD OF INCREASING ENERGY EFFICIENCY OUTDOOR LIGHTING

It is necessary to bring the lighting parameters with the set of rules 52.13330 and other regulatory documents and thus reduce the power consumption with respect to the current level.

There is a number of ways not previously used, improving lighting efficiency, which does not contradict the set of rules 52.13330 but it's technically feasible at this time.

a. Using multi-zone control of the luminous flux of outdoor lighting installations throughout the day.

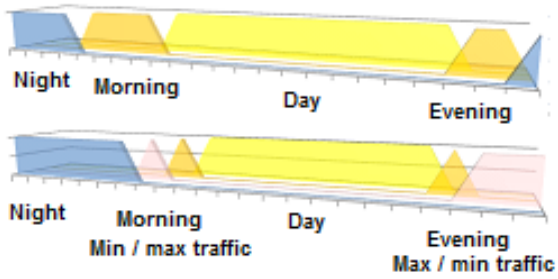


Fig. 3. The four and six-zone graphic works of outdoor lighting.

Time zone "evening" and "morning" is very long (from 2 to 7 hours). In such a large time interval volume of traffic varies very considerably [6]. Rationally it would share these time zones into multiple smaller sub-zones with less traffic scatter values in each.

b. Switching the lighting from "evening" in the "night" mode, and from "night" to "morning" mode during real decrease (increase) of traffic.

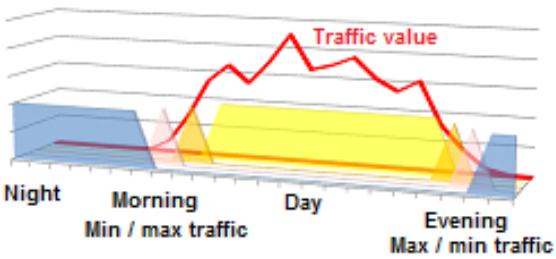


Fig. 4. Six-zones schedule outdoor lighting and value of traffic intensity.

The intensity of traffic has a certain dependence on the day of the week, time of the year, but every day it will have an individual trend. According to this trend, the switching lighting to "night" mode (illuminance decrease) will be carried out on different days at different times [7]. "Night" time zone becomes a "floating" rather than fixed one (from 00:00 to 06:00), and it becomes necessary to measure the value of traffic (number of vehicles per unit of time).

c. The use of dimming devices in lighting installations to reduce the luminous flux in night mode.

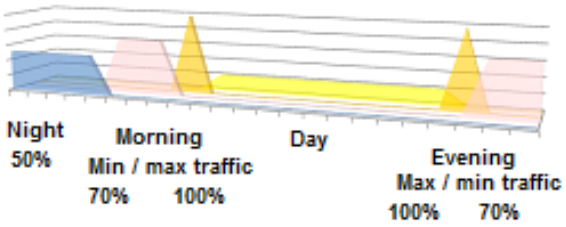


Fig. 5. Six-zones schedule outdoor lighting with different values of the luminous flux in different zones..

The use of dimming devices (electronic ballast), in addition to their other positive properties, is allowed for group control lamps with their union in arbitrary groups that are independent from the electrical wiring (phase). As a result, it becomes possible to reduce the luminous flux at the required luminaires without turning them off. This increases the uniformity of road brightness, removes overload "night" phase, it is possible to install an arbitrary value of the luminous flux from each group of luminaires.

d. Using multi-zone control of the luminous flux with different boundaries of time zones for different groups of lighting installations.

According to the set of rules 52.13330 streets are divided into 7 categories (A1, A2, B1, B2, C1, C2, C3) for the purpose of the valuation value of the average illuminance and the average brightness of the road surface. In addition to the various streets volume of traffic has individual character, i.e., at one time on the same street it is already necessary to start "night" lighting mode and the other must still work "evening" mode [4]. Thus, the lighting modes and the luminous flux must be individualized for each street (its homogeneous area) at a time [3].

e. Smooth (multi) transition from one zone to another on the basis of the value of the actual volume of traffic.

Above 6 time zones were allocated corresponding to the various modes of operation of outdoor lighting (light flux).

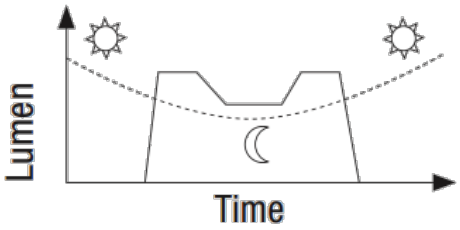


Fig. 5. Smooth change of light flux [5].

The transition from one zone to another is supposed to step, while the value of road traffic - is changing, and the possibility of electronic ballasts is allowed for a gradual decline of the luminous flux. Smooth (multi) transition from one lighting mode to another will improve the energy efficiency of lighting (for proactive move).

f. Exact simultaneous automatic switching on and off lights throughout the city depending on the ambient light value.

The amount of ambient light in different parts of the city at the same time can have a sufficiently large variation (up to 20 lux). This is determined by differences of weather conditions and different geographical longitude of a few city parts [2]. For precise simultaneous switching of the lighting it is necessary to measure the outdoor lighting at several points in the city and switching the lighting depending on average value of illuminance.

III. INTEGRATED USING KNOWN METHODS OF INCREASE OF OUTDOOR LIGHTING ENERGY EFFICIENCY

None of the methods discussed above, as well as already implemented, will not solve the problem **alone**.

Next, we will see **combined** application of these methods, i.e. control of the power level of outdoor lighting by dimming devices, simultaneous management of time borders of zones of different lighting power levels, depending on the traffic value and brightness, splitting power borders and time zones in different parts of the streets.

a. Exact simultaneous automatic switching on and off lights throughout the city depending on the ambient light.

Switch on or off the lighting automatically on real illumination will guarantee savings of 5 to 15 minutes of

burning every day, without the risk of exceeding the time-limit to switch on or off the lighting. Besides the quality of system performance will be improved - the lighting will be switched on even when it is needed.

Savings are achieved by precise lighting control for illuminance value and due to simultaneous operation of all power points that are not observed in the present manual (and semi-automatic) mode.

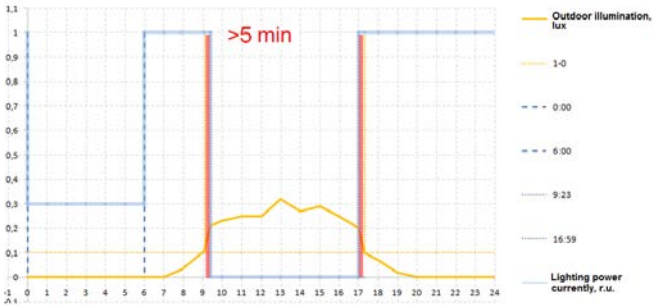


Fig. 7. Reduction of the total burning time when the lighting switching is based on the average data from the ambient light sensor.

b. Multi-zone (6 zones) regulation of the luminous flux depending on the actual value of the traffic fixed by sensors.

According to the set of rules 52.13330 p.7.43 and p.7.32 there is an allowed reduction of the luminous flux on the streets of categories A and B, if the actual illumination is more than 15 lux.

TABLE 3
ELECTRONIC BALLAST CONTROL METHODS IN 6 TIME ZONES OF THE DAY

Mode	Boundary condition of time zone	"Per phase" control	Electronic ballast control
Day		Three phases is off	All luminaries is off
Evening (morning) high traffic	illuminance lower than 20 lux	Three phases is on	100% light
Evening (morning) low traffic	Traffic lower than 34% of it maximal value	Two phases is on	70% light
Night	Traffic lower than 20% of it maximal value	One phase is on	50% light

The borders of time zones with different luminous flux will be determined on the basis of indications from the traffic sensors.

During the day the sensor continuously detects the intensity of the traffic as the ratio of the total time of the car motion for 30 minutes to 30 minutes. Here is a sample for a half an hour: if the car moving stream is 5 minutes, the sensor will value: 5/30 = 0.167. Resolution of calculating traffic intensity values shall be equal to 5-10 minutes, as with such an accuracy comes the definition of boundaries of time zones.

According to the set of rules 52.13330 p.7.43 it is necessary to fix the maximum value (MX) from the sensor during the day and calculate two borders: MX * 0,2 and MX * 0,33 (Figure 8).

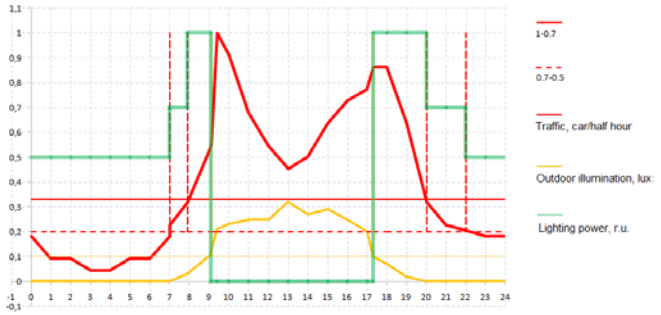


Fig. 8. Reduction of the total burning time when the light switching based on the average value from traffic sensors.

Obviously, depending on the traffic, time zone borders will be different for different parts of streets.

In addition, it should be taken into account that there are individual schedules of major social facilities (hypermarkets, schools, universities, etc.) And a decrease in the luminous flux is unacceptable at lighting of pedestrian crosses and crossroads.

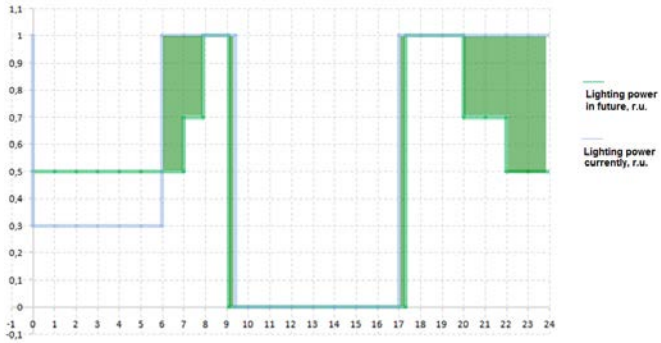


Fig. 9. Reduction of the total burning time when the light switching based on the average value from traffic sensors and outdoor illumination sensors.

These methods, with their complex application, are economically and socially efficient (reduce the amount of electricity consumed up to 50%, increase equipment life, improve lighting conditions and the uniformity of the lighting on the streets, the lighting is provided in accordance with the set of rules [1]). However, the exact calculation value of the effect is required.

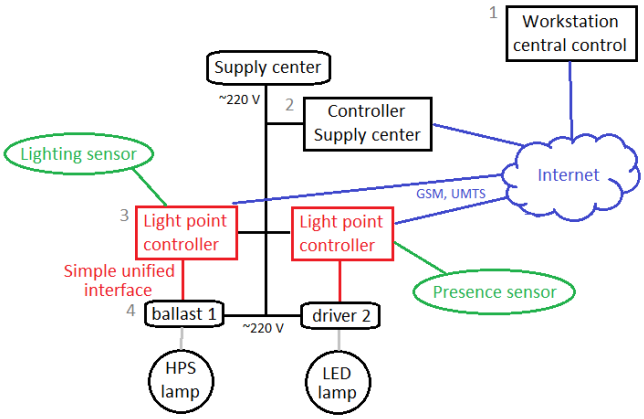


Fig. 10. Structure of lighting control system [2].

The use of these methods is impossible without the usage of electronic ballasts with individual control and use of the global automated control system, allowing control ballasts with algorithms given, when control modes for each light point vary depending on the ambient light, traffic density or pedestrians and others [8].

IV. CONCLUSION

The main reason for the low efficiency of existing modes of lighting is their static nature. In this paper, the main methods of increasing the dynamic modes of outdoor lighting are considered. Here the comprehensive utilization is proposed ways are discussed and a rough analysis of the energy of the event is made. It confirms the need for light points controllers.

As part of this work a micro layout of outdoor lighting systems is developed, which approves the proposed Article lighting control algorithms and hardware layout. Further research directions are also defined in terms of the development of outdoor lighting control systems, lighting control study of algorithms, calculations of economic efficiency.

ACKNOWLEDGMENT

I want to thank the employees of the companies ELLIPS, RADAR, Prof. Dr. E. Siemens for providing materials and support in this work.

REFERENCES

[1] SR 52.13330.2011 "Natural and artificial lighting - The updated edition of SNIP 23-05-95"

[2] A. Bachurin "City Outdoor Lighting System Control. Problems and Prospects" Proceedings of the 3rd International Conference on Applied Innovations in IT (2015). Jg. III. Koethen : Hochschule Anhalt, 2015, pp. 20-24

[3] DE102010049121 (A1) "Method for lighting e.g. road, involves switching on the lamp on detection of movement of person, and sending message to neighboring lamps through communication unit"

[4] S. Zinov, E. Siemens "The Smart Lighting Concept" Workshop on Problems of Autonomous Power Systems in the Siberian Region. Koethen : Hochschule Anhalt, 2013

[5] Light management systems. http://www.osram.com

[6] Folles E, Hogema J, Van der Horst R. 2004. Dynamic public lighting (DYNO). Report. Ministry of Transport, Public Works and Water Management. The Netherlands. [ILE] Institution of Lighting Engineers. 2006. Technical report 27-code of practice for variable lighting levels for highways.

[7] Jusle´n H, Hautala P, Holmstro¨m O, Saari M, Eloholma M, Halonen L. 2006. Experiences with intelligent road lighting. In: Proceedings of International Urban Conference. Athens. pp 236–238.

[8] Decentralized wireless control system outdoor lighting, Obraztsov S.A., Panfilov D.I., Light & engineering, 2012, № 1, pp 32-36.

The Problem Sensitization Robotic Complex Drilling and Milling of Sandwich Shells of Polymer Composites

Aleksandr Dudarev
Perm National Research Polytechnic University – Aerospace Department
Komsomolsky Ave. 29, 614990, Perm, Russia
E-mail: dudarev@pstu.ru

***Abstract*—In clause is given robotic a complex for drilling and milling sandwich shells from polymeric composites. The machining of polymeric composite materials has technological problems. At drilling sandwich shells there is a probability of destruction of a drill from hit of the tool in a partition. The system sensibilization robotic complex for increase of reliability of work of the cutting tool of the small size is offered.**

***Keywords:* Robot, Robotic Complex, Polymer Composite, Drilling, Milling, Sandwich Construction, Shell.**

I. INTRODUCTION

In all world it is known robotic complexes for manipulations of products, colouring, chemical processing, assembly-welding processes etc. In Russia robotic complexes on the basis of modern programmed robots having six degrees of freedom, for the decision of industrial tasks became distributed since 2003. The leaders among the manufacturers of robots for an industry were designated also: Kawasaki, Fanuc (Japan), Kuka (Germany), ABB (Sweden) etc. Till 2008-2010 years the application of robots for machining was exotic [1][2].

In 20 century the new class of constructional materials - polymeric composite materials has appeared which are widely applied in aircraft building, rocket production, shipbuilding. There are technological problems arising at machining polymeric composite materials [1][3][4].

With the expanded use of polymer composites, we need to develop the corresponding theory and organize their production, with appropriate technology, equipment, and tools.

In cutting polymer composites, we should note the following features [3][5]:

(1) Peeling and disintegration of the polymer composite (as a rule, at points of tool exit) on account of the poor adhesion of the filler to the binder.

(2) Difficulty in obtaining satisfactory surface quality (roughness), on account of anisotropy of the properties. Individual selection of the cutting conditions and tool is required.

(3) Low heat conduction of the material and correspondingly poor heat extraction from the cutting zone. (The tool absorbs 80–90% of the heat.)

(4) Intense tool wear due to abrasion by the solid filler. Mechanical and hydrogen induced wear are also present.

(5) Destruction of the polymer binder in cutting. Mechanical and thermal loads lead to chemical breakdown of the filler.

(6) Low productivity, on account of the low cutting rates. The use of lubricant and coolant fluid is limited, because the polymer composite absorbs moisture.

(7) Shrinkage on account of the elastic properties of the polymer composite. (The contact area is greater at the rear surface of the tool.)

(8) Specific safety requirements, associated with the release of toxic volatile particles of material on cutting.

The literature largely lacks systematic information regarding the machining of polymer composites.

The urgency of application of robots for machining polymeric composites in comparison to variants of application of machine tools explains to the following:

- Small forces of cutting, in comparison with processing metals, at drilling and milling of polymeric composites;
- The absence of large batches of let out products, that demands fast changeover the technological equipment;
- Occurrence of interfaces of control systems in robots at higher level, in comparison with machine tools. It allows to coordinate work of several components of the process equipment from the different manufacturers.
- Base cost robotic a complex is much less, than cost of multiaxis machines.

II. RELATED WORK

In whole all robots have anthropomorphic structure, geometrical similarity and technological similarity, and analog design [1][2][6].

On Fig. 1 shows the robots KUKA KR 60 HA, Kawasaki RS060N and ABB IRB 4600-60/2/05.

a) *b*) *c*)

Fig. 1. Robots:
a – KUKA KR 60 HA (Germany); *b* – Kawasaki RS060N (Japan); *c* – ABB IRB 4600-60/2/05 (Sweden)

Basic characteristics of robots - manipulators are given in Table1:

TABLE 1
CHARACTERISTICS OF ROBOTS-MANIPULATORS

Model		Kuka KR 60 HA		Kawasaki RS060N		ABB IRB 4600-60/2/05	
Country		Germany		Japan		Sweden	
Type		Articulated		Articulated		Articulated	
Number of axes		6		6		6	
Maximum reach, mm		2033		2100		2050	
Positioning repeatability, mm		±0,05		±0,07		±0,05-0,06	
Maximum total load, kg		60		50		60	
Range of motion software-limited		Range, °	Angular speed, radian/s	Range, °	Angular speed, radian/s	Range, °	Angular speed, radian/s
Number axis	Axis 1	±185°	2,23	±180°	3,14	±180°	3,05
	Axis 2	+35°, -135°	1,78	+140°, -105°	3,14	+150°, -90°	3,05
	Axis 3	+158°, -120°	2,23	+135°, -155°	3,22	+75°, -180°	3,05
	Axis 4	±350°	4,53	±360°	4,53	±400°	4,35
	Axis 5	±119°	4,27	±145°	4,53	+120°, -125°	4,35
	Axis 6	±350°	5,61	±360°	6,27	±400°	6,27
Maximum linear speed flange 6 axis, mm/s		No information		13400		No information	
Weight, kg		665		555		435	

For today at robots the existing characteristic of accuracy of positioning repeatability 0,05 mm at repeating of 0,05 mm (Table 1), and with the help of program calibration it is possible to achieve repeatability of 0,01 mm [7].

Existing problems of introduction of industrial robots for machining: difficult algorithm of programming, that demands the programmers mathematician of a high category.

Algorithm of programming robotic complex the following: creation 3D of model of object of processing a spelling of the managing programs in – CAM-SYSTEM for the processing centre, transformation of the managing programs for the processing centre in the managing programs for the robot. As it is visible, the circuit of programming robotic complex, as against the traditional processing centre, is longer on one step. Business in distinctions of degrees of freedom of the processing centre and robotic complex. The managing program is written initially independently for what - processing machine centre or robotic complex, and then for robotic complex exists special software adaptable the programs in the component programs for the robot.

For the sanction of industrial problems of manufacturing of sandwich construction, in view of technological features of processing polymer composites the tasks were put:

- Preliminary high-technology of a complex for machining products of the intricate geometrical detail form (5 axial processings) from polymeric composite materials;
- Application of technologies for drilling (punching apertures) and milling;
- Project of a technique of an estimation of integrity of the cutting tool during drilling (observation of breakages of drills of a small diameter at hit in partitions sandwich shells of polymer composites);
- Research of questions of the machine control behind a trajectory of movement of working bodies and opportunity of updating of positioning of the tool at coordinate processing;
- Creation of mathematical models of products as a cloud of points;
- Rational components of a complex ensuring necessary kinematics of processes,
- Release equipment for fastening products;
- System design of algorithms of the managing programs;
- Selection of furnishing working bodies of a complex: high speed a spindle, tool, systems of ventilation, auxiliary components of the adaptive control.

In [2][6], a robot created by the complex, allowing to carry out the perforation and milling operations in sandwich shells of polymer composite materials and the project is now implemented at JSC «Permsky zavod «Mashinostroitel»

After working through the complex technical solutions specification includes the following components of the complex:

- industrial robots Kuka KR 60 HA, load capacity of 60 kg;
- dust sealed cover of the robot;
- the control panel to the controller;
- computer software package for working with CAM-files;
- high-speed servo spindle 8 kW with a maxi mum rotational speed of 24000 rev / min;
- positioner single-axis (rotary table) with a vertical axis of rotation, carrying capacity of not less than 500 kg;
- system scanning laser sensors tracking the path of the tool relative to the work piece surface;
- control system zero point of the tool;
- a system of small-sized sensors tracking tool breakage;
- ventilation system with local suction and vacuum filter SET Coy;
- automatic tool changer magazine for ten instruments (replacement cartridges for the spindle, collet for various instruments (diameters of 2, 4, 6, 8, 10, 12, 16 mm), the tool;
- a device for securing the items;
- protecting fence and security locks with a mounting kit.

With the help of three-dimensional computer environment modeled robotic complex with all components and machined sandwich shell.

Robotic project of the complex is shown in in Fig. 2.

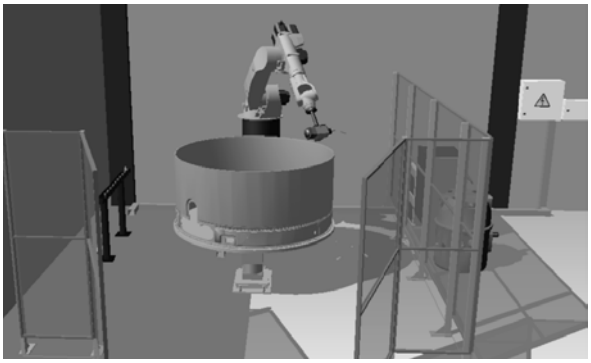


Fig. 2. Model 3-D of robotic complex

Dimensions and weight of processed parts: the diameter of 2500 mm; height of 1500 mm; Product weight up to 1000 kg.

The developed system allows to perform punching and milling in products such as sandwich shells.

Materials processed products: polymer composite, various non-metallic materials (plastic, wood, etc.).

The structure of the robotic complex includes active and passive safety systems, to prevent staff in the hazardous area of the industrial equipment.

The complex has a turntable (positioner KUKA KPF1-V500V2). The positioner works as an external axis of the complex. And as part of the robot-aided complex based KUKA robot, there are two options to connect the positioner:

- Asynchronous operation, thus, there is no mathematical connection with the robot;
- Mathematical connection with the kinematics of the robot.

An example of a mathematical relation is shown in Fig. 3.

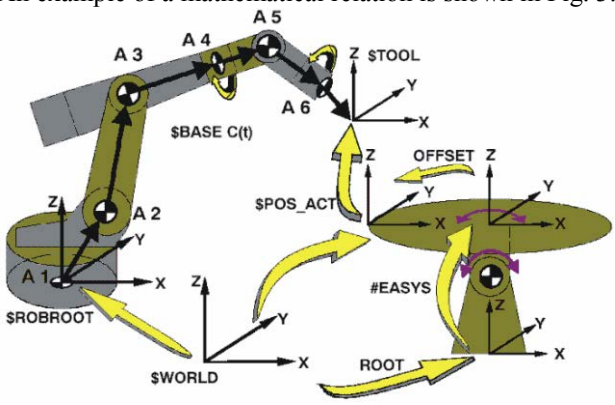


Fig. 3. Robot with external axes and extended kinematic system
Axes A1, A2, A3, A4, A5, A6 - mobile units. World coordinate system (World). Coordinate System tools (Tool). The coordinate system base (Base).

With connectivity options with a mathematical relationship, the drive continuously monitors the movement of the external axes. The mathematical relationship allows programming of complex processes.

Moving on each axis (Fig. 3), the robotic system can be a positive or negative direction.

When programming using three coordinates:

- Coordinate System World (World). Cartesian coordinate system with the on-roan reference at the base of the robot;
- Coordinate System tools (Tool). Cartesian coordinate system with the beginning of the report on the instrument;
- Coordinate System Base (Base). Cartesian coordinate system with coordinate basic origin report on the work piece.

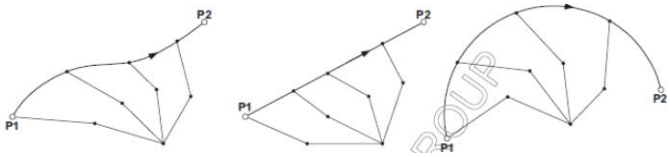
In practice, it is important that robotic system had been aligned in all three coordinate systems. Robotic system developed for the treatment of sandwich constructions connects and uses the base coordinate system and the coordinate system of the tool and for changing the cutting tool is used the world coordinate system (linking the base of the robot to the posts with the multipocket tool holder magazine for storage and tool change).

Another feature is the inclusion in the programming kind of movement of the robot, there are three:

a) Movement of the axes from point to point (PTP) (Fig. 4a). The tool moves on a fast path to the end point.

b) Motion for a linear path (LIN) (Fig. 4b). The tool is moved at a predetermined speed along a straight line.

c) Moving on a circular path (CIRC) (Fig. 4a). The tool moves with a given speed along a circular path.



a) b) c)

Fig. 4. Types of robot motion

Each of the three types of motion has submitted his interpolation speeds. All three motion graphics speeds leading axes of the robot consists of three phases: the phase of acceleration, constant phase, the phase of deceleration.

When choosing a software option (PTP) i. e from point to point, for each point on the path uses the maximum allowable speed.

When choosing a software option for linear motion tasks trajectory (LIN), the tool moves to the end point in a straight line while in motion on a trajectory is constantly changing from the initial position to the final orientation of the tool. This is achieved by turning and rotating the tool tip.

Similarly, when you select (CIRC) tool orientation in space is constantly changing from the initial position to the end.

When processing sandwich construction, namely, when drilling holes in sandwich construction possible breaking of the cutting tool of small size. The problem of prevention of tool breakage author dedicated to the work [8]. However, new technical solution to this problem is the implementation of sensitization, which is as follows.

Systems with proportional sensitization sensors allow automated processing drilling operation if it enters the walls of sandwich constructions. It requires dosing feed force based on an analysis of the situation in the zone of contact of the tool with the work piece. Consider a feature of such a system sensitization robot equipped with a force sensor, under the terms of technical operations necessary to adjust the feed rate

of the tool.

When you install the force sensor in the collet tool in the drill tool coordinate system has three degrees of freedom, so the sensor measures the moments about the three axes.

Drilling cycle begins with the supply of drill holes to a coordinate axis of the sandwich construction, to achieve a set point efforts to drill axis X (in the coordinate system tools TOOL). Due to the partial entry into the drill hole when the drill hit in the partition there is some microscopic skewed relative to the axis of the drill holes allowed by the sensor elements. From the ratio of forces generated while the Y and Z, we can determine the actual presence of the partition. Instantly, the feed rate of the drill is to slow down and be included in the work routine that tapping on the drill coordinate this with the transition to the main program loop work piece.

This system solves the very difficult technological problem and can significantly improve the reliability of the cutting tool. The sensitivity of the system to errors in the orientation of the axes of the holes may increase when using the six-component force sensor, which measures the three components of force along the coordinate axes and three points with respect to these axes.

Analysis of the complete picture of the interaction of the power to evaluate terms of contact with the drilling operations of the sandwich construction, aimed to organize the search for a match axes of the holes and the walls and avoid tool breakage.

III. MATHEMATICAL INTERPRETATION OF SENSITIZATION

An important sub-task in the system is the process of slowing down the robot drill feed, which is the use of so-called artificial compliance [9].

In the simplest case, when the interaction force F of the elastic sensor with the object, in this case, the isolating partition sandwich shell of the movement described by the equation of the second order

$$M\ddot{x} + k_4\dot{x} + F = k_3U\,, \qquad (1)$$

where x – coordinate translational movement; M – the reduced mass of the moving parts of the robot; U – controlled by the signal applied to the actuator; k_4 – coefficient of friction in the system; k_3 – coefficient of amplification in the system control loop.

If we denote the coordinate of the point of contact of the sensor with an elastic wall sandwich construction and, if we neglect the mass of the sensor, in accordance with Hooke's law determine the force F

$$F = k(x - x_0) \text{ at } x \ge x_0\,, \qquad (2)$$
$$F = 0 \qquad x \le 0\,.$$

The control law also assume a linear, such as

$$U = -k_1(F - F_{np}) - k_2\dot{x};\ k_1 > 0\,,\ k_2 > 0\,, \qquad (3)$$

where F_{np} – software enhancement of interaction of the sensor with the object formed by the control system of the robot; k_1, k_2 – coefficient of amplification.

Force F_{np} corresponds to a certain program coordinates defining the expression $x_{np} \ge x_0$, defined by the expression (2), i.e $F_{np} = k(x_{np} - x_0)$.

Thus, to create some control loop, which is not regulated by the position of the final level, and some effort to link the robot interaction with the environment. Indeed, the equation (1) with (2) and (3) can be transformed to the following vat:

$$M\ddot{x} + (k_4 + k_3k_2)\dot{x} + k_3k_1k(x_0 - x_{np}) = 0$$
provided $x \le x_0$, (4)

$$M\ddot{x} + (k_4 + k_3k_2)\dot{x} + k(x - x_0) + k_3k_1k(x - x_{np}) = 0$$
provided $x \ge x_0$. (5)

The term in $k_3k_1k(x_0 - x_{np})$ в (4) shows that $F_{np} > 0$ and $x \le x_0$ the system operates in a constant force proportional F_{np}. This means that the position of the hand, wherein the sensor probe is not in contact with the sandwich construction, is variable, and the link will move in a direction determined by the signs of the coefficients in (3). After reaching the contact with the sandwich construction in the drill begins elastic interaction with the treated surface of the work piece past the sensor probe in Step drilling layers (in the equation (5) is expressed in terms) $k(x - x_0)$).

Interaction is to achieve a predetermined force, if the partition will meet, and the full stroke of the drill into the work piece, if the partition is encountered. This corresponds to the normal action of a mechanical spring (sensor hardness) and "programmable" spring (in the equation (5) shown by the term $k_3k_1k(x - x_{np})$, of equivalent mechanical rigidity k_3k_1k).

Thus, the linear feedback force affects the movement unit in the same way as a conventional mechanical spring. The advantage of this spring is the ability to program control its stiffness by changing the ratio k_1. Those, when modifying the sandwich construction (thickness variation, variation in strength and other characteristics of the material) required minimum time for changeover robotic system.

In the position of static balance is provided by a constant

force F in contact equal $k_3 k_1 F_{np} /(1 + k_3 k_1)$.

The effect on the power of the drive servo system obtained based on the expressions (4) and (5) solves the problem of the sensitivity of the drive actuator - when spindle motion robotic complex that is necessary for the operations of machining of sandwich constructions. This feature allows you to set sensitizing robotic drilling walls and stiffening ribs of sandwich constructions made of polymer composite materials without sudden breakage of the cutting tool when need to dispense feed force of the cutting tool along the axis of the drill during drilling. In addition, the sensitivity of complex robotic prevent possible deformation of the sandwich constructions during milling.

IV. CONCLUSION

It showed a robotic system for drilling and milling sandwich constructions made of polymer composites. The main feature of the complex is developed that solves the proposed robotic complex technological problems arising in the processing of articles made of polymeric composites. When drilling sandwich constructions there is a risk of damage of the cutting tool in contact with the tool in the partition. The proposed system is complex robotic sensitization increases operational reliability of the cutting tool for small size. Sensing robot can protect the sandwich structure from undesirable deformation during milling windows, grooves or elements.

By development of tracking systems on force for robotic complex, it is necessary also to consider stability of all system at detailed elaboration of dynamic model, oscillatory processes and other effects of a control system. But it is themes of separate researches.

As the level of development of a robotics raises, constantly there are new opportunities of application of robots in manufacture. The specialized software are developed for new tasks of machining which raise a management efficiency and level of automation of preparation of the managing programs for robots.

In Russian Federation of analogues created robotic complex is not present. The basic foreign alternative variant is the system LASERDYNE SYSTEMS of the company PRIMA North America, Inc. (USA) used by corporation Boeing [1].

ACKNOWLEDGEMENT

[1]This work was supported by the Ministry of Education and Science of the Russian Federation (the contract №02.G25.31.0016) within the framework of Decree-governor of the RF №218 «On measures of state support of development of cooperation of Russian higher educational institutions and organizations implementing complex projects on creation of high-tech production".

REFERENCES

[1] A. S. Dudarev, «Problems of automation at performance of operations of punching of apertures insound-proof panels of air engines from polymeric composite materials», Engineering and automation problems, no. 3, pp. 63-68, Mar. 2012.

[2] A. S. Dudarev, V.I. Svirshchev, M.A. Bayandin, «The robotized complex for holes punching and milling aircraft engines sound-absorbing panels from composite polymeric materials», Automation and modern technology, no. 1, pp. 9-14, Jan. 2013.

[3] A. Phadnis, F. Makhdum, A. Roy, V. Silberschmidt, «Drilling in carbon/epoxy composites: Experimental investigations and finite element implementation», Composites Part A: Applied Science and Manufacturing, Vol. 47, pp. 41–51, April 2013.

[4] A. Faraz, D. Biermann, K. Weinert, "Cutting edge rounding: An innovative tool wear criterion in drilling CFRP composite laminates", International Journal of Machine Tools and Manufacture, Vol. 49, no. 15, pp. 1185-1196, 2009.

[5] M. Marcos, A. J. Gomez-Lopez, M. Batista Ponce, A. J. Salguero, «Roughness based study of milled composite surfaces», in Proc. Annals of DAAAM 2011, Volume 22, No. 1, pp. 0153-0154.

[6] A. Dudarev, «Machining automated system for drilling and milling of polymer composite materials», International Journal of Innovative and Information Manufacturing Technologies, no. 2, pp. 5-9, Feb. 2015.

[7] R. DeVlieg, T. Szallay, «Applied Accurate Robotic Drilling for Aircraft Fuselage», Electroimpact, Inc., U.S. Mukilteo, WA 98275, Jan. 2010. [Online]. Available: https://www.electroimpact.com/WhitePapers/2010-01-1836.pdf

[8] A.S. Dudarev, «Damage prevention methods of the small-sized cutting tool for holes shaping in sandwich structures of polymer composites», Automation and modern technology, no. 5, pp. 27-31, May 2013.

[9] A.I. Korendjasev, B.L. Salamandra, L.I. Tyves, «Theoretical Foundations of Robotics», vol. 2, S.M. Kaplunov, Ed. Moscow: Nauka, 2006. 376 p.

Electric Motor Life Estimate Based on Statistic Data

Anton Kaverin[1], Sergey Bochkarev[1], Sergey Buhanov[2]
[1]Perm National Research Polytechnic University - Electrotechnical Department Komsomolsky Ave. 29, 614990, Perm, Russia
[2]Bauman Moscow State Technical University – Computer Automation Systems Department ul. Baumanskaya 2-ya, 5, 105005, Moscow, Russia
E-mail: kaaperm59@gmail.com, bochkarev@msa.pstu.ru, bukhanov@yandex.ru

***Abstract*—Many different methods based on both planned inspection and health inspection for estimate of electrical equipment health are used. The estimation method of residual life of electric motors by their health in pulp and paper industry is considered.**

Keywords: **distribution function, gamma-percentile probability, lognormal distribution, residual life, technical maintenance, electric motor.**

I. Introduction

Any industrial plant is a complex of sophisticate technical products. Among them, electrical motors play important part. Maintenance expenses of electrical motors are a good proportion of total operating expenses of plants which use thousands of electrical motors. A technology of electric motors maintenance by their health is considered in this paper. Problems related with electric motor health estimate, timely detection of fault and pre-fault situations of electrical motors, residual life estimation are of critical importance. An operability estimation on basis of fault probability by means of finding out the defects and troubles is attained. That allows to exclude faults during the operational period of electrical equipment.

II. Key Features Related with Changeover to Maintenance Technology by Equipment Health

Changeover to maintenance technology by equipment health allows us to control current health of electrical motor and its repair quality, to optimize financial and operation expenses in service, to decrease demands in spare parts, to avoid of unexpected production suspension, to schedule periods and conditions of technical maintenance. Repair start and reconditioning volume by health of electrical motors are determined in service.

Using of various diagnostic methods in industrial operating conditions is limited because access to diagnosable object and mounting of control facilities are not always possible. Thus, methods based on analyzing the electric motors health statistical data have the advantage in practical realization of diagnostic systems. These methods allow to handle data without direct access to diagnosable electric motor and mounting of primary transducer next to it [1].

Consequently, developing the new diagnostics methods based on statistical data of electric motor is actual, scientific and technical problem.

III. Method Description and Problem Formulation

There is a huge class of electric equipment, which does not have complete and sufficient statistical data of reliability. That is highly-reliable and unique equipment without analogs [2]. The method considered is focused on data processing of certain amount of similar production equipment units. The equipment can have limited amount of faults during longtime period. Operation times of non-fault equipment for the same period are addition data about reliability of objects. These operations times are called censored ones.

Initial period of equipment operation is considered when statistical fault data is not accumulated. Even so, there is no possibility to apply more effective and accurate methods but there is a possibility to get initial estimate function shape of fault probability distribution of equipment by initial unit facts of equipment fault. Equipment data for using the method is collected. The data includes starting operation date of equipment, fault equipment date, health of equipment at the time of collecting data, and operation time of equipment in hours or days. Parameters of density probability distribution function by data collecting results are defined. Required reliability indexes of equipment are forecasted on the basis of the function.

An algorithm based on quantiles method is showed on Figure 1. The algorithm allows to gain efficient estimates at the earliest operation stages (under fault of first equipment). Estimates due to faults of each following equipment are detailed. The quantiles method deals with finding values of corresponding parameters, which ensure maximum credibility of curve shape in accordance with points received.

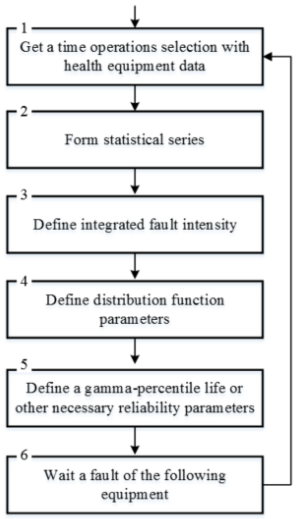


Fig 1. Algorithm of residual life equipment estimation.

Input data for calculation are operation times of each electric motor including fault ones. The operation time of electric motor when fault occurs is time interval from starting operation moment to the moment when electric motor is in fault condition. Operation data of all electric motors at following electric motor fault moment are fixed. The data can be changed inhomogeneously. The example of input data is showed in Table 1.

Actual electric motors health by various methods and ways can be determined. For example, methods based on analyzing the mechanical parameters such as electric motor vibration are used. These methods require proximate mounting the measuring equipment. Diagnostics of increased vibration causes of electric motor should be executed in the case of possible severe load. Furthermore, methods based on analyzing the electric parameters such us current, voltage, power demand and power delivered are used. These methods do not require mounting of special sensors at electric motor, but require special measuring equipment and they are used more often [3][4].

One-dimensional fault time vector (*Noi*) and additional vector of censored operation times (*Nci*) are plotted based on received data of electric motors health. Each next fault time point is fixed at the moment of the following electric motor fault. Operation time of electric motor is descended from censored vector to fault time vector under fault of the following electric motor. Both vectors cannot contain operation time of the same electric motor simultaneously. The example of one-dimensional fault time vector and additional vector of censored operation time are showed in Table 2.

Fault intensity of equipment under investigation is calculated after one-dimensional vector is obtained. Integrated fault intensity is calculated by the following equation:

$$p_i = \left(m_i^f\right)/(m_i^f + m_i^{c*}),$$

where

m_i^c – amount of fault objects with the time operation less than q_i, m_i^{c*} – amount of objects with time operation more than q_i.

Statistical series under each calculation based on data of all equipment operation times at various moments which correspond to these moments are formed. Both fault operation time and censored operation time can be taken as *i* point (Table 3).

Reliability parameters based on distribution function selection and its parameters are determined. The following distributions such as normal distribution, lognormal distribution, Weibull distribution are applied more often for technical objects in regulatory and scientific literature [5].

The distribution function shape of each type by parameters values is given in Table 4.

The distribution parameters search in solving the equation system by using of least square method is implemented:

$$\begin{cases} \int_0^{q_1} f(t,\mu,\sigma)dt - p_1 = a_1 \\ \vdots \\ \int_0^{q_n} f(t,\mu,\sigma)dt - p_n = a_n \\ \sum_{i=1}^{n} (a_i)^2 \to \min \end{cases}$$

where

f(t, μ, σ) – distribution function; pi – integrated fault intensity; αi – difference between theoretical and actual integrated fault intensities.

The following shapes of curves depend on distribution function type for one and the same values are showed on Figure 2.

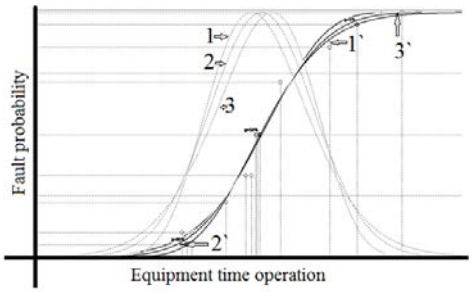


Fig 2. Example of distribution curve: 1 – density of lognormal distribution; 2 – density of normal distribution; 3 – density of Weibull distribution, 1' – integral function of lognormal distribution, 2' – integral function of normal distribution, 3' – integral function of Weibull distribution.

The next to last stage of determining the residual life of equipment is calculation of gamma-percentile indexes of life. The gamma-percentile life is total operation time when object under study does not break down with g probability expressed in percentage. The gamma-percentile fault operation time tg is determined by the equation:

$$\int_0^{t_g} f(t,s_1,s_2)dt = 1 - g$$

The algorithm presented on Figure 1 under fault of the following equipment for residual life determination should be repeated.

TABLE I
EXAMPLE OF EQUIPMENT STATISTICAL DATA

Object name	Fault cause	Object health	Starting operation date	Fault date	Operation time (in days)
Name 1	Cause 1	Health 1	Starting operation date 1	Fault date 1	Amount 1
Name 2	Cause 2	Health 2	Starting operation date 2	Fault date 2	Amount 2
…	…	…	…	…	…
Name N	Cause N	Health N	Starting operation date N	Fault date N	Amount N

TABLE II
EXAMPLE OF ONE-DIMENSIONAL FAULT VECTOR AND ADDITIONAL VECTOR OF CENSORED OPERATION TIMES

Object name	Name 1	Name 2	Name i	Name i+1	…	Name N-1	Name N
Fault operation times	No1	No2	Noi		…		
Censored operation times				Nci+1	…	Ncn-1	Ncn

TABLE III
EXAMPLE OF INTEGRATED FAULT INTENSITY

Objet name	Name 1	Name 2	Name i	Name i+1	…	Name N-1	Name N
Fault time operations	No1	No2	Noi		…		
Censored time operations				Nci+1	…	Ncn-1	Ncn
Quintile number (q)	1	2	i	i+1	...	N-1	N
Integrated fault intensity (Pi)	P1	P2	Pi	Pi+1	...	Pn-1	Pn

TABLE IV
DISTRIBUTION FUNCTION SHAPE AND DISTRIBUTION PARAMETERS

Distribution	Equations for parameter search	Distribution parameters
Normal	$\frac{1}{\sigma\sqrt{2\pi}}\int_0^{q_i} e^{-\left(\frac{(t-\mu)^2}{2\sigma^2}\right)} dt = p_i$	μ, σ
Lognormal	$\frac{1}{\sigma\sqrt{2\pi}}\int_0^{q_i} e^{-\left(\frac{(\ln t-\mu)^2}{2\sigma^2}\right)} dt = p_i$	μ, σ
Weibull	$\int_0^{q_i} \left(\frac{k}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{(k-1)} e^{-\left(\frac{t}{\lambda}\right)} dt = p_i$	k, λ

where μ, σ - expectation value and mean squared departure correspondingly; t – control parameter; k, λ – shape factor and scale factor correspondingly

TABLE V
STATISTICAL DATA OF OBJECTS UNDER INVESTIGATION

Object name	Fault cause	Object health	Starting operation date	Fault date	Operation time (in days)
VN#1	Stator insulation breakdown	Fault	14.06.2010	16.07.2012	750
VN#2	Bearing runout	Run	12.02.2013	25.12.2013	316
VN#3	Bearing runout	Run	26.03.2013	05.02.2014	316
VN#4	Shaft breakdown	Run	05.05.2013	18.02.2014	289

IV. METHOD IMPLEMENTATION

Four same-type electric motors in pulp and paper production are used as test objects. Collected data for research are showed in Table 5.

Based on collected data on the VN#1 operation it should be noted that this electric motor is operated at different time interval. The VN#1 at the moment of other electric motors operation was not ued. For this reason, VN#1 will be used in calculation circumstantially. Consequently, the operation time of VN#1 is 750 days and it will not change in following calculations. VN#2 and VN#3 and VN#4 electric motors were used under similar conditions. Therefore, all calculations for the last three electric motors will be described in the paper.

Operation times of electric motors are calculated by collected data and one-dimensional fault time vectors when electric motor breaks down and censored vectors are generated. The following vectors under second, third and fourth faults are showed in Table 6.

TABLE VI
ONE-DIMENSIONAL FAULT VECTOR UNDER VN#2, VN#3, VN#4 FAULTS

	Second object fault				**Third object fault**				**Fourth object fault**			
Object name	VN#1	VN#2	VN#3	VN#4	VN#1	VN#2	VN#3	VN#4	VN#1	VN#2	VN#3	VN#4
Fault operation times	750	316			750	316	316		750	316	316	289
Censored operation times			269	230				271				

The fault operation time vector includes operation times of fault electric motors. The censored operation time vector includes operation times of running electric motors. The operation time of VN#2 is 316 days at the moment of its fault. Operation times of VN#3 and VN#4 are 269 and 230 days consequently at the censored vector because they work at the moment of VN#2 fault. Under the following fault vectors are generated in the same manner.

The integrated fault intensities are calculated based on fault time vectors and censored vectors given. The following fault intensities of VN#2, VN#3, VN#4 per quintiles are obtained correspondly (Table 7).

TABLE VII
INTEGRATED FAULT INTENSITY UNDER VN#2, VN#3, VN#4 FAULTS

	Second object fault				**Third object fault**				**Fourth object fault**			
Object name	VN#1	VN#2	VN#3	VN#4	VN#1	VN#2	VN#3	VN#4	VN#1	VN#2	VN#3	VN#4
Fault operation times			316	750		316	316	750	289	316	316	750
Censored operation times	230	269			271							
Quintile number (q)	1	2	3	4	1	2	3	4	1	2	3	4
Integrated fault intensity (Pi)	0	0	0,33	1	0	0,25	0,5	1	0,2	0,4	0,6	1

It should be noted that an same-type electric motor with 20% probability after 289 days and with 60% probability after 316 days based on calculation of integrated fault intensity breaks down.

In this case parameters search for data under investigation is realized simultaneously by few quintiles corresponding to current operation times of certain electric motors. A fault is fixed only for one electric motor. For this example, the highest informativeness is reached by lognormal distribution.

The curve shape of lognormal distribution under VN#4 fault as a result of calculation is given on Figure 2.

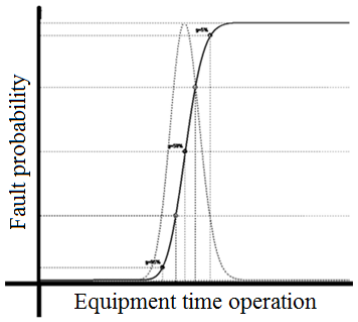


Fig 3. Lognormal distribution curve. Solid line – distribution density f(t); doted line – integral function of probability distribution F(t).

Unprecise parameter values of distribution function under the first electric motor fault are determined. These parameters under the following faults are clarified. The following parameter values after VN#4 fault as the result of equation system solution by least square method are obtained:

$$\mu=8,89$$
$$\sigma=0,07$$

These values under calculation of gamma-percentile operation time at the moment of current equipment fault are used. For determination of operation time g is 95% because this values is more often used for determination of objects reliability. In this way, when step, maximum value, accuracy are known and the following equation is calculated by numerical method:

$$\int_0^{t_g} f(t,\mu,\sigma)dt = 1-0{,}95$$

then the gamma-percentile operation time tg4 after the fourth fault is 271 days. It should be noted that each following electric motor must be checked not later than 271 days after gamma-percentile indexes calculation. The gamma-percentile operation time after each new spontaneous equipment fault will be corrected. That will increase probability calculation accuracy of the same electric motor fault.

V. Conclusion

The algorithm of equipment residual life determination has been developed. This algorithm allows to calculate fault probability of each equipment unit before its fault. Statistical data of electric motors operations in pulp and paper production are considered. Dates of new electric motors Fault have been computed based on statistical data.

References

[1] Benbow, Donald and Hugh Broome, “The Certified Reliability Engineer Handbook”, 2 ed., ASQ Quality Press. 2013.

[2] Universal model for reliability prediction of electronics components, PCBs and equipment, IEC Standard 62380, 2004.

[3] Breyfogle, Forrest, “Statistical Methods for Testing, Development, and Manufacturing”, New York: Wiley, 1992.

[4] Neufelder, Ann Marie, “Ensuring Software Reliability”, New York: Dekker, 1993.

[5] Abernethy, Robert B., “The New Weibull Handbook”, 5th ed., Houston, TX: Dr Robert Abernathy, 2006.

Optimum Design of Electromagnets for Magnetic Levitation of Transport Systems based on the Inverse Problem Solutions

Yuriy Bakhvalov, Valeriy Grechikhin, and Anna Yufanova
Platov South-Russia State Polytechnic University (NPI)
Prosvesheniya Str. 132, 346428 Novocherkassk, Russia
E-mail: imo.npi.tu@gmail.com, vgrech@mail.ru, ufanova@otlook.com

***Abstract*—Article is devoted to design of optimum electromagnets for magnetic levitation of transport systems. The method of electromagnets design based on the inverse problem solution of electrical equipment is offered. The method differs from known by introducing a stage of minimization the target functions providing the stated levitation force and magnetic induction in a gap, and also the mass of an electromagnet. Initial values of parameters are received, using approximate formulas of the theory of electric devices and electrical equipment. The example of realization of a method is given. The received results show its high efficiency at design. It is practical to use the offered method and the computer program realizing it as a part of system of the automated design of electric equipment for transport with a magnetic levitation.**

***Keywords:* optimization, inverse problems, finite elements method, electromagnet, magnetic field, design automatization.**

I. Introduction

Now more and more strict requirements of ecological purity and profitability are imposed to modern transport. In this regard searches of the new principles of transport systems creation are conducted. One of the perspective directions of their development considers applying of a magnetic levitation of trains that allows, eliminating mechanical contact with road structure, to increase movement speed up to 500 km/h [1]–[3].

Intensive works on creation of transport with a magnetic levitation have been conducted from 70th of the XX century in Germany, Japan, Great Britain and in other countries [4–7]. More than 25 modifications of experimentation`s trains are created. In China and Japan commercial transport systems with a magnetic levitation are successfully operated [8,9]. The Shanghai express build up speed of 430 km/h, a gap between magnetic levitation and iron rail is 15 mm.

It should be noted that magnetic levitation receives the increasing use in mechanical engineering (magnetic bearings, positioning systems, etc.) [10][11].

Process of design of levitation transport systems divided into a number of stages. At the first stage the electric power necessary for power supply of electromagnets and the linear traction engine is estimated. At the second stage the system of power supply for transfer on train of necessary power is selected (values of voltage, current-collectors, etc.). At the third stage design of traction system and magnetic levitation is carried out. Thus, having information on tension size, AC to DC converters for supply of electromagnets are designed, and then all system of electromagnetic levitation in a complex with the converter, a control system and an electromagnet, and also traction system is optimized.

The materials reported in this article are intended to be used at the first design stage.

In last years the inverse problems solution in design of technical objects [12] and creation on their basis systems of the automatization design of the electrotechnical equipment were widely used. In this regard an actual task is creation of methods of the automatization design of the electromagnets applied in levitation transport systems with the use of such approach.

II. Design of Electromagnet

Designs of the electromagnets fixed on the special console connected with train are presented in fig. 1 [4]. In figure: 1 – iron rail of levitation system; 2 – bearing; 3 – levitation electromagnet core; 4 – coil; 5 – coil; 6 – core of an electromagnet of side stabilization; 7 – iron rail of side stabilization system; 8 – console.

We will consider design of a levitation electromagnet. We consider known the mass of train **M**, quantity of levitation electromagnets **n**, their length **L**. It is required to determine the sizes of an electromagnet section and a magnetic potential difference (MPD) of the coil (fig. 2) so that the demanded levitation force, the minimum mass of an electromagnet and reliable control system were provided.

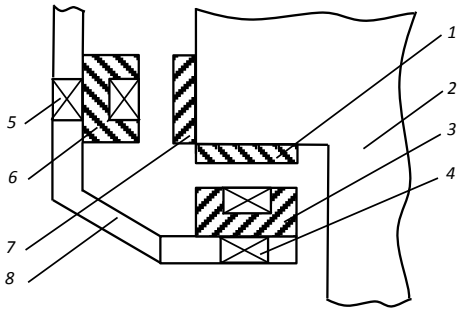


Fig. 1. Designs of electromagnets and their placement on the console, onnected with train.

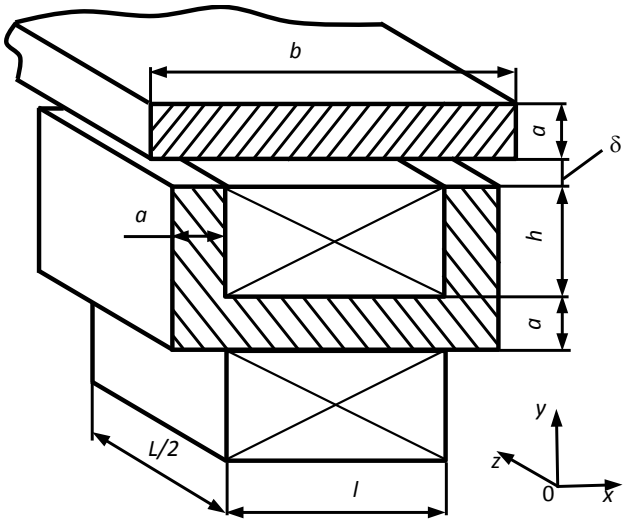


Fig. 2. Central section of an electromagnet. L – electromagnet length on an axis 0z.

III. MATHEMATICAL MODEL

We will determine levitation force on one electromagnet by a formula

$$F_l = Mg/n \tag{1}$$

where g – acceleration of gravity.

Using an approximate formula from the theory of electric devices for electromagnet attraction force [13]

$$F_l = 2\frac{B_\delta^2 S_p k_b}{2\mu_0} \tag{2}$$

where B_δ – average magnetic induction in an air gap of δ; k_b – the coefficient considering a buckling of a magnetic flux; $k_b = 1.2 \div 1.5$; μ_0 – a magnetic constant, $\mu_0 = 4\pi \cdot 10^{-7}$ H/m, we will receive initial approach for the pole area

$$S_p^{(0)} = La^{(0)} = \frac{\mu_0 F_l}{B_\delta^2 k_b} \tag{3}$$

Taken set values for B_δ, L, k_b we obtain on the basis of (3) formula for initial value of the size «*a*»

$$a^{(0)} = \frac{S_{\text{п}}^{(0)}}{L} = \frac{\mu_0 F_l}{B_\delta^2 k_{\text{в}} L} \tag{4}$$

Let permeabilty of ferromagnetics be infinite. Using a known ratio [14]

$$\oint_l \vec{H} d\vec{l} = iw$$

We will receive a formula for definition of initial approach of a magnetic potential difference (MPD) $F_{\text{м}} = iw$ of the coil

$$iw^{(0)} = \frac{B_\delta}{\mu_0} 2\delta \tag{5}$$

We will determine the area of a port of the coil by a formula

$$S_{\text{c}}^{(0)} = l^{(0)} h^{(0)} = \frac{iw^{(0)}}{k_z j}, \tag{6}$$

where j – the chosen current density in the coil, k_z – coefficient of filling of a port of the coil on copper, $k_z = 0.7$ is accepted.

From (6) follows

$$l^{(0)} = \frac{S_c^{(0)}}{h^{(0)}} = \frac{iw^{(0)}}{k_z j h^{(0)}} \tag{7}$$

At calculation of magnetic fields we consider a curve of magnetization of steel is unique, isotropic mediums.

We will determine the mass of an electromagnet by a formula

$$M_{em} = \left[2ah + \left(\frac{S_c}{h} + 2a \right) a \right] L\rho_{st} + 2S_c k_z L\rho_{cu} \tag{8}$$

where ρ_{st}, ρ_{cu} – density of a ferromagnetic (steel) and copper respectively.

In a formula (8) a variable value is only h.

Necessary and sufficient conditions of a minimum of M_{em} as shown here:

$$\frac{dM_{em}}{dh} = 0; \ \frac{d^2 M_{em}}{dh^2} > 0 \tag{9}$$

On the basis of (9) we will receive

$$\frac{dM_{em}}{dh} = \left[2a - \frac{S_c a}{h^2} \right] L\rho_{st} = 0; \ \frac{d^2 M_{em}}{dh^2} = \frac{S_c a}{h^3} L\rho_{st} > 0 \tag{10}$$

The value h at which M_{em} is minimum, is defined from (10)

$$h^{(0)} = \sqrt{\frac{S_c^{(0)}}{2}} = \sqrt{\frac{iw^{(0)}}{2k_z j}} \quad (11)$$

We will determine width of an iron rail by a formula

$$b^{(0)} = l^{(0)} + 2a^{(0)} + 0.02 \quad (12)$$

IV. Solution Algorithm

At minimization of mass of M_{em} with use of expression (8) it is necessary to consider two restrictions providing operability of system of magnetic levitation

$$B_\delta^{(k)} \le B_\delta^* \quad (13)$$

$$F_l \le F_l^{(k)} \le F_l\left(1 + \varepsilon_1\left(F_l\right)\right) \quad (14)$$

where $B_\delta^{(k)}$ and $F_l^{(k)}$ – the values calculated on k`s algorithm step.

Usually $B_\delta^* = 0.4 - 0.7$ T, that ensures functioning of system in the nonsaturated mode.

We will transform restrictions to following target functions

$$J_1^{(k)}(iw) = \left(B_\delta^{(k)} - B_\delta^*\right)^2 \quad (15)$$

$$J_2^{(k)}(a) = \left[\left(1 + \varepsilon_1\left(F_l\right)\right)F_l - F_l^{(k)}\right]^2 \quad (16)$$

We have a multicriteria task with three target functions (15), (16) and (8). We will execute the solution of a task by a lexicographic method of ordering [15], taking into account features of a task.

The iterative algorithm of an electromagnet design consists of the following stages:

1. We determine levitation force of one electromagnet Fl by a formula (1). We assign values to sizes of a gap of δ, current density in coils *j*, airgap flux density B_δ^*, electromagnet length *L*, coefficients k_b and k_z, errors of magnetic induction definition $\varepsilon_1\left(B_\delta^*\right)$ and force $\varepsilon_2(F_l)$.

We determine by formulas (4), (5), (7), (11), (12) initial values $a^{(0)}$, $iw^{(0)}$, $h^{(0)}$, $l^{(0)}$, $b^{(0)}$.

Further for each k=0,1,2,…

2. We solve a direct problem of a stationary magnetic field calculation and force F_l by finite elements method, using

$$a = a^{(k)},\ iw = iw^{(k)},\ h = h^{(k)},\ l = l^{(k)},\ b = b^{(k)}.$$

3. We minimize functionalities $J_1^{(k)}$ and $J_2^{(k)}$ by method of gradient descent, checking performance of conditions

$$J_1^{(k)}(iw) \le \left[\varepsilon_1\left(B_\delta^*\right)\right]^2 \quad (17)$$

$$J_2^{(k)}(a) = \left[\varepsilon_2\left(F_l\right)\right]^2 \quad (18)$$

As a result we define values iw^{k+1} and a^{k+1} at which conditions (17) and (18) are carried out.

4. We minimize the mass of an electromagnet.

We determine parameters *h* and *l* by formulas

$$h^{(k+1)} = \sqrt{\frac{iw^{(k+1)}}{2k_3 j}} \qquad l^{(k+1)} = \frac{iw^{(k+1)}}{k_3 j h^{(k+1)}}$$

Further we calculate

$$b^{(k+1)} = l^{(k+1)} + 2a^{(k+1)} + 0.02$$

V. Computing Experiment Results

We will review an example. It is required to determine the sizes and MPD of electromagnets of train levitation system with the mass $M = 60$ t, including $n = 20$ electromagnets, located in two ranks; length of each $L = 1$ m. Train length – 27 m. We choose: $j = 3 \cdot 10^6$ A/m^2; $B_\delta = 0.7$ T; $k_b = 1.5$; $k_z = 0.7$; $\delta = 0.015$ m; $\varepsilon_1\left(B_\delta^*\right) = 0.01$; $\varepsilon\left(F_l\right) = 0.01$.

We carry out the first stage of algorithm: $F_l = 30$ kN; $a^{(0)} = 0.034$ m; $iw^{(0)} = 16711$ A; $h^{(0)} = 0.063$ m; $l^{(0)} = 0.126$ m; $b^{(0)} = 0.215$ m.

Go over to the stages 2-4 execution.

As a result we will receive parameters of the electromagnet of the minimum weight on the fourteenth iteration providing magnetic induction in a gap of an electromagnet $B_\delta = 0.7$ T and levitation force $F_l = 30$ kN with the set error: $a^{(14)} = 0.054$ m; $iw^{(14)} = 14180$ A; $h^{(14)} = 0.058$ m; $l^{(14)} = 0.116$ m; $b^{(14)} = 0.245$ m.

The sizes specified above and MPD of an electromagnet are received at motionless train. At the movement vortex currents are induced.in a ferromagnetic rail. Distribution of eddy-currents in a rail according to the law of electromagnetic inertia is so that their field is directed opposite to an electromagnet field that leads to reduction of uplift power.

The method of field and forces calculation of the electromagnet moving along a ferromagnetic plate was considered in work [16]. The famous software ANSYS Maxwell allows making such calculations. By means of this software it is determined that for creation of the electromagnet providing levitation force F_l at the movement of train over iron rail with a speed $V = 400$ km/h shall be

$$F_{lcalc} = k_{ec} F_l$$

where $k_{ec} = 1.35$ at specific electrical conduction of rail $\gamma = 2 \cdot 10^6$ $(\Omega \cdot \text{m})^{-1}$ and a gap of $\delta = 0.015$ m. For the considered electromagnet $F_{lcalc} = 40.5 \cdot 10^3$ N.

Thus, uplift force decreases by 35 % in the considered device at a speed $V = 400$ km/h because of induced eddy-currents.

Value k_{ec} depends on speed, for example, at $V = 200$ km/h $k_{ec} = 1.21$, at $V = 800$ km/h – $k_{ec} = 1.5$.

Using the algorithm described above, we will determine the sizes and MPD of an electromagnet for $F_{lcalc} = 40{,}5 \cdot 10^3$ N,that will provide the required force of a levitation $F_l = 30 \cdot 10^3$ N.

On the thirteenth iteration parameters of an electromagnet of the minimum mass $M_{em} = 344$ kg are determined providing magnetic induction in a gap of an electromagnet $B_\delta = 0.7$ T (fig. 3) and levitation force $F_л = 40.5$ kN with the set error: $a^{(13)} = 0.091$ m; $iw^{(13)} = 16990$ A; $h^{(13)} = 0.064$ m; $l^{(13)} = 0.127$ m; $b^{(13)} = 0{,}328$ m.

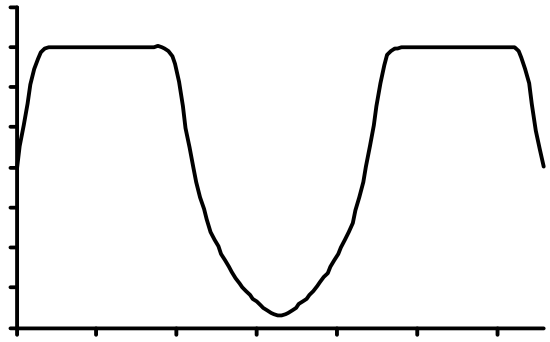


Fig. 3. Distribution of the module of magnetic induction in an electromagnet gap.

CONCLUSIONS

The method of optimum design of electromagnets of magnetic levitation on the basis of the solution of conditionally correct inverse problems with use of two target functions received on the basis of restriction for magnetic induction in an electromagnet gap on tongue attraction force is offered. Minimization is carried out by method of gradient descent. At the same time magnetic potential difference of the magnetizing coil and the geometrical sizes of a pole of an electromagnet are defined. Minimization of the third target function (mass of an electromagnet) carried out analytically with use of necessary and sufficient conditions allowed to receive formulas for calculation of its optimum sizes. Iterative algorithms of the inverse problems solution of magnetic fields are constructed. The offered method allows reducing considerably time of design of electromagnets in comparison with the known methods, for example, by penalty function method. It is practical to use the offered method and the realizing computer program as a part of the computer-aided engineering system of levitation transport systems.

The results of the research are obtained under support of the project1.2690.2014/K "Methods of inverse problems solutions for diagnostics of complex systems (in engineering and medicine) on the basis of full-scale and model experiment" which is component of state project task.

REFERENCES

[1] M. Murai and M. Tanaka, “Magnetic levitation (maglev) technologies,” *Japan Railway & Transport Review*, vol. 25, pp. 61–67, 2000.

[2] R. Schach, P. Jehle, and R. Naumann, *Transrapid und Rad-Schiene-Hochgeschwindigkeitsbahn*. Berlin Heidelberg: Springer-Verlag, 2006, 428 p.

[3] H. W Lee, K.C. Kim, and J. Lee, “Review of maglev train technologies,” *IEEE Trans. Magn.*, vol. 42, no. 7, pp. 1917–1925, 2006.

[4] V. I. Bocharov, Yu.A. Bahvalov, and I.I. Talya, *Bases of design of an electrorolling train with a magnetic levitation and the linear traction electric drive*. P.1. Rostov N / D.: Publishing house of RND university, 1992, p. 432.

[5] Y. H. Tzeng and T.C. Wang, “Optimal design of the electromagnetic levitation with permanent and electro magnets,” *IEEE Trans. Magn*., vol. 30, no. 6, pp. 4731–4733, 1994.

[6] S. Guangwei, C. Weigong, and R. Meisinger, “The Research on the Model of a Magnetic Levitation System,” *Electric Machine and Control*, vol. 9, no. 2, pp. 187–190, Feb. 2005.

[7] P. K. Biswas and S. Bannerjee, “Analysis of U-I and U-U type rail and actuator used in electromagnetic levitation system using FEM software,” *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 5, pp. 32-39, 2012.

[8] Y. Yasuda, M. Fujino, and M. Tanaka, “The first HSST maglev commercial train in Japan,” in: *Maglev 2004 Proceedings*, Shanghai, China, 2004, pp. 76-85.

[9] X. M. Wu, *Maglev Train*. Shanghai: Shanghai Science and Technology Press, 2003: pp. 1-10.

[10] G. Schweitzer and E.H. Maslen, *Magnetic Bearing Theory: Design, and Application to Rotating Machinery*. Berlin: Springer, 2009, 535 p.

[11] W. J Kim and H. Maheshwari, “High-precision control of a maglev linear actuator with nanopositioning capability,” in: *Proc. 2002 American Control Conf.* , May 2002, pp. 4279–4284.

[12] Yu.A. Bahvalov, N.N. Gorbatenko, and V.V. Grechikhin, *Inverse problems of electrical equipment*. Novocherkassk: Publishing house of the magazine "News of higher education institutions. Electromecanics", 2014, p. 211 .

[13] A. T. Morgan *General Theory of Electrical Machines*. London: Heyden & Sons, 1979, 448 p.

[14] K. Simonyi, *Theoretische Elektrotechnik.* Leipzig: Barth Verlag, 1993, p. 974.

[15] J. Nocedal and S.J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999, 636 p.

[16] O.V. Tozoni, “The calculation of the field of the electromagnet moving along the electromagnetic band,” *Proceedings of the Russian Academy of Sciences. Power Engineering and Transport*, vol. 6, pp. 96–109, 1977.

Optimization of the Modular Educational Program Structure

Alexandr Ivanchenko, Anastasia Kolomiets, Dmitriy Grinchenkov, Nguyen Van Ngon
Platov South-Russian State Polytechnic University (Novocherkassk Polytechnic Institute)
Information Technologies and Control Department
Prosvescheniya Str. 132, Rostov region, 346428, Novocherkassk, Russia
E-mail: grindv@yandex.ru

***Abstract*—Optimized structure of the educational program consisting of a set of the interconnected educational objects is offered by means of problem solution of optimum partition of the acyclic weighed graph. The condition of acyclicity preservation for subgraphs is formulated and the quantitative assessment of decision options is executed. The original algorithm of search of quasioptimum partition using the genetic algorithm scheme with coding chromosomes by permutation is offered. Object-oriented realization of algorithm in language C ++ is described and results of numerical experiments are presented.**

Keywords: **e-learning, educational program, educational object, computer testing, acyclic weighed graph, number partition, graph partition, random permutation, genetic algorithm, object-oriented approach.**

I. INTRODUCTION

While study processes planning of large volumes of a training material with use of electronic training technologies (for example, for preparation and professional development of the industrial enterprises personnel) all training material of the educational program (EP) is presented usually in the separate portions – educational objects (EO), on each studying completion intermediate certification is carried out (usually in the form of computer testing – CT).

EO are connected by the relation of precedence-consequence, and successful certification for one EO is an admission condition to studying of another EO. Natural model of such EP is the directed acyclic weighted graph $G=(V,E,W_E)$, where V – a set of n vertex presenting EO, E – a set of edges (precedence- consequence conditions) and – W_E a set of edges weight [1][2][3]. The weight of w_{ij} defines "activation threshold" of an edge: if the assessment on EO v_i is more or equal to w_{ij}, studying of EO v_j is possible, otherwise repeated testing on v_i is necessary.

In practice the quantity of EO as a part of EP can be rather big, especially in cases of the corporate educational structures connected with training of employees of the enterprises on various specialties. Under these conditions problem of integration of EO in the educational modules (EM) by graph partition G on a small amount of k subgraphs is becoming actual:

$$V = \bigcup_{p=1}^{k} V_p,\ \bigcap_{p=1}^{k} V_p = \emptyset. \quad (1)$$

Rather obvious optimality criteria of such partition are minimization of "interconnections" between EM while preservation of acyclicity for subgraphs.

Before formalization of optimality criteria and condition of acyclicity we will make necessary explanations. We will call the *scheme of graph partition* $s_i(n,k)$ any number partition n into k parts:

$$s_i(n,k) : n = a_1 + a_2 + \cdots a_k. \quad (2)$$

It is known [4, p.71], that total of such partition P is defined recursively:

$$P(n,k) = P(n-1,k-1) + P(n-k,k), \quad (3)$$

where: $P(i,i) = 1, P(i,1) = 1, \forall i,\ P(i,j) = 0, j > i$.

Generation algorithms of all number partition n into k parts are given in [4, p 70]. Thus, when graph partition from n vertex on k subgraphs there is P (n, k) schemes of partition (for example, P (20,3) = 33).

We will define, how many various options of graph partition for the scheme set (assign $Q(s_i(n,k)$ to that quantity). As there generally can be repeating elements in the scheme of partition, the general view of the scheme will be:

$$s_i(n,k) : n = q_1 a_1 + q_2 a_2 + \cdots q_m a_m, \quad (4)$$

where q_j>=1, m<=k, all a_j varies.

The formula is fair:

$$Q\big(s_i(n,k)\big) = \frac{n!}{q_1!(a_1!)^{q_1} q_2!(a_2!)^{q_2} \ldots q_m!(a_m!)^{qm}} \quad (5)$$

For example, $s_i(4,2)$: 4 = 2+2. Then:

$$Q\big(s_i(4,2)\big) = \frac{4!}{2!\,(2!)^2} = \frac{2\cdot 3\cdot 4}{2\cdot 4} = 3.$$

It will be such partitions: {(1,2), (3,4)}, {(1,3), (2,4)}, {(1,4), (2,3)}.

One more example: P(8,3)=5. It will be scheme:

$s_1(8,3)$=1+1+6, $s_2(8,3)$=1+2+5, $s_3(8,3)$=1+3+4, $s_4(8,3)$=2+2+4, $s_5(8,3)$=2+3+3

Total number of graph partition from 8 vertex on 3 subgraphs:

$$N(8,3) = \sum_{i=1}^{P(8,3)} Q\big(s_i(8,3)\big) = \frac{8!}{2!6!} + \frac{8!}{2!5!} + \frac{8!}{3!4!} + \frac{8!}{2!(2!)^2 4!} + \frac{8!}{2!2!(3!)^2} = 966.$$

The formula for total number of graph partition from n vertex on k subgraphs is fair:

$$N(n,k) = \sum_{i=1}^{P(n,k)} Q\big(s_i(n,k)\big). \qquad (6)$$

Let us denote $r_l(n,k) = \{V_1^l, V_2^l, \dots V_k^l\}$ for current graph partition. Then we will consider the best for the set n and k - partition minimizing the following criterion (total weight of interconnections):

$$F(n,k) = \min_{r_l(n,k),\, l\in[1,N(n,k)]} \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}, \quad v_i \in V_x^l, \ v_j \in V_x^l, x \neq y. \qquad (7)$$

It is required to keep an acyclicity condition for subgraphs by its partition. We will consider that partition $r_l(n,k)$ keeps property of acyclicity if for any couple of subgraphs V_x^l, V_y^l all edges connecting them go into one direction.

We will notice that formulated problem of graph optimum partition differs from known graph partition problems [6,7] by the condition that initial graph is directed acyclic and it is required to keep an acyclicity condition for subgraph.

II. Algorithm for Solution

Algorithm 1. Search of optimum graph partition from n vertex on k subgraphs by straightforward enumeration.

[It is supposed that procedure of generation of partition schemes is available (described in literature) and generation procedure of all graph partition for the partition scheme set is available (not presented in literature)]

Step 1. Regular scheme of partition is generated.

Step 2. The next partition for the current scheme is generated.

Step 3. If partition is accepted (provides acyclicity for subgraphs), we pass to a Step 4, else to Step 5.

Step 4. The total weight of interconnections for the current partition is calculated and correct the minimum value of criterion.

Step 5. If not all partitions for the current scheme are received, transition to a Step 2, else to a Step 6.

Step 6. If not all schemes of partition are received, transition to a Step 1, else Stop.

The main shortcomings of this algorithm are: need of development of generation procedure for all graph partitions for the partition scheme set and the exponential growth of algorithm operating time at increase in number of graph vertexes.

It is possible to offer simple algorithm of quasioptimum partition search **(Algorithm 2)**, free from the first shortcoming which is overcome by generation of random permutations of a set of graph vertexes and "appoint" it - the current partition. Of course, there is a problem of repeatability of partition appears, as it is obvious that two various permutations can correspond to the same partition for the scheme set.

For example, for the scheme of 4=2+2 permutation (1,3,2,4) and (4,2,3,1) will be identical because they set the same graph partition into subgraphs with vertexes (1,3) and (2,4). However repeatability of partition can be referred to the property of a quasioptimality of this algorithm. It is obvious that the end condition of the offered algorithm of random permutations (step 5) is performance of the set number of iterations (for example, 10000).

Offered algorithm of random permutations is simplest and it is possible to achieve much better effect from use of random factor in case of application the standard scheme of the genetic algorithm (GA) [7]-[10] which can be "built in" algorithm 1 instead of steps 2-5 (we will call it **Algorithm 3**).

Obvious way of decisions coding (chromosomes on terminology of GA) in the considered task are permutation, as in this case the phenotype and a genotype coincide and coding/decoding operations aren't necessary, and the total weight of interconnections for the current partition will be used as fitness function.

It should be noted that at generation of the next decision (chromosome in the form of permutation) when forming initial population or as a result of performance of mutation/crossing operations it is necessary to carry out two checks:

1) whether this decision is admissible from the point of view of ensuring acyclicity for subgraphs;

2) whether this decision (chromosome) coincides with one of already available in population, or with a chromosome before a mutation, or from one of parental chromosomes.

III. Practical Realisation of Algorithm

For formalization we will use the formal-language notation on the basis of language C++ that allows to apply exclusively effective in this case object-oriented approach [11-13]. The object model of the considered task is based on several interconnected classes.

Class **Graph** encapsulates a set of *vertex (set<int> vertex)* and weighted edges (*vector<Edge> edges*, where *Edge = pair<pair<int,int>, int>*) and contain in open part necessary functionality:

- function *void push_back(Edge),* including edge in graph,
- function *size_t size(),*returning quantity of vertex of the graph,

- function *bool isEdge(pair<int,int>)*, checking whether there couple of numbers that corresponds to any edge of the graph;
- function *bool one_direction(vector<int>, vector<int>)*, checking whether all edges connecting vertexes of two subgraphs have one direction;
- function *double bind_size(vector<int>, vector<int>),* returning the total power of interconnections of two subgraphs.

Class **Partition** encapsulates partition – array of subgraphs *(vector<vector<int>> P)*. Besides, static fields are stored in open part of this class: the scheme of graph partition (*static vector<int> Scheme*) and the index on graph (*static Graph* G*) that establishes between the classes Partition and Graph the dependence relation. Functionality of the class Partition is provided with the following functions:

- two constructors: by default *Partition()* and constructor of type transformation *Partition(vector<int>),* allowing to initialize a class sample by permutation;
- overloaded operation of assignment *void operator= (vector <int>*), interfaced to the constructor of type transformation;
- overloaded operation of checking of two partitions into identity *bool operator==(const Partition&);*
- function *bool isAcyclic*(), checking partition into acyclicity;
- function *int bind_size(),*returning total power of interconnections for all subgraphs.

For organization of calculation with usage of classes Graph и Partition it is necessary:

- to declare static fields in global area of visibility *vector<int> Partition::Scheme* and *Graph* Partition::G*;
- to create graph in the body of main program *Graph Gr*, fill it with data (edges) through of *Gr.push_back(...)* and to pass a pointer to *Gr* in class Partition: *Partition::G = &Gr*.

In the subsequent it will be necessary to transfer to the class Partition the current scheme of partition by value assignment to the static field *Partition::Scheme*, for example: *Partition::Scheme = {4,5,6}*.

The object model of genetic algorithm includes the following classes:

Class **Chromosome** encapsulates a chromosome as an array of genes *(vector <int> genes)* and an assessment of "fitness" of a chromosome (*double fitness*), as well as the static index on fitness function which will be set in the main program (*static F fit_fun* where *F = double (*) (vector <int> &)*). Functionality of the class Chromosome is provided with the following functions:

- two constructors: by default *Chromosome()* and constructor of type conversion *Chromosome (vector<int>)*, allowing to initialize chromosome by permutation and on the fly and to assess its "fitness" with the help of fitness-function *fit_fun*;
- two functions for “retrieval” values of fields: *vector<int> get_genes()* and *double get_fitness()*;
- overloaded external functions of comparison of two chromosome (==, < and >).

Class **Population** encapsulates population as an array of (*vector<Chromosome> pop*). Functionality of the class Population is provided with the following functions:

- constructors by default *Population()*;
- function of chromosome inclusion in population *void insert(const Chromosome& c)*, which automatically maintains orderliness of chromosomes in population on increase of value of fitness;
- overloaded operation of access to population elements on index *Chromosome operator[](int i);*
- function of obtaining size of population *int size();*
- function of the table frequencies (roulette wheel) creation for population elements *vector<double> get_freq();*
- function of the next population creation *void next_pop()*.

The most difficult is the function *next_pop()* which realizes the GA scheme, consistently by following steps:

- select "elite" (the most adapted chromosomes) and includes it in the next population (next generation);
- the rest of the next generation is filled with chromosomes descendants which are formed as a result of application crossing operation (crossover) to two "parental" chromosomes selected from the current population by means of "roulette wheel";
- mutation operation is applied to some chromosomes from the created new generation.

Developed and realized set of classes has allowed to make record of the main algorithm of the problem solution exclusively compact: it is only necessary to create initial population by means of the generator of casual permutation and then the set number of times (for example, 10000) to call the *next_pop* function ().

Fig.1 shows graph from 20 vertexes and its optimum partition into subgraphs A, B and C according to the scheme {5,6,9}. The size of population has been set 50, and the number of iterations (generations) – 500. We will notice that there are 77 597 520 options of partition for that graph according to the scheme {5,6,9}.

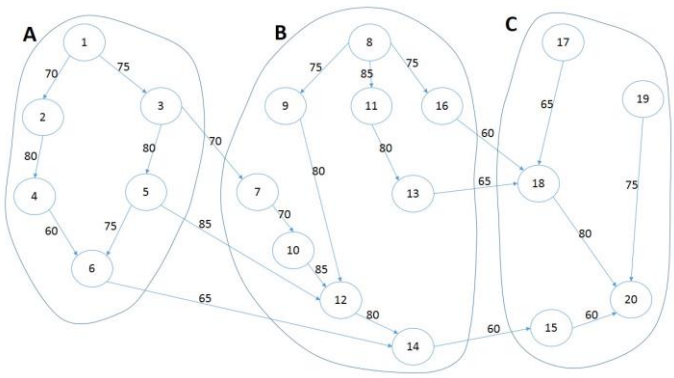


Fig. 1. Initial graph and its optimal partition by scheme {5,6,9}

The integrated representation of initial EP in the form of three modules is shown on the Fig.2. Similar structurization allows optimizing process of training by allocation of a small amount of logically complete educational modules and orienting student to the ordered passing of all set of educational objects [15].

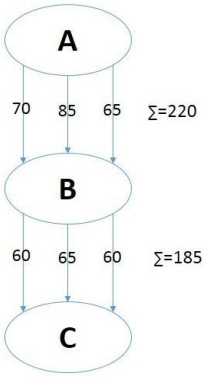


Fig. 2. Graph representation in the form of three subgraphs (modules).

IV. CONCLUSIONS

The numerical experiments executed by authors on model examples have shown high efficiency of the offered algorithm of creation of quasioptimum graph partition with use of genetic algorithm and coding of decisions by permutations at rather large number of vertexes. Optimized structure of the educational program consisting of a set of the interconnected educational objects is offered by means of problem solution of optimum partition of the acyclic weighed graph. The condition of acyclicity preservation for subgraphs is formulated and the quantitative assessment of decision options is executed. The original algorithm of search of quasioptimum partition using the genetic algorithm scheme with coding chromosomes by permutation is developed.

REFERENCES

[1] Thanassis Hadzilacos, Dimitris Kalles, Dionysis Karaiskakis, Maria Pouliopoulou. Using Graphs in Developing Educational Material // Proceedings of the 2nd International Workshop on Building Technology Enhanced Learning Solutions for Communities of Practice. TEL-CoPs'07. Sissi, Lassithi Crete. Greece, 18 September, 2007. URL: http://ceur-ws.org/Vol-308/paper04.pdf

[2] Ivanchenko A. N., Nguyen Van Ngon. Simulation modeling of the process study of the modular curriculum // *University News. North-Caucasian Region. Technical Sciences Seri*es. 2015; 3: 28-33

[3] Ivanchenko A. N., Nguyen Van Ngon. Modelling of e-learning using temporal and random graphs // *University News. North-Caucasian Region. Technical Sciences Seri*es. 2016; 1: 15-19

[4] D.L. Kreher and D.R. Stinson, Combinatorial Algorithms: Generation, Enumeration and Search, CRC press LTC , Boca Raton, Florida, 1998. 340p.

[5] W. Kocay and D.L. Kreher, Graphs, Algorithms, and Optimization, Chapman \& Hall/CRC Press, Boca Raton, Florida, 2005.

[6] CME342/AA220/CS238 - Parallel Methods in Numerical Analysis. Graph Partitioning Algorithms. Stanford University. Stanford, California. URL: http://web.stanford.edu/class/cs238/lect_notes/lecture12-13-05.pdf

[7] Gladkov L.A. The bioinspired methods in optimization / L.A. Gladkov, V. V. Kureychik, V. M. Kureychik, P. V. Sorokoletov. – M.: FIZMATLIT, 2009. - 384 pages.

[8] Kureychik, V. V. Theory of evolutionary calculations / V. V. Kureychik, V. M. Kureychik, S. I. Rodzin. – M.: Fizmatlit, 2012. – 260 pages .

[9] Rutkovskaya D. Neural networks, genetic algorithms and indistinct systems / D. Rutkovskaya, M. Pilinsky, L. Rutkovsky. – M.: The hot line – Telecom, 2013. – 384 pages.

[10] Karpenko A.P. Modern algorithms of search optimization. The algorithms inspired by the nature: manual / A.P. The Karpenka – M.: MGTU publishing house of N.E Bauman, 2014. – 446 pages.

[11] Nicolai M. Josuttis. The C++ Standard library: A Tutorial and reference (2nd edition) / Nicolai M. Josuttis // Publisher Addison-Wesley Professional. - 2012. - 1128p.

[12] Musser D.A., Stepanov A.A. Generic Programming / David A. Musser, Alexander A. Stepanov // Proceeding of International Symposium on Symbolic and Algebraic Computation, vol. 358 of Lecture Notes in Computer Science, pp. 13–25, Rome, Italy, 1988.

[13] Stepanov A.A., Mac-Jones P. Beginnings of programming: Transl. from English / A.A. Stepanov, P. Mac-Jones//M.: LLC I.D. Williams, 2011. – 272 pages.

[14] Petrochenkov A. Practical Aspects of Genetic Algorithms' Implementation in Life Cycle Management of Electrotechnical Equipment. Proc. of the 3rd International Conference on Applied Innovations in IT, (ICAIIT), March 2015. URL: http://icaiit.org/proceedings/ 3rd_ICAIIT/1.pdf

[15] Grinchenkov D.V., Kushchy D. N. Methodological, technological and legal aspects of use of electronic educational resources. // *University News. North-Caucasian Region. Technical Sciences Seri*es. 2013. No. 2 (171). Page 118-123.

Electric Submersible Pumps in Oil Production and Their Efficiency Analysis

M. I. Khakimyanov, I. N. Shafikov, F. F. Khusainov
Ufa State Petroleum Technological University - Faculty of Electrical Engineering and Electrical Equipment,
Kosmonavtov Str.1, 450062, Ufa, Russia
E-mail: hakimyanovmi@gmail.com, shafikov_in@mail.ru, fanur1991@rambler.ru

***Abstract*—The paper describes the structure of electricity consumption of electric submersible pumps. Analyzed the loss of all elements of the pumping unit: the motor, the cable line, transformer, control stations and pump. Discusses how to optimize power consumption of oil production.**

***Keywords:* electric submersible pump, submersible motor, power consumption, power, power loss, oilwell.**

I. Introduction

At the present time issues of energy and resource saving become especially actual in almost all industries. Fuel-power complex, which are engaged in extraction, preparation, transportation and refining of oil and gas are no exception [1]. Oil and gas companies are forced to develop and implement the whole complexes of energy-efficiency measures. Energy audit of all technological processes should be performed. This helps to identify units where the efficiency of the use of energy resources is not high enough and there is the potential for savings [2].

Optimization of well pumps operation allows not only to reduce energy consumption, but also extend the life cycle of electrical equipment [3][4][5].

With this purpose automatic systems for commercial accounting of power consumption have become widely implemented at the industrial enterprises, allowing continuous monitoring of the efficiency of energy consumption by equipment.

II. The Structure of the Electricity Consumption by at the Oil and Gas Companies

The structure of the energy consumption by oil and gas companies is shown in Fig. 1. The most energy-intensive technological process in all oil and gas industry is artificial oil lift by borehole pumps - up to 55 ... 62% of the total electricity consumption [6]. The main method of oil well operation is using of electric submersible pumps (ESPs). Over 54% of all wells in Russia operated by ESP, with these oil-wells about 75% of total oil is extracted [7]. Therefore, optimization of artificial oil production process of by using ESP can provide significant energy saving effect for the enterprise.

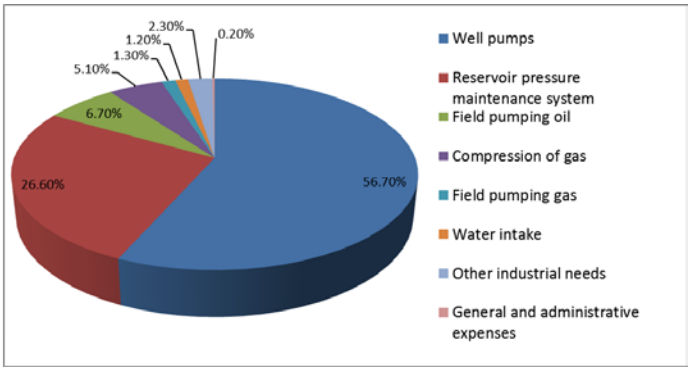


Fig. 1. The structure of the electricity consumption by various technological processes at the oil and gas companies.

III. Structure of Power Losses in Well Pump

It is necessary to analyze energy consumption by electrical drives of the submersible pumps to estimate efficiency of energy using during electricity consumption planning. For this purpose it is necessary to have a methodology for calculating the energy consumption depending on the installed equipment and technological parameters of the oil-well [8].

The structure of the electric power losses in ESP units is given in Fig. 2.

As seen in Fig. 2, the power consumption of ESP consists of useful power P_{UP} which is pump uses to lift the well fluid, as well as losses in all elements of the unit: in the pump ΔP_{CP}, in the upstream device ΔP_{UD}, in a wear sleeve ΔP_{WS}, in the submersible electric motor ΔP_{SEM}, in a cable line ΔP_{CL}, in a transformer ΔP_{TR}, in a control station ΔP_{CS}, in a line filter ΔP_{LF} and input filter ΔP_{IF} if any:

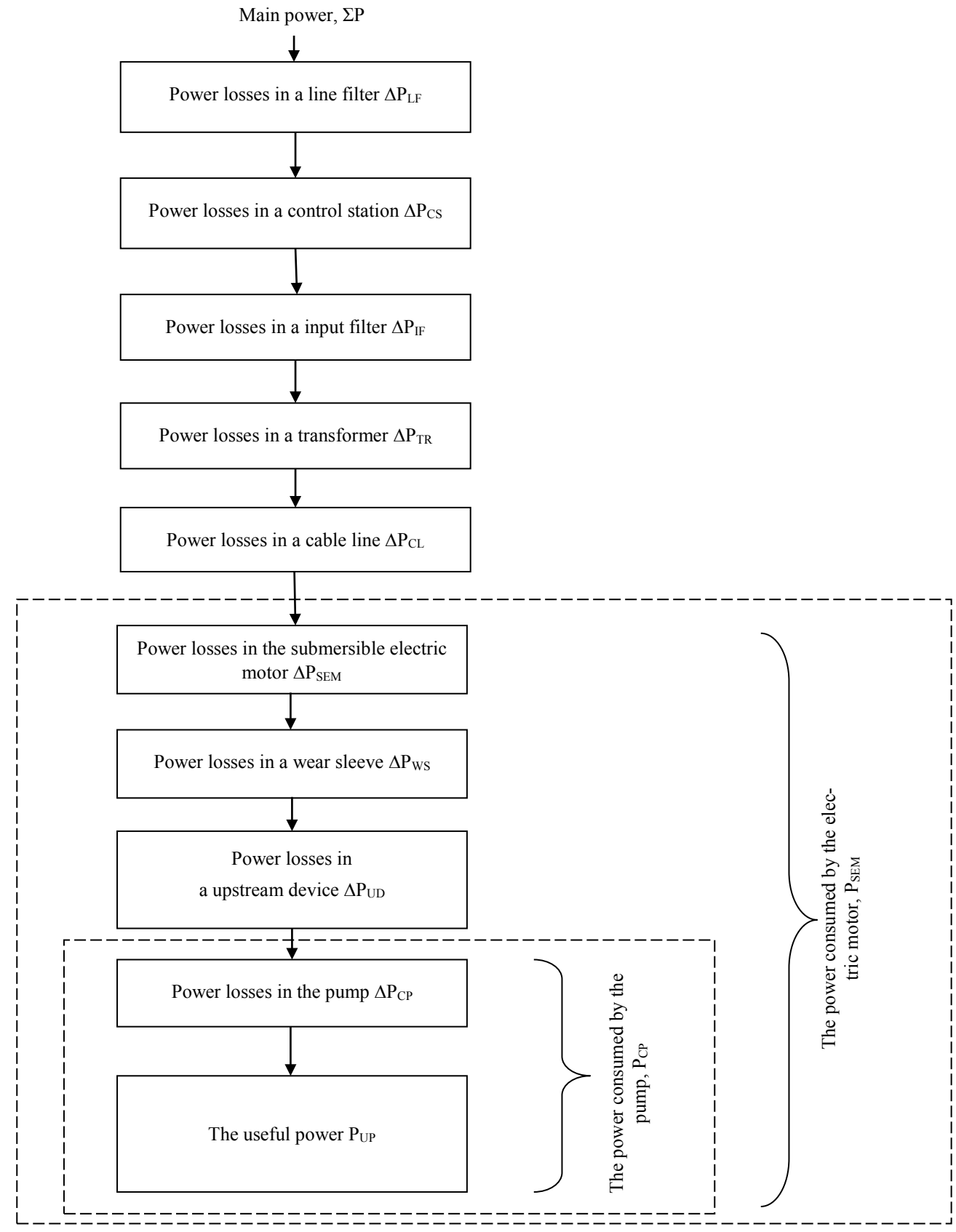


Fig. 2. Structure of power losses in ESP.

$$P_{\Sigma} = P_{UP} + \Delta P_{CP} + \Delta P_{UD} + \Delta P_{WS} + \\ + \Delta P_{SEM} + \Delta P_{CL} + \Delta P_{TR} + \Delta P_{CS} + \Delta P_{LF} + \Delta P_{IF}. \quad (1)$$

The first five terms in the expression (1) form a power consumed by the motor:

$$P_{SEM} = P_{UP} + \Delta P_{CP} + \Delta P_{UD} + \Delta P_{WS} + \Delta P_{SEM}. \quad (2)$$

Useful power P_{UP} used by a pump to lift the well fluid depends on many factors: pump rate, liquid density, dynamic head, casing head pressure, annular pressure, hydraulic resistance of the CNT, saturation pressure, gas oil ratio and other [9]. Therefore, the accuracy in the calculation of useful power will depend on the completeness of available data on the parameters of the well and equipment operating mode.

Power loss in the pump ΔP_{CP} is determined by its coefficient of performance (COP). In this case coefficient of efficiency of the centrifugal pump varies depending on outflow performance, peaking in nominal mode. Therefore, for definition of coefficient of efficiency of ESP for this outflow performance necessary to use the pump characteristic.

The power loss in such elements as the upstream device ΔP_{UD}, the wear sleeve ΔP_{WS}, the control station ΔP_{CS}, line filter ΔP_{LF} and the input filter ΔP_{IF}, are taken from the work documents. In the absence of typical values based on the installed capacitance can be taken.

The losses in the electric motor ΔP_{SEM} are determined by its coefficient of efficiency, which varies depending on the load. The value of motor efficiency at a given load is on the running characteristic [10].

The power loss in the cable ΔP_{CL} comprise a significant part of total energy consumption in ESP. The power losses in the cable lines are proportional to the square of the current:

$$\Delta P_{CL} = \frac{1{,}732 \cdot \rho \cdot L_{CL}[1 + \alpha(t - 20)] \cdot I^2}{F}, \quad (3)$$

where ρ – electrical resistivity of the cable, Ohm·m (for copper $\rho = 0{,}0195 \cdot 10^{-6}$ Ohm·m); α – temperature coefficient of expansion of copper (for copper $\alpha = 0{,}0041$); L_{CL} – cable length, m; t – average temperature of the cable, ° C; I – operating current, A; F – area of cross-sectional, m^2.

Power losses in double-wound transformer formed from iron loss and copper loss:

$$\Delta P_{TR} = \Delta P_{L} + \Delta P_{CL} = \Delta P_{L} + \Delta P_{SC}\left(\frac{S}{S_{NOM}}\right)^2, \quad (4)$$

where ΔP_{L} – no load loss given in the published data, W; ΔP_{CL} – copper losses, W; ΔP_{SC} – short-circuit losses, given in the published data, W; S – power of transformer load, V·A; S_{NOM} – nominal power of the transformer, V·A.

IV. Results of Calculation of the Power Loss

Results of calculation of the power loss in ESP elements using formulas (1), (2), (3) and (4) for the well № 1608 are shown in Table 1. Fig. 3 shows a diagram of illustrating percentage distribution of power loss between the setup items. Note that in this configuration there are no filters well.

As you can see from the diagram in Fig. 3, the useful power that is consumed on lifting well fluid, is only 22.7% of the total power consumption of ESP. Thus the loss in the centrifugal pump (47% of total consumption) is more than two times higher than the useful power. This is due to the relatively low efficiency of the centrifugal pump at rate - 32%. Also, significant part of energy consumption in the ESP comprise the motor loss (12.3%), in the cable line (6.4%) and the transformer (7.5%). The power consumed by the wear sleeve and the losses in the control station can be considered small (1.9% and 2.0%).

Table 1 - Power lost to the elements of ESP installations for the well № 1608

The element, wherein the power loss	Designation	Value	
		kW	%
The useful power	P_{UP}	4,77	22,7
Losses in the pump	ΔP_{CP}	9,90	47,0
Power of the wear sleeve	ΔP_{WS}	0,40	1,9
The motor loss	ΔP_{SEM}	2,60	12,3
The losses in the cable line	ΔP_{CL}	1,36	6,5
The losses in the transformer	ΔP_{TR}	1,58	7,5
The losses in the control station	ΔP_{CS}	0,44	2,1
The total power consumption of the installation	ΣP	21,05	100,0

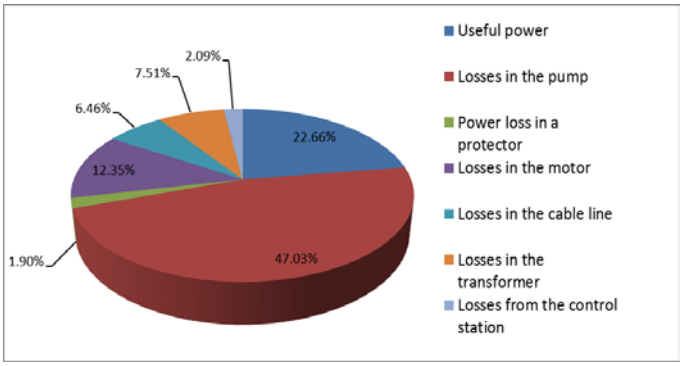


Fig. 3. Diagram of power losses on elements of ESP unit for the well № 1608.

V. Conclusion

The above distribution of power losses for the well № 1608 can be considered typical for wells operated by ESP. High electricity consumption for artificial oil lift of hydrocarbons crude are forced oil producers to look for ways to optimize energy consumption. The main areas of improvement of the ESP are the introduction variable-frequency drive, the use of the submersible telemetry units, and replacement of asynchronous motors at permanent magnet synchronous motor. Blocks submersible telemetry are transmitted to surface of the information about technological parameters on the bottom hole that allows to choose optimum operating mode for each a well. The permanent

magnet synchronous motor have an efficiency of 6 ... 8% higher than asynchronous motors [11], that, respectively, reduces the losses in the motor, power consumption and current. By reducing the current is also declining and the remaining loss: in the cable, transformer and control stations.

An important factor of optimizing operating modes of ESP is using intelligent control stations with a controller [12].

Thus and so, can be drawn the following conclusions:

1. Artificial oil lift is the most energy-intensive technological process in oil and gas companies, and therefore has the greatest potential for the introduction of energy saving measures.

2. Among the methods of artificial oil lift most relevant to optimize the energy consumption is using of ESP, as this method is exploited most of the wells of the Russian Federation (over 54%), which provide up to 75% of all produced oil.

3. The useful power of ESP expended by on lifting wellbore fluid, is only 20 ... 25% of total consumption, while the losses in the centrifugal pump can reach 47% or higher. Also, considerable power losses occur in the electric motor (12%), of the cable line (6%) and the transformer (7%).

4. Reducing the power consumption of the motor reduces losses in the following elements: cables, transformers, control station due to the current reduction.

REFERENCES

[1] A. R. Brandt, “Oil depletion and the energy efficiency of oil production: The case of California”, Sustainability. – 2011. – vol. 3. – №. 10. – pp. 1833-1854.

[2] M. Vazquez et al., “Global Optimization of Oil Production Systems, A Unified Operational View”, SPE Annual Technical Conference and Exhibition. – Society of Petroleum Engineers, 2001.

[3] A. Petrochenkov, “Regarding to Implementation of Genetic Algorithms in Life Cycle Management of Electrotechnical Equipment”, Proc. of the Second International Conference on Applied Innovations in IT, E. Siemens (editor in chief) et al. Kothen, Anhalt University of Applied Sciences, 2014. – pp.79-83. doi: 10.13142/kt10002.13.

[4] A. Petrochenkov, "Practical Aspects of Genetic Algorithms Implementation for Life Cycle Management System of Electrotechnical Equipment", Proc. of the Third International Conference on Applied Innovations in IT, E. Siemens (editor in chief) et al. Kothen, Anhalt University of Applied Sciences, 2015. – pp. 1-6. doi: 10.13142/kt10003.01.

[5] A. Petrochenkov, "Methodical Bases of the Integrated Electrotechnical Complexes Life Cycle Logistic Supports", Proc. of First International Conference on Applied Innovations in IT, E. Siemens (editor in chief) et al. Kothen, Anhalt University of Applied Sciences, 2013. – pp. 7-12. doi: 10.13142/kt10001.02.

[6] M.I. Khakimyanov, “Analysis of energy consumption for mechanized oil production electric submersible pumps”, Elektrotekhnicheskiye i informatsionnyye kompleksy i sistemy.– 2013.– vol. 9.– №3.– pp. 37-41. (in russian)

[7] M. I. Khakimyanov, “Energy intensity in artificial lift of sucker rod pumping units”, Vestnik UGATU. – 2014. – vol. 18. – №. 2 (63). – pp. 54–60. (in russian)

[8] J. L. Wu, Y. T. Liu, H. N. Yang, “A New Analytical Solution of the Productivity Equation for a Vertical Fractured Well in 3D Anisotropic Oil Reservoirs”, Petroleum Science and Technology. – 2014. – vol. 32. – №. 4. – pp. 433-441.

[9] J. L. Wu, Y. T. Liu, H. N. Yang, “A New Analytical Solution of the Productivity Equation for a Vertical Fractured Well in 3D Anisotropic Oil Reservoirs”, Petroleum Science and Technology. – 2014. – vol. 32. – №. 4. – pp. 433-441.

[10] G. Takacs, “Ways to Decrease Production Costs for Sucker-rod Pumping”, Internal report, Petroleum Engineering Department, University of Miskolc, Hungary, 2000.

[11] M.I. Khakimyanov, F.F. Khusainov, I.N. Shafikov, “Technological Parameters Influence of Oilwells on Energy Consumption Sucker Rod Pumps”, Electronic Science magazine “Oil and Gas business”.– 2015.– №1.- pp.533-563. URL: http://ogbus.ru/issues/1_2015/ogbus_1_2015_p533-563_KhakimyanovMI_ru_en.pdf (date of treatment: 02.02.2016)

[12] As.H. Rzayev, M.H. Rezvan, M.I. Khakimyanov, “Automation artificial lift systems in the CIS”, Electronic Science magazine “Oil and Gas business”.– 2013.– Issue 5.– pp. 19-29. URL: http://www.ogbus.ru/eng/authors/RzayevAsH/RzayevAsH_1.pdf (date of treatment: 02.02.2016)

One Approach to the Problem Solution of Specialized Software Development for Subject Search

Dmitriy Grinchenkov, Darya Kushchiy, Anastasia Kolomiets
Platov South-Russian State Polytechnic University (NPI)
Prosveshchenie Str. 132, 346400, Novocherkassk, Russia
E-mail: grindv@yandex.ru, dkushchiy@rambler.ru, anastasia.srstu@gmail.com

***Abstract*— in the article relevance of system development for subject search using computational linguistics is considered. The basic principles of system functioning are defined. The principle of grammar development for information retrieval from the partially structured text in a natural language is considered. The ranging principle of results of information search is defined.**

***Keywords:* information retrieval, subject search, semantic web, grammar, natural languages, formal languages.**

I. Introduction

Nowadays Internet is one of the most important sources of information in all fields of knowledge. However constantly growing volume of various structure text data makes more difficult process of subject search and determination of result practical importance for the end user. This circumstance negatively affects quality of information support for research of any sort. Further, as an example, we will consider subject search of electronic educational resources [1] in texts of working programs.

The main modern search services on the Internet such as subject directories and search services by keywords show a number of shortcomings when scientific information search. Firstly, correlation of the document with one or another category is not fully automatized, and secondly search results are influenced by limitation of demand assignment language. Thus, development of the specialized software for subject search is rather actual nowadays. [2]

II. Structure of Search System

In very general case the system of subject search can be presented in the form of the information and operating structural elements. Their interrelation is carried out by some mechanism L (Fig. 1).

A set of internal states of system is formed by a set of analyzed internet resources IR. Their processing is defined by transition function H from one resource to another (by change of internal states) (1):

$$H:(Q \bullet IR)^h \Rightarrow IR. \quad (1)$$

There are not only a set of entrance values Q (system request) and output values R (search results), but also operators of transformation its forms of representation for the user and the system F_Q and F_R in the structure.

Functioning of subject search system is time-spaced process T of information transformation from entrance value Q to output value R (2):

$$T:\{T \to Q; T \to R\}. \quad (2)$$

Generation a set of results is realized by algorithm called for system exit function E (3):

$$E:\{(Q \bullet IR) \to R\}. \quad (3)$$

Feedback mechanism for system regulation is denoted by parameter F. Thus system can be represented as an ordered set of elements (4):

$$S =< Q, R, IR, F_Q, F_R, L, H, T, F >. \quad (4)$$

The main distinctive feature of analytical search system from simple search system is the way of demand assignment and rating of found resources considering context of chosen subject field. View of documents search form can be fully or partially taken from the traditional search system [3].

III. Form of Query View

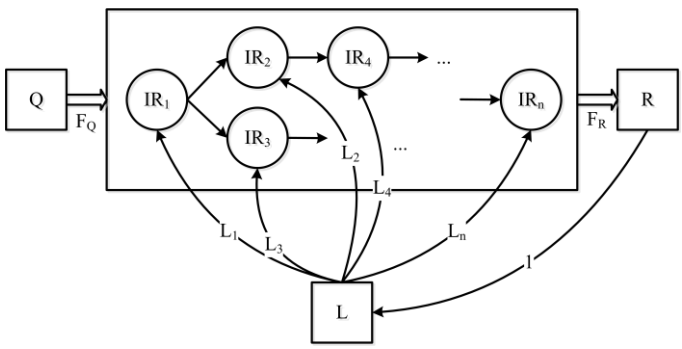


Fig. 1. Generalized structure of a search system.

In subject search inquire represents the model of subject field, where key words and relationship is determined. It is practical to use for it methods of computing linguistics [4].

As it was mentioned above, the working program of educational module serve as a source for search field in our case.

Formation of model sampling for creation of inquiry manually demands certain knowledge in subject domain and time. Those are the reasons for need in automation of this process.

Working programs of the Russian higher education institutions represent partially structured text which reflects subject domain and contains additional information on educational process with the repeating blocks. This fact causes low efficiency application of statistical methods for allocation of keywords by Tsipf method and the subsequent formation of terminological builds by t-score and MI-test [5][6][7]. In other words, for different disciplines will be formed almost identical retrieval of valid and syntactic words without display of a subject context For example, discipline, subject, section, literature, way, definition, target.

This problem is seen to be overcome by use of linguistic methods. In such documents linguistic analysis could be applied correctly only within the isolated text blocks. However despite some structuring, this text remains written in a natural language, and complexity of its analysis is caused by lack of the formal representation.

As it has been noted in some works [8][9], not all natural languages can be distinguished by regular grammars. Moreover, the existing languages may contain structures are recognizable only by context-dependent grammars [10].

However, use of context-free grammars at rules modeling provides good approximation to the truth and allows solving the majority of applied problems [11]. The semantic analysis of the text of the working program requires creation of the formal unified text structure, allowing fully displaying contents of structural blocks.

At the same time types and fields layout in the document from which text information is taken, define sense and communications of this information with other information in the document.

One more complexity consists of slippery character of information of a set of structural blocks. One reason lay in distinctive types of an academic load for different disciplines. For example, for one discipline the curriculum has only lectures and laboratory researches, and for another – lectures, laboratory researches, a practical training and a term paper. Other reason is covered in alternativeness and mobility of language norm borders and in statistical nature of separate information types [12].

It should be noted separately reasons for linguistic incomplete at formalization:

1. Continuous development of a natural language. It includes appearance of new language units, character change of the existing units and rules of their compatibility. Especially it is noticeable in sublanguages of new subject domains with not well-established terminology.
2. Language features of separate native speakers which could not be described and formalized today.

In Russian language there is much tension around this problem due to lack of rigidly regulated sentences construction.

Carrying out the text analysis of working programs the following stylistic features has been found: the text doesn't contain figural expressions, estimative adjectives, almost no adverbs; the natural language polysemy is minimized by use of in advance defined terms. The main language construction of text blocks is the grammatical basis with a number of additions.

The developing grammar of the working program leads to right-linear context-free grammar because of the choice of highly specialized area of a natural language and existence of attributes in grammar.

Actually this grammar is used for splitting a source text of the document for sections and processing most important of them for our task. For this purpose accurate observance of structure of the document is required, the working program consists of in advance defined sequence of sections.

Top level production rules serve for analysis of top level sections. Rules for analysis of sections consist of two parts: the first part serves for analysis of the section name, the second – for analysis of the section text content. Symbols of such grammar can have syntactic attributes. Names of semantic attributes are specified in attributes of nonterminal symbols. In attributes of terminal symbols syntactic text attributes can be specified in addition. Comparison of words at analysis is made taking into account their morphology.

We will consider a fragment of the developed grammar, provided by xml-format:

```
…<global-rule id="Section4" comment =
"4.SODERZHANIE DISCIPLINI ">
<rule><rulerefuri="#Section4Name"/>
<rulerefuri="#Section4x"/></rule>
  </global-rule>
  <global-rule id="Section4Name" sectionPart ="Name"
comment="Заголовок раздела 4"><rule><clauseType=
"NEOPRED"/></rule></global-rule><global-rule
id="Section4x" frame= "SubTitle" frameSlot="Title"
comment="4.1 KONTAKTANAYA AUDITORNAYA
RABOTA""or"comment="4.2 SAMOSTAYATELNAYA
RABOTA " "or"comment="4.3 KONTAKTNAYA
VNEAUDITORNAYA RABOTA"
</rule><ruleref uri="#Section4xContent" /></rule>
  </global-rule>
  <global-rule id="Section4xContent" section-
Part="Content"
comment=""><rule><rulerefuri="#Section42xInputs"
minOccurs=""/><rulerefuri="#Section42xOutputs"minOccu
rs=""/></rule>
  </global-rule>
  <global-rule id="Section4xInputs" comment="TEMA">
<rule><sentence/><clause/><rulerefuri="#Input"maxOccurs
="unbounded"/>
  </rule></global-rule> …
```

In structure of the working program keywords are highlighted. Those keywords determine ownership of section to the certain nonterminal.

In rules of grammar there are syntactic attributes and attributes which specify degree of the rule implementation:

1. Name – the text contain the name of the section.
2. Content – the text describe section contents;
3. Clause – clause;
4. Clause NEOPRED – the clause which does not have sense for the description of system structure;

5. Clause TIRE – a fragment with a dash;
6. GENIT_IG group – the nominative group connected by a genitive case.

The nominations accepted in the Penn Treebank project are used in names of nonterminal symbols. [13]

Sematic text representation consist of semantic representation of separate sentences, which elements are definition retrieved from the analyzed text and its semantic relations (Fig 2) [14][15].

Semantic representation of separate sentences is described by algebraic system, similar to the graph with definitions as a vertexes, any edge is marked to the semantic relations and connects those vertex-definitions which are with each other in this relation.

IV. Assessment of Search Results Relevance

The selection constructed according to the working program by search is considered as the reference text. Text representation of inquiry is formed using linguistic analysis methods at a search query. Data selection is carried out using normal form of words from special data structure containing information on word usage in texts of documents collection in which search is carried out. This structure of data contains information on documents texts according to text representation. To fill this structure texts of documents were preliminary analyzed by linguistic method [16][17].

Unlike of exact-match search function [18][19] offered approach will solve problem of sentential search by finding sentences corresponding to inquiry lexically and syntactic, but differ in a form and order of word usage (Fig 3).

By sentential search the sentences in the found documents are compared with the inquiry sentence so that at least one phrase will match to one in inquiry.

Under the match of phrases we mean presence of all

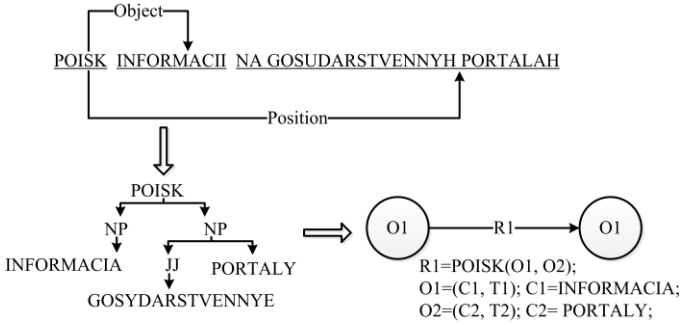


Fig. 2. Semantic image of offers.

syntactic links at the corresponding word usage as a part of phrases in reference sentence and found sentence.

The principle of results ranging of information search is defined (5) where the following notations have been accepted:

ref(r_i) – the document weight given by basic algorithm of ranging BM25 [20];

List – amount of the given results on the search page;

AllList – amount of all search results satisfying to inquiry *q*;

col_key(K_i,C_q) – the weight coefficient characterizing collocations among matches of a set of keywords to the text of the document.

$$F(r_i) = ref(r_i) + col\_key(K_i, C_i);$$

$$yes\_col(r_i) = ref(r_i) \cdot \left(1 - \left(\max ref - ref(r_i)\right)\right);$$

$$not\_col(r_i) = \left(\max ref - ref(r_i)\right) \cdot \left(1 - ref(r_i)\right);$$

$$col\_key(K_i, C_i) = \begin{cases} \left(2^{yes\_col(r_i)} - 1\right) \big/ norma, K_i \in C_i \\ \left(-2^{yes\_col(r_i)} - 1\right) \big/ norma, K_i \notin C_i \end{cases};$$

$$norma = 2^{\max ref} \sum_{j=1}^{List} \log_2(1 + j);$$

$$\sum_{j=1}^{AllList} ref(r_i) = 1; \max ref = \max_R \left(ref(R)\right).$$

(5) Ranging function has to satisfy orderliness property, which means that for couple of documents (r_i, r_j) holds $F(r_i) \geq F(r_j)$, if the document rr_i more corresponds to inquiry q, than the document rr_j.

The complex algorithm of relevant information finding is presented with the help of the flowchart (Fig 4).

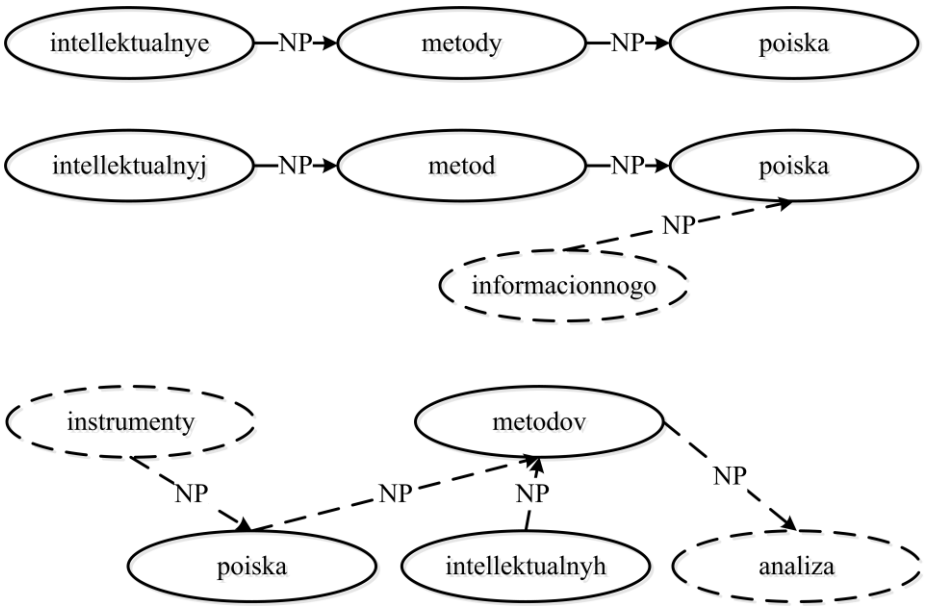


Fig. 3. Options of relevant search results.

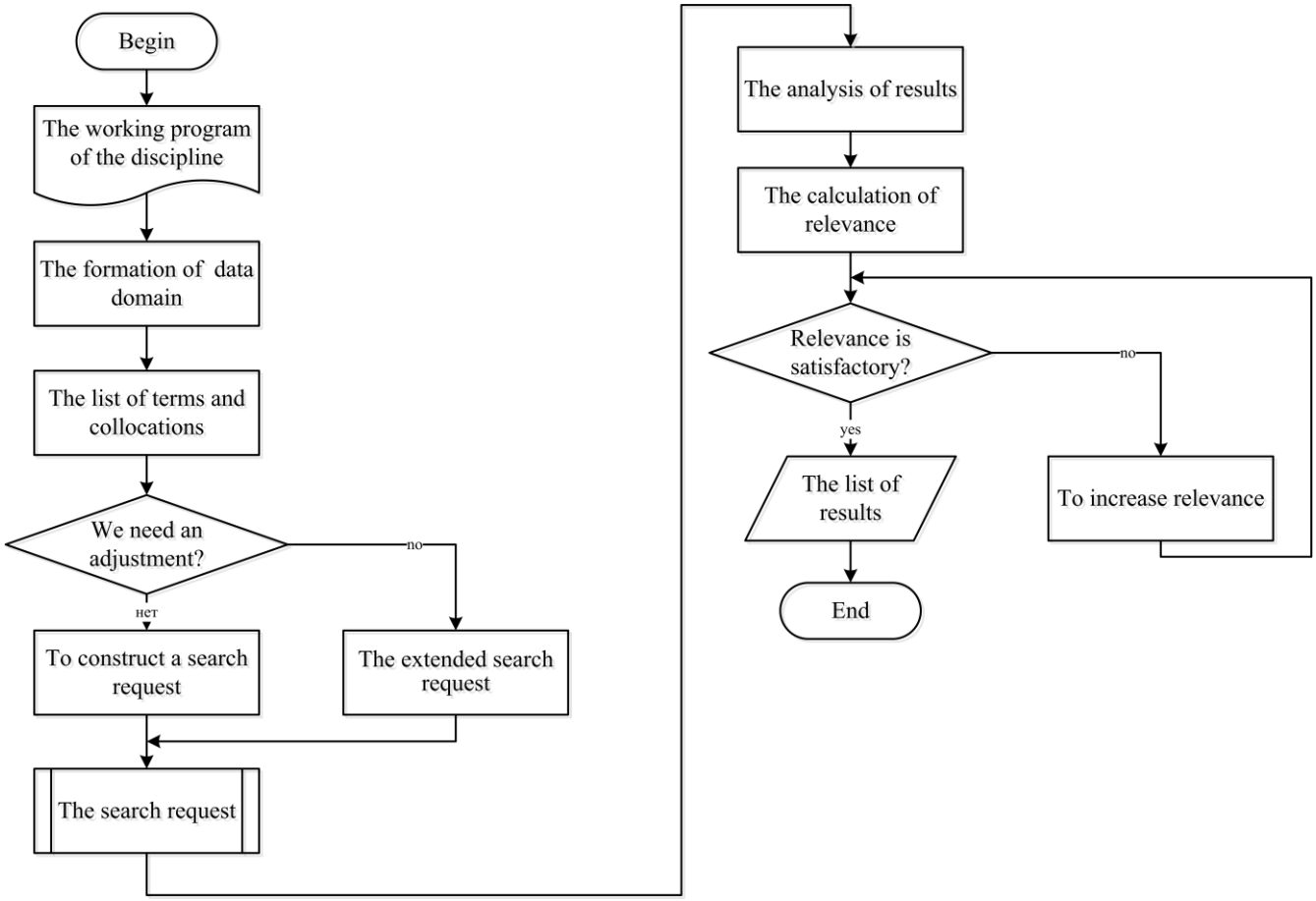


Fig. 4. Complex algorithm of relevant information finding.

V. CONCLUSION

Comparison of efficiency of simulated system and traditional information retrieval systems has been carried out with the aim of efficiency assessment of the developed models and algorithms [21]. Results were compared to a reference system – the hypothetical system finding all available relevant to this inquiry documents. Comparison was carried out by the number of the relevant documents issued by systems. (Fig 5)

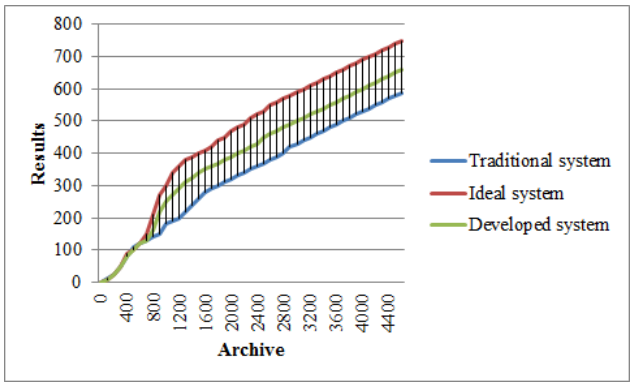


Fig. 5. The number of the issued relevant documents

It should be noted that distinctions in efficiency are seen in process of increase in volume of the worked out array of data. At the small sizes of archives (up to 800 documents) distinctions in results is becoming more and more obvious. The collection of pages of Wikipedia (about 2500 documents) was used for simulation. At such size of archive the difference in number of the issued relevant documents makes about 15-20%.

The further plan of work includes working out increase in efficiency of the developed algorithms on large volumes of text collections.

REFERENCES

[1] D.V. Grinchenkov, D.N. Kushchiy, "The methodological, technological and legal aspects of using the e-learning resources," J. University News. North-Caucasian Region. Technical Sciences Series, no.2, pp.118-123, 2013.

[2] D.V. Grinchenkov, D.N. Kushchiy, "Topicality and principles of construction of the intellectual information system for forming methodical support of educational disciplines on the basis of internet resources," J. University News. North-Caucasian Region. Technical Sciences Series, no.3, pp.114-119, 2014.

[3] D.V. Landeh, A.A. Snarskij, I.V. Bezsudnov, Internetika: Navigation in difficult networks: models and algorithms, Moscow: LIBROKOM (Editorial URSS) Publ., 2009, pp. 142-153. (rus)

[4] I. A. Tikhomirov, I.V. Smirnov, "Applying linguistic semantics and machine learning methods to search precision improvement in search engine "Exactus"," Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue-2009"(Bekasovo, Russia, May 2009), issue 8(15), 2009, pp. 483-487.

[5] D. Gildea, D. Jurafsky, "Automatic labeling of semantic roles," J. Computational Linguistics, vol. 28, no. 3, pp. 245-288, 2002.

[6] D.V. Grinchenkov, D.N. Kushchiy, "On selection of keywords in the contents of the framework curriculum of academic disciplines," Traditions of the Russian engineering school: yesterday, today, tomorrow: collection of scientific articles on problems of higher school, Novocherkassk, Platov South-Russian State Polytechnic University (NPI) Publ., Nov. 2015, pp. 114-117. (rus)

[7] D.V. Grinchenkov, D.N. Kushchiy, "The formation of fuzzy collocation on the basis of semantic analysis of the framework curriculum of academic disciplines," Problems of modernization of engineering education in Russia: collection of scientific articles on problems of higher school, Novocherkassk, Platov South-Russian State Polytechnic University (NPI) Publ., Oct. 2014, pp. 298-300. (rus)

[8] N. Chomsky, Syntactic Structures, 2nd ed., Berlin: De Gruyter Mouton, 2002, pp. 112-115.

[9] R. Huybregts, "The weak inadequacy of context-free phrase structure grammars," In Proc. Van Periferie naar Kern, Dordrecht, The Netherlands, 1984, pp. 81-99.

[10] M. Shieber, "Evidence against the context-freeness of natural language," J. Linguistics and Philosophy, vol. 8(3), pp. 333-343, Aug. 1985.

[11] M.M. Gavrikov, A.N. Ivanchenko, D.V. Grinchenkov, Theoretical bases of development and realization of programming languages. The manual for students of the higher education institutions which are trained in "The software of computer facilities and the automatized systems" direction of training of certified specialists "The information scientist and computer facilities", Moskow: KNORUS, 2010, pp. 134-149. (rus)

[12] M. Pennacchiotti, P. Pantel, "Entity Extraction via Ensemble Semantics," Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009), Singapore, Aug. 2009, pp. 238-247.

[13] M.P. Marcus, M.A. Marcinkiewicz, B. Santorini, "Building a large annotated corpus of English the Penn treebank," J. Computational Linguistics, vol. 19, no. 2, pp. 313-330,1993.

[14] R. McDonald, F. Pereira, K. Ribarov, J. Hajic, "Non-projective dependency parsing using spanning tree algorithms," Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05), Stroudsburg, PA, USA, 2005, pp. 523-530.

[15] R. Mcdonald, J. Nivre, Y. Quirmbach-Brundage et al., "Universal dependency annotation for multilingual parsing," Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, vol. 2, Aug. 2013, pp.92-97.

[16] A.V. Sokirko, " Bystroslovar: morphological prediction new Russian words using very large corpora," Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue-2010"(Bekasovo, Russia, May 2010), issue 9(16), pp. 450-456, 2010.

[17] Leontyeva N. N. About theory of automatic understanding of natural texts. P.3: Semantic component. Local semantic analysis, Moscow: MSU Publ., 2002, pp. 27-32. (rus)

[18] D.V. Grinchenkov, D.N. Kushchiy, "Principles of software development for support of decision-making based on integreted expert estimates ," J. University News. Electromechanics, no. 5, pp. 69-73, 2012.

[19] R. Johansson, P. Nugues, "Dependency-based syntactic-semantic analysis with PropBank and NomBank," Proceedings of the Twelfth Conference on Computational Natural Language Learning (CoNLL '08), Stroudsburg, PA, USA, pp. 183-187, 2008.

[20] K. Sparck Jones, S. Walker, S. E. Robertson, "A probabilistic model of information retrieval: development and comparative experiments (part 1)," J. Information Processing and Management, vol. 36(6), pp. 779-808, 2000.

[21] I. Luzyanin and A. Petrochenkov "Regarding Information Systems Dependability Analysis" Proceedings of the 3rd International Conference on Applied Innovations in IT (2015). Jg. III. Koethen : Hochschule Anhalt, 2015, pp. 7-11 (DOI: 10.13142/kt10003.02, Anhalt University of Applied Sciences Digital library).

Algorithm for Estimation of Electricity Consumption Reduction by Using Variable Frequency Drive of Main Line Pumps (February 2016)

V. A. Shabanov, E. F. Khakimov,
Ufa State Petroleum Technological University – Automation Department,
Kosmonavtov Str. 1, 450062, Ufa, Russia
E-mail: shabanovva1@ya.ru

***Abstract*—The article is about using of variable frequency drives for reduction oil pumping main line pumps energy consumption. Block diagram of developed computer program is shown in the article. The computer program allows to determine the reduction of energy consumption and to estimate payback period of variable frequency drives.**

***Keywords:* main line pump, oil pipeline, variable frequency drive, efficiency, oil pumping station.**

I. Introduction

Nowadays main line pumps are connected in series for purpose of oil pumping. In this process dozens of pumps are used on a process section, consisted of a number of oil pumping stations (OPS), in "from pump to pump" mode. Number of working pumps depends on a required pipeline capacity. Regulation of the pipeline capacity by means of changing of number of working on a process section pumps or their parameters are main methods of operating modes regulation. [1][2]. However, using this method implies that pipeline capacity may differs from rated pumps capacity. Thus in this case main line pumps will have efficiency less than rated. When required capacity cannot be achieved by changing of working pumps number, method of cyclic pumping and pressure reducing valves (PRV) are used. Changing of working pumps number and using of PRV implies additional power losses and ineffective energy using [3][4]. That is why using of variable frequency drives (VFD) on main line pumps is possible solution for reduction of OPS energy consumption [5][6]. The objective of this article is to estimate reduction of power consumption for oil pumping by means of using VFD.

II. Object Analysis

The main documents for operated oil pipeline are: technological modes map and modes compliance information. Technological modes map contains all pipeline operation modes that can be established. Modes compliance information contains information about used in fact modes and used modes measured parameters.

Evaluation of pumps energy consumption changing in case of using VFD is very important for estimation of advantages VFD of pipeline pumps. One of the biggest problems in this issue is processing of bulky database. This problem is also vital for other IT sections [7][8].

In case of using VFD with main line pumps reduction of energy consumption is achieved by two main reasons: elimination of power losses in PRV and increasing of pumps efficiency by means of speed reduction in modes with low pipeline capacity [9][10].

Power losses in the valves ΔP_{PRV} are in direct ratio with the pressure decreasing in the valves Δp_{PRV} when PRV are being used

$$\Delta P_{PRV} = \Delta p_{PRV} \cdot \mathrm{Q_p}, \quad (1)$$

where . $\Delta p_{PRV} = p_m - p_o$ (2)

p_m - pressure in the OPS manifold, N/m2.

p_o - OPS output pressure, N/m2

Qp – pipeline capacity, m3/s;

If VFD are being used, than there is no need of PRV using, because pressure decreasing is achieved by decreasing of rotation speed. Thus energy consumption is reduced by a value of losses in PRV.

In modes with low capacity efficiency of rotating with rated speed pumps is lower compare to efficiency of rotating with lower than rated speed pumps. Thus in such modes consumed by pump with decreased by VFD rotation speed power is lower than consumed by unregulated pump power.

$$\Delta P_{eff} = p_{dif} \cdot \mathrm{Q_p} \cdot \left(\frac{1}{\eta_{unr}} - \frac{1}{\eta_r} \right), \quad (3)$$

Where p_{dif} – differential pressure that equates difference

between pump output pressure p_o and pump input pressure p_i, N/m2

$$p_{dif} = p_o - p_i; \quad (4)$$

η_{unr} – efficiency of rotating with rated speed pump; η_r - efficiency of rotating with lower than rated speed pump.

Efficiency of rotating with lower than rated speed pump can be calculated using equation [4]

$$\eta_r = \eta_{rtd} - (q - \nu)^2 \cdot \eta_{rtd} / \nu^2, \quad (5)$$

where η_{rtd} – rated efficiency; $q = Q/Q_{rtd}$ – relative value of pump capacity; Q and Q_{rtd} – actual and rated capacities of pumps as following; $\nu = \omega/\omega_{rtd}$ - relative value of pump rotation speed; ω and ω_{rtd}– actual and rated rotation speed of pumps as following.

In case of using VFD instead of RPV reduction of power losses equates power losses in (1) and (3) in total.

$$\Delta P_{VFD} = \Delta P_{PRV} + \Delta P_{eff}. \quad (6)$$

The problem of using equations (1) - (6) for operating pipeline is necessity of using input data from technological modes map and modes compliance information. Technological modes map and modes compliance information consist of millions of numerical parameters about operating modes for a year. That is why for purpose of automatically calculations of energy consumption computer program must have algorithm of searching and picking information from different databases on every step of calculations. Main steps for performing calculations based on designed algorithm are shown on Fig. 1

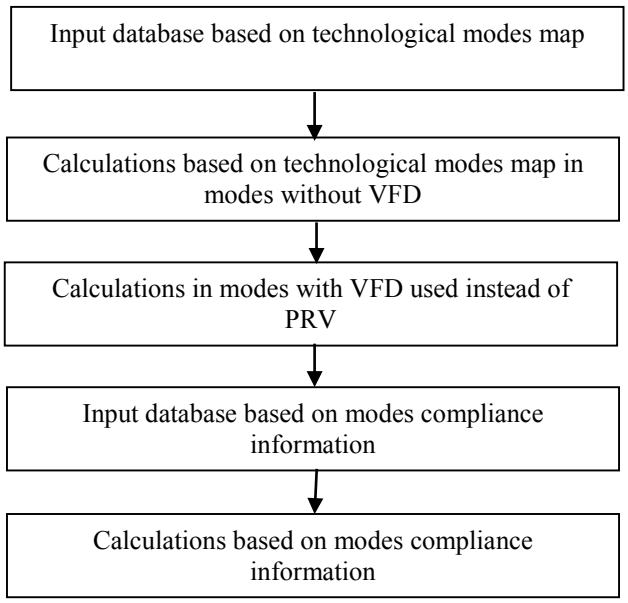


Fig. 1. Algorithm of calculation of energy consumption on an operating pipeline pumps

III. CALCULATIONS

Developed program for evaluating of energy consumption reduction in case of using VFD instead of PRV consist of following steps:

1. Creation database based on technological modes map and equipment information

1.1. Making «technological modes map» database. Part of this database is shown in table 1.

In the first column of the technological modes map number 001 means that there is only one main line pump working on the whole process section. The second column contains information about pipeline capacity per year, day, and two hours. For every OPS there is information about input, output, and manifold pressure, type (NM10000x210) and quantity (1) of using pumps. 1ch+№2(3) is pumps working mode. 1ch means that there is 1 working charging pump on the first OPS, №2(3) means that main line pump number 2 or 3 can be in work on the first OPS.

1.2 Equipment database creation that contains rated parameters and head-capacity curves coefficients for every pump

2. Calculations based on technological modes map in modes without VFD

2.1. Working database (WDB) is formed for calculations.

2.2. Head calculations for every pump

$$H = \Delta p_{dif} / (\rho \cdot g), \quad (7)$$

where g- gravity acceleration;

ρ – pumped oil density.

2.3. Efficiency for every main pipe pumps and charging pumps is calculated (ν=1 as pump are unregulated) by using (5). Q_{rtd} for every pump is taken from equipment database depending on pump number.

2.4. Calculation of power consumption for every electric drive. For these calculations electric drive efficiency, electric drive utilization factor $k_{u.f.}$, and electric drive load power $P_{u.f.}$ should be calculated.

2.5. Total amount of power consumption by all pumps P_Σ and specific energy consumption (energy consumption per hour, kW*h divided by pipeline capacity, t/h) are calculated.

$$W_S = \frac{P_\Sigma}{Q}. \quad (8)$$

An example of energy consumption calculations in modes without VFD is shown in Table 2

3. Calculations of same modes but in case of using VFD instead of PRV. Needed information is taken from calculations based on technological modes map in modes without VFD.

3.1. Calculation of power losses and head losses in PRV.

$$\Delta H_{PRV} = \Delta p_{PRV} / (\rho \cdot g). \quad (9)$$

3.2. Calculation of needed for elimination of the PRV head losses relative rotation speed of the main line pump,

$$\nu = \sqrt{1 - \frac{\Delta H_{PRV}}{a}}. \quad (10)$$

3.3. Using (5) pump efficiency in case of using VFD is calculated

3.4. Head produced in case of using VFD is calculated using equation

$$H = a \cdot \nu^2 - b \cdot Q^2. \quad (11)$$

In (10) and (11) a and b are coefficients of head-capacity curve

3.5. Power that consumed by regulated drives (with VFD) is calculated. Efficiency of the frequency transformer is taken into account

TABLE 1
PART OF «TECHNOLOGICAL MODES MAP» DATABASE FOR ONE OF THE MODES

Oil viscosity n = 14,43802 mm²/s, oil density ρ = 865,4 kg/m³								**Specific energy consumption**
Mode №	**Capacity**	**Parameter name**	**OPS1**	**OPS2**	**OPS3**	**OPS4**	**OPS5**	
1	2	3	4	5	6	7	8	9
	t/2h							kW·h/t
001	7891,67	Type and quantity of pumps	NM10000x210; 1	-	-	-	-	1,58
		Pumps working mode	1ch+№2(3)	-	-	-	-	
		P_i, kg/sm²	9,2	11,3	9,9	11,5	1,0	
		P_m, kg/sm²	33,7	11,3	9,9	11,5	1,0	
		P_o, kg/sm²	26,0	11,3	9,9	11,5	1,0	
		Power, kW	6225					

TABLE 2
AN EXAMPLE OF ENERGY CONSUMPTION CALCULATIONS IN MODES WITHOUT VFD

Mode	**Head using (7), m**	**Pump efficiency (5)**	**Pump input power, kW**	**Drive utilization factor**	**Drive input power, kW**	**Power losses in PRV, kW**	**Power consumption total amount P_Σ, kW**	**specific energy consumption kW*h/(t/h)**
004	233,7	0,88	5731,5	0,90	5878,4	1512,9	26282,7	3,33
001	282,8	0,63	4634,9	0,56	4519,4	1101,4	5937,1	1,50

TABLE 3
AN EXAMPLE OF POWER CONSUMPTION CALCULATIONS IN CASE OF USING VFD

Mode	**Q , (t/2h)**	**Δp_{PRV}, kg/sm²**	**ΔH_{PRV}= ΔP_{PRV}/(g*p), m**	**Pump rotation speed, v**	**∑P, kW**	**Specific power W_S, kW*h/(t/h)**
004	15750,00	5,3	61,1	0,87	24827,5	3,15
001	7891,67	7,7	88,89	0,81	3746,0	0,94

TABLE 4
MODES COMPLIANCE INFORMATION DATABASE

Date	**Operational mode**		**Working time, h**		**Capacity, t*1000/day.**	
	On a plan	**In fact**	**On a plan**	**In fact**	**On a plan**	**In fact**
12.01	001	001	10	10	37,9	38,6
13.01	002	002	5	5	33,5	32,9
17.01	004	001	18	18	141,8	142

TABLE 5
PART OF ENERGY CONSUMPTION REDUCING CALCULATIONS

Date	**Operating modes**	**ΔWs, kW*h/(t/h)**	**Q. t/2h**	**Working time, t, h**	**Reducing of energy consumption for the time t. ΔW, kW*h**	**Reducing of electricity cost for the time t, rubles*1000 / dollars**
01.01.2011	004	0,185	15750,00	7	10186,6	30,6/ 383
01.01.2011	001	0	10666,67	17	0	0
…	…	…	…	…	…	…
12.01.2011	001	0,564	7891,67	10	22244,3	66,7/834
…	…	…	…	…	…	…
Total for the month					298000,0	893,0/11163
…	…	…	…	…	…	…
Total for the year						7900,0/98750

3.6 Power that consumed by all left drives (without VFD) is calculated.

3.7 Total amount of power consumed by all regulated and all unregulated pumps is calculated.

3.8 Specific energy consumption (Power consumption by all pumps divided by pipeline capacity) is calculated.

Part of the power consumption calculation in case of using VFD instead of PRV on the first OPS for modes shown in Table 2 is shown in Table 3.

As can be seen, when comparing table 2 and 3, in mode № 004 specific power decreased by 0,18 kW*h/(t/h) (5%) by using VFD instead of PRV. In mode № 001 specific power decreased by 0,56 kW*h/(t/h) (37%) by using VFD instead of PRV

3.9 Difference between specific power in case of using PRV and specific power in case of using VFD ΔWs, kW*h/(t/h) is calculated.

This difference is reducing in power consumption by means of using VFD.

4 Creation of database based on modes compliance information for the chosen pipeline. Part of the database is shown in table 4.

5 Calculations based on modes compliance information.

5.1 Needed information for every used in fact mode is taken from previous databases

5.2 Decrease of power consumption by using VFD instead of PRV is calculated for every used mode for a month or year.

5.3. Total reducing of power consumption for the whole chosen period is calculated

5.4. Total reducing of electricity cost and payback period are calculated.

Part of total reducing of electricity cost and payback period calculations is shown in table 5. Taken price of 1 kW*h is 3 rubles (3,75 cents).

Calculations based on chosen process section showed that in case of low pipeline load (from 0,65 to 0,75 of rated) elimination of PRV using by means of VFD can reduce electricity costs by 8 million rubles (100 000 dollars) . In this case payback period would be 6,3 years (taken VFD cost is 50 Million rubles (625 000 dollars). If pipeline load is higher than 0,75, than payback period would be more than 20 years).

IV. CONCLUSION

1 Algorithm of calculation of main line pump power consumption reduction in case of using VFD in case of using PRV on oil pumping process section was developed.

2 Calculations showed that reducing of energy consumption and payback time for VFD depends on operating mode. Calculations based on chosen process section showed that in case of low pipeline load elimination of PRV using by means of VFD can reduce electricity costs by 8 million rubles (100 000 dollars). In this case payback period would be 6,3 years

REFERENCES

[1] A. A. Korshak, A. M. Nechval, Pipeline transportation of oil, oil products and gas. Ufa, DizaynPoligrafServis, 2001, pp. 76–132.

[2] L. A. Zaitsev, G. S. Jasinski, Regulation of the main oil pipelines modes. Moscow, Nedra, 1980, pp. 23-73.

[3] V. A. Shabanov, O.V. Kabargina "Advantages and prospects for the use of frequency-controlled electric drive of the main pumps on the PS," Quality management in the oil and gas sector, vol. 2. pp. 63-66, Dec. 2011.

[4] A. P. Tumansky "Optimization of modes of pumping through pipelines with pumping stations equipped with variable frequency drive," Transport and storage of petroleum products, vol. 8. pp. 11-14, 2005.

[5] V. A. Shabanov, O.V. Kabargina, Z. H. Pavlova. (2011, Nov) Evaluating the effectiveness of frequency control of main pumps. Oil and Gas Business [online]. vol 6, pp 24-29. Available: http://www.ogbus.ru/authors/Shabanov/Shabanov_8.pdf

[6] V. A. Shabanov, A. A. Ahmetgareev "To the selection of the optimal mode of operation of the main pump with VFD," Transport and storage of petroleum products and hydrocarbons, vol 3. pp. 7-10, May, 2012.

[7] I. Luzyanin, A. Petrochenkov "Regarding Information Systems Dependability Analysis" Proceedings of the 3rd International Conference on Applied Innovations in IT (2015). Jg. III. Koethen: Hochschule Anhalt, 2015, pp. 7-11 (DOI: 10.13142 / kt10003.02, Anhalt University of Applied Sciences Digital library).

[8] A. Petrochenkov "Regarding to Implementation of Genetic Algorithms in Life Cycle Management of Electrotechnical Equipment" Proceedings of 2nd International Conference on Applied Innovations in IT. Kothen, pp. 79-83 2014 (DOI: 10.13142 / kt10002.13).

[9] V.A. Shabanov, E.F. Khakimov, N. P. Pirozhnik “Energy Efficiency Analysis of variable-frequency electric current to the OPS by increasing the efficiency of the main pumps,” Oil and gas business. Scientific and technical journal. vol. 10, pp.55-60. Oct, 2012.

[10] V.A. Shabanov, E.F. Khakimov, S. F. Sharipova (2013, Feb) Analysis of the efficiency of the main pipeline pumps operated using a frequency controlled electric pressure regulators in function. Oil and Gas Business [online]. vol. 1. pp 324-333. Avaliable http://www.ogbus.ru/authors/Shabanov/Shabanov_16.pdf

[11] V.A. Shabanov, E.F. Khakimov, S. F. Sharipova (2013, Feb) “Algorithm for evaluating the effectiveness of variable frequency drive pumps trunk pipelines operated by the criterion of reducing energy consumption,” Electrical and information systems and systems. vol. 2, pp. 34-42. Sep, 2013.

High-speed UDP Data Transmission with Multithreading and Automatic Resource Allocation

Dmytro Syzov, Dmitry Kachan, Eduard Siemens
Anhalt University of Applied Sciences - Faculty of Electrical, Mechanical and Industrial Engineering,
Bernburger Str. 57, 06366 Köthen, Germany
E-mail: {dmytro.syzov, d.kachan, e.siemens}@emw.hs-anhalt.de

***Abstract*— In this paper a utilization of the high data-rates channels by threading of sending and receiving is studied. As a communication technology evolves the higher speeds are used more and more in various applications. But generating traffic with Gbps data-rates also brings some complications. Especially if UDP protocol is used and it is necessary to avoid packet fragmentation, for example for high-speed reliable transport protocols based on UDP. For such situation the Ethernet network packet size has to correspond to standard 1500 bytes MTU[1], which is widely used in the Internet. System may not has enough capacity to send messages with necessary rate in a single-threaded mode. A possible solution is to use more threads. It can be efficient on widespread multi-core systems. Also the fact that in real network non-constant data flow can be expected brings another object of study –- an automatic adaptation to the traffic which is changing during runtime. Cases investigated in this paper include adjusting number of threads to a given speed and keeping speed on a given rate when CPU gets heavily loaded by other processes while sending data.**

***Keywords:* high-speed data transport, threading, automatic resource management.**

I. INTRODUCTION

High-speed content delivery is a service that is more and more demanded by society over the time. And for some purposes, like a transmission of huge amount of data, TCP may be not appropriate and another reliable transport protocol is needed. UDP serves as a base for such protocols.

Protocols that are built on top of UDP (for example: RBUDP [2], UDTv4 [3], RWTP [4], RMDT [5]) rely on the message consistency. Thus, it is crucial that MTU of the generated message will not be bigger than maximum size of the data field of the frame used in channel layer. This requirement especially important for high-speed transport protocols and solutions [6][7]. For Ethernet standard MTU size is 1500 bytes.

Unlike TCP sockets, UDP sockets preserve message boundaries [8]. This is why it is safe to queue multiple calls to the system as message consistency will not be violated. This fact allows creation of a multi-threaded UDP sender and receiver, which can be used in different kinds of UDP-based protocols for reliable transmission. Flexible thread handling with automatic resource allocation and control over threads parameters can simplify development.

For threads management there are certain problems to be solved:

1. Data rate control.
2. Automatic resource allocating.

For send rate regulation certain means need to be implemented for the ability to generate messages with a constant data-rate and change it in a runtime. To automatically allocate resources – an algorithm of making decisions about allocation that are based on information which can be collected without significant overhead should be added . Because of the fact, that traffic parameters can vary or amount of system resources, that are available for the application, may be changed while the application is running – mentioned functionality is important for such a system.

II. RELATED WORK

Research [9] shows the basic problems of traffic generating for a 40 Gbps channel. It is comparing different traffic generators for network testing: DITG, packETH, Ostinato. For experiments held in [9] payload varies from 64 to 8950 bytes. Both TCP and UDP traffics were measured. Results describe the exact problem that is studied further in this work: none of traffic generators is capable of achieving the full bandwidth utilization, unless packets with the high payload were used. In case of D-ITG even using 8950 bytes payload was not sufficient. Although, authors in [9] use 40 Gbps link, the problems remains for a 10 Gbps link, as on packet size of 1500 bytes, traffic generators were unable to achieve 10 Gbps.

Solution, suggested in [9], is to utilize system resources by using threading. It shows that, when D-ITG is using 16 threads, the higher data-rate can be achieved with less payload per packet, in comparison with D-ITG running in a single-threaded mode. Same approach for achieving higher data-rates is used in this work and researched deeper.

In paper [10] authors also suggest using multi-threading for sending and receiving. Among other subjects studied, research shows bottlenecks of achieving close to channel capacity performance. Using threading helps to get high bandwidth utilization for channels with high data-rates. The handling of such systems is studied in [10] by researching the effects of system parameters on a throughput.

In this work handling of multiple threads and management of them is studied further from algorithmic and practical point of view.

III. METHODS

Several algorithms for solving problems described in introduction are presented in this section. They are implemented using "High Precision Timer" library [11] to get accurate time as they heavily rely on time stamps and to implement precise thread sleeping mechanism.

Data rate can be controlled in different ways:

1. Basic principle is blocking(force the thread to sleep some time) each thread for a specific time to decrease data-rate. They may have the same delay time or different if necessary. After each sending, time of the next message sending is calculated. It can be done by adding to the previous expected time a given time delay.
2. Other way is to add delay to the current time after sending.

The first approach will have different effect than the second. If a sender process is constantly delayed for some time due to some external influence, for example the other process is using the same CPU resource, and after that resource has been freed – the thread will start sending packages frequently until the real sending time will meet ideal expected time. The second approach does not give rate more than requested and application will try to keep the data rate constant.

To get the ability to change speed during runtime threads periodically update information about size of its inter-packet time interval. Threads do not necessarily have the same send rate, it can be assigned individually.

Initial inter packet interval time for each thread is calculated by formula (1):

$$T_{interval} = \frac{MSS \times 8 \times N_{th}}{R}, \quad (1)$$

where $T_{interval}$ – an inter-packet interval for a thread; MSS – MSS of a packet to be sent by a thread in bytes; 8 – constant amount of bits in one byte; N_{th} – given amount of threads; R – a data-rate of the whole multi-threaded system in bytes per second.

Other method is used to automatically handle threading. The information, used to make a decision about running or stopping sender thread is:

- the time of a full send loop (that can include, apart from send() system call, message generation, additional calculations, etc.),
- time of sending,
- time lag – the difference between actual sending time and expectation time (time when message should be sent to achieve given speed).

More precise, the time lag can be sufficient for a decision to spawn send processes, but additional time-related data give more information about sender behavior.

Approach is based on the assumption that if a total time lag of all threads is higher than zero, then capacity of existing threads is not enough. If it is less than zero, it is assumed that existing threads have more capacity than needed to achieve requested rate.

Theory behind this assumption is that if a thread can not send within a given time interval it will accumulate within each iteration the difference between target time for sending and factual time – time lag. Another case, when thread can perform the send operation within shorter period than the given inter-packet interval, this will result in an accumulation of the time it has to wait by blocking itself manually. Resulting lag – positive or negative difference – is used to evaluate current performance. Flow chart that illustrates algorithm is presented in the Fig. 1.

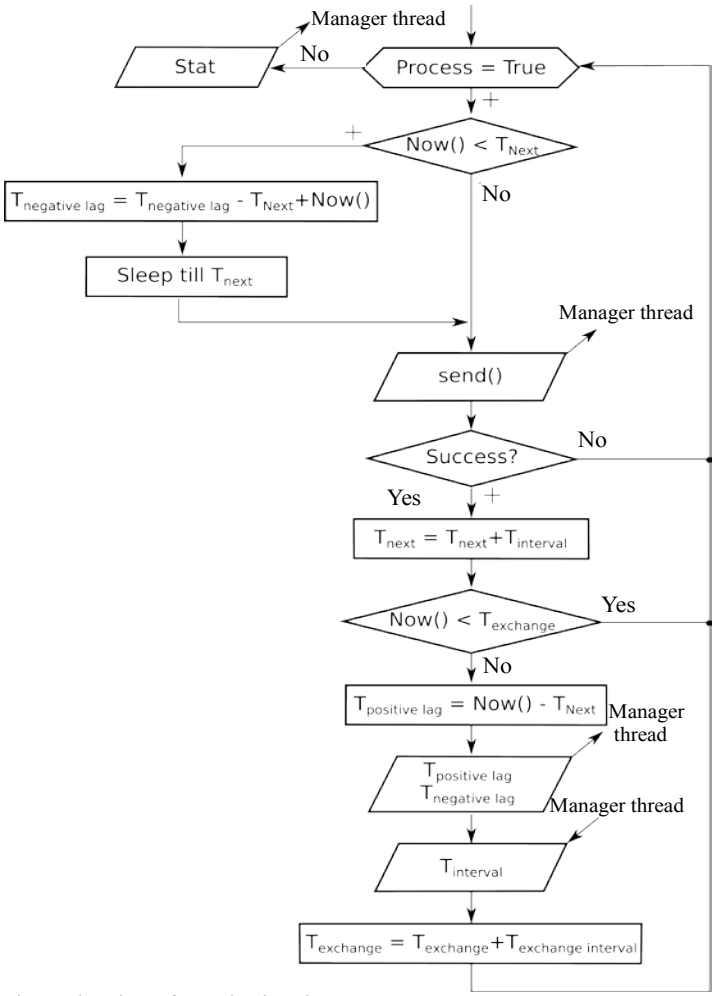


Fig. 1. Flowchart of a sender thread

in the Fig. 1:

$T_{interval}$ – the inter-packet interval for a thread; $Stat$ – a structure that contains the information about time parameters and collected at the end of the sending session; $T_{positivelag}$ – variable that accumulates the difference between expected and factual time of sending messages; $T_{negativelag}$ – variable that accumulates the time spent on waiting, caused by manual blocking; $Process$ – boolean variable that is used to stop thread; T_{next} – time when next send call should be made; $Now()$ – "High Precision Timer" function that provides current time; $Send()$ – UDP send call; $T_{exchange}$ – time when next data exchange between sender and main threads should be

made; $T_{exchangeinterval}$ –interval between such exchanges.

One specific thread manages all others to perform the functionality described previously: collecting information from other threads, adjusting necessary parameters and initiating spawn of a thread. This manager thread is blocked most of the time, and it unblocks periodically for performing its functionality.

An ideal case is when necessary rate is achieved without blocking and, thus, system resources are maximally utilized. If the requested data rate is not equal to a multiplication of a certain amount of threads' maximum generating rate – then, to achieve given rate, some thread blocking must be performed and at least one of the threads will have time lag less than zero.

IV. Testbed Topology Description

The core element of the tested topology is WAN emulator AppositeNetropy 10G [12] that can be used to create an emulation of WAN links with different impairments such as packet loss ratio up to 100%, delays of up to 100000ms and delay jitter with an accuracy of about 20ns. The Emulator allows a transmission of Ethernet traffic with an overall throughput of up to 21 Gbps on both, copper and fiber optic links. Apart from Netropy, setup contains two PC servers. They are connected via an Extreme Networks Summit x650 10 Gbps Ethernet switch and the WAN Emulator. Fiber optics with 10 Gbps bandwidth acts as a medium for transmission between compartments. There is no background traffic used for experiments, since in the focus of presented investigations is research of the pure traffic generation.

Each server is equipped as follows:

−CPU: Intel Xeon X5690 @3.47GHz;

−RAM: 42 GiBytes (speed 3466 MHz);

−OS: Linux CentOS 6.3;

−NIC: Chelsio Communications Inc T420-CR, 10Gbps

Also, for comparison of the performance on a different system, some tests were performed on servers with different CPU: *Intel Xeon E2630 @2.30GHz.*

V. Experimental Results

Firstly, the performance of the sender is compared for a different amount of threads, MSS and speed. MSS is taking following values: 1024, 1472, 8972 bytes. Data-rate that were tested is 10 Gbps.

First experiment is for MSS of 1472. Results of the experiment for 10 Gbps rate are presented in Fig 2. Since 1472 bytes MSS corresponds to 1500 bytes MTU in Ethernet [1] this test is especially important to the experiment as its results are of interest for transport protocols, used in TCP/IP networks, which often use Ethernet technology on the channel level. As was mentioned in the introduction, the consistency of a message is a hard condition for the protocols built on top of the UDP protocol.

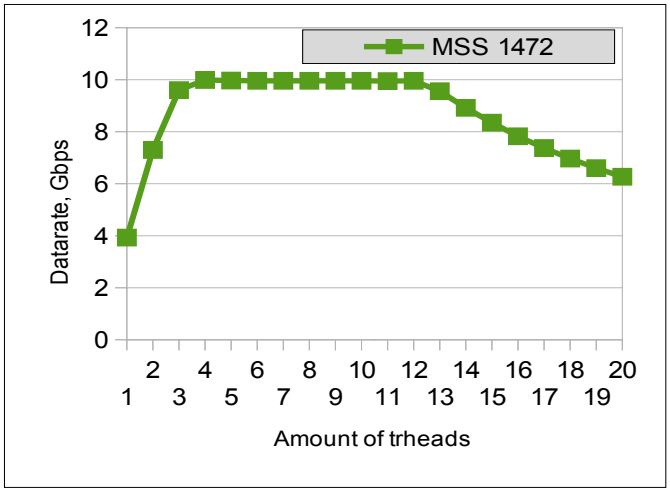


Fig. 2. Dependence of the datarate from the amount of threads, MSS 1472

As can be seen on the graphs, frames with size of 1500 bytes does not meet speed requirements when only one sender is present. When the amount of threads is higher than number of cores (or Hyper-threads) the opposite effect can be observed. The overhead of context-switches decreases performance rapidly. Thus, the data-rate of traffic generating is limited by the amount of CPU threads.

Additional test, run on the CPU with lower frequencies but same amount of CPU threads, gave same behavior, but higher number of senders is needed to achieve 10 Gbps.

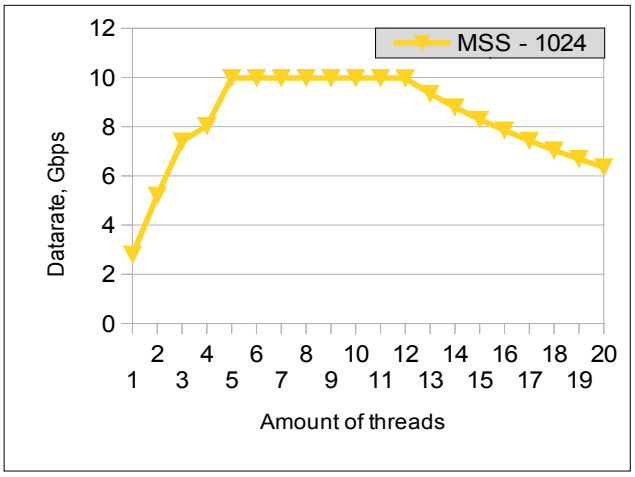


Fig. 3. Datarate dependence from the amount of threads

Next, two traffics with significantly different MSS are tested to check if the same pattern can be observed for other packet sizes, and discover possible dependencies.

Results for 10 Gbps are presented on the Fig. 3 and Fig. 4.

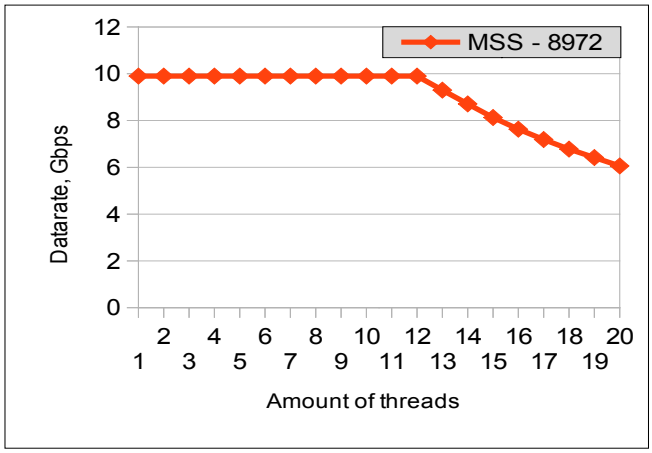


Fig. 4. Datarate dependence from the amount of threads.

It is clearly visible that the bigger MSS makes achieving higher speed easier. Also, whereas at lower number of threads there is a clear difference in speed, when CPU's threads limit is reached the speed drop is approximately the same in both cases. Thus, conclusion is made that using maximal MTU for the channel is beneficial and has no negative influence.

Receivers were also tested and, as they handle traffic much faster than senders, 3 reception threads are sufficient for any MSS or speed, used for experiments held in this work. Thus, this part of system is less critical than sender side as it requires less resources and for tested system does not have large difference in number of threads in the studied cases. It is clearly dependent from speed – higher rates need more threads. In this experiment no automatic receivers' allocation was implemented – testing is performed by manually stating the number of receivers.

Test was made with an application that does not control data rate by manual blocking to compare how system handles multiple senders by itself. This resulted in a different form of traffic as it has more fluctuations than traffic generated by application that implemented manual blocking threads. Other result is higher number of errors caused by trying to access a resource which is taken by another process.

For comparison, simple send threading was also performed on a PC that has processor with lower frequency (number of CPU's threads is still 12). Rate generated by each thread is lower on CPU with lower frequency. A sender on a 2.3 GHz processor has 3.2 Gbps rate (MTU – 1500). A sender on a 2.47 GHz CPU creates 3.9 Gbps traffic.

Auto-spawning sender threads were tested with different speeds, initial amount of time, and time interval between evaluations. The fluctuations of the time lag are around one sender loop time, which corresponds to the mean value of 3.5 µs with standard deviation of 1.4 µs on the tested setup. Thus, whereas the time lag that is considered to be sufficient for spawning new thread is dependent on a particular situation, it is unsafe to take the decision based on a sign of a time lag, as it may lead to changing amount of threads because of a random small deviation. Some limit must be given to prevent unjustified thread spawning. Although not all causes for the fluctuations are clear from this experiment it is clear that higher limit gives less probability to spawn redundant process, but higher probability of not getting requested speed. Higher thread-spawning limit of lag gives wider interval of allowed data rates. The requested speed is only one value from this range.

At last, behavior of the application under utility "stress", which is an utility for imposing load on a system for Linux [13], is checked. Tests showed that new thread is started on the first information analysis (which is performed by main thread) after "stress" uses cores that are already used by senders. Interval between information gathering and analysis used it tests is 0.5 second. At this stage of development for each 0.5 seconds only one thread can be initialized. Thus it takes 1 second to achieve 10 Gbps rate if only one sender was initialized at the start of the application.

VI. Conclusions

The behavior of a single-thread traffic generator is examined in a real network. With MTU of 1500, using single sender, 10 Gbps data rate is not achievable on a tested topology. Using multi-threaded send and receive methods proved to be a working solution as it allows to achieve 10 Gbps speed which is full bandwidth of a tested topology. Theoretically it allows to get any rate, supported by network equipment, though it is limited by the amount of CPU threads. Increasing MSS results in higher speed per thread.

When the amount of threads is higher than number of CPU threads the opposite effect can be observed. The overhead of context-switches decreases performance rapidly. This is observed for all tests with different packet sizes and data rates. Thus, the data-rate of traffic generating is limited and if necessary data rate is not achieved with CPU, fully loaded by application, the conclusion about impossibility to provide requested rate can be made.

[Dmitry Ka1]The automatic thread management basic algorithms proved to be working in a simple environment with constant traffic, although some deviations from theoretical behavior were experienced as, for example, time jitter or accidental rate decrease of a particular thread, while the others have the expected rate.

VII. Future Work

Possible continuation of this work is developing and testing more complex algorithm with advanced thread management and smart statistical data evaluation.

First logical improvement of the existing application is the implementation of the automatic temporary stopping or permanently terminating thread. In this work only cases when rate is not achieved, but not the case of decreasing speed in time which brings necessity to free resources occupied by redundant threads. Also, algorithm of automatic receiver threads management has to be developed.

More tests should be run for different setups. Of special interest are tests with constantly changing traffic. Adaptation to such kinds of traffic is one of the main goals for algorithms described in this paper. Based on the results of such testings, they have to be improved to be able to handle variety of situation correctly.

Next step could be combining functions of pure send and receive with other, often used, operations – for example I/O.

Finally, if all functionality will be proved to work correctly, it can be tested as a part of an UDP-based

transport protocol for high speed data transmission such as, for example, RMDT.

REFERENCES

[1] "RFC 894 - A Standard for the Transmission of IP Datagrams over Ethernet Networks." [Online]. Available: https://tools.ietf.org/html/rfc894. [Accessed: 04-Mar-2016].

[2] E. He, J. Leigh, O. Yu, and T. A. DeFanti, "Reliable blast UDP: Predictable high performance bulk data transfer," in *Cluster Computing, 2002. Proceedings. 2002 IEEE International Conference on*, 2002, pp. 317–324.

[3] Y. Gu and R. Grossman, "Udtv4: Improvements in performance and usability," in *Networks for Grid Applications*, Springer, 2008, pp. 9–23.

[4] S. Höhlig, "Optimierter Dateitransfer über 100 Gigabit/s," in *100-Gigabit/s-Workshop in Mannheim, Mannheim*, Germany, Sep-2011.

[5] "Big Data Transmission | F I L A." [Online]. Available: https://fila-lab.de/index.php/our-work/big-data-transmission/. [Accessed: 09-Mar-2016].

[6] D. Kachan, E. Siemens, *Comparison of Contemporary Protocols for High-speed Data Transport via 10 Gbps WAN Connections.* Proceedings of 2nd International Conference on Applied Innovations in IT. Köthen, pp. 21-27, 2014 (DOI: 10.13142/kt10002.04);

[7] D. Kachan, E. Siemens, V. Shuvalov, *Available bandwidth measurement for 10 Gbps networks.* Proceedings in 2015 International Siberian Conference on Control and Communications (SIBCON), 2015, pp. 1–10.

[8] *Linux Programmer's Manual,* Linux, p. "socket".

[9] S. Srivastava, S. Anmulwar, A. M. Sapkal, T. Batra, A. Gupta, and V. Kumar, "Evaluation of traffic generators over a 40Gbps link," in *Computer Aided System Engineering (APCASE), 2014 Asia-Pacific Conference on*, 2014, pp. 43–47.

[10] V. Vishwanath, T. Shimizu, M. Takizawa, K. Obana, and J. Leigh, "Towards terabit/s systems: Performance evaluation of multi-rail systems," in *High-Speed Networks Workshop, 2007*, 2007, pp. 51–55.

[11] I. Fedotova, E. Siemens, and H. Hu, "A high-precision time handling library," *J. Commun. Comput.*, vol. 10, pp. 1076–1086, 2013.

[12] "Apposite Technologies :: Linktropy and Netropy Comparison." [Online]. Available: http://www.apposite-tech.com/products/index.html. [Accessed: 04-Mar-2016].

[13] *Linux Programmer's Manual,* Linux, p. "stress".

Connection Establishment Algorithm for Multi-destination Protocol

Sergii Maksymov, Dmitry Kachan, Eduard Siemens
Department of Electrical, Mechanical and Industrial Engineering
Anhalt University of Applied Sciences
Bernburger Str. 55, 06366 Köthen, Germany
E-mail: {s.maksymov, d.kachan, e.siemens}@emw.hs-anhalt.de

***Abstract*—connection establishment is a fundamental function for any connection-oriented network protocol and the efficiency of this function defines the flexibility and responsiveness of the protocol. This process initializes data transmission and performs transmission parameters negotiation, what makes it mandatory process and integral part of entire transmission. Thus, the duration of the connection establishment will affect the transmission process duration. This paper describes an implementation of a handshake algorithm, designed for connection with multiple peers, that is used in Reliable Multi-Destination Transport (RMDT) protocol, its optimization and testing.**

***Keywords:* multi-destination; handshake; connection establishment; network protocol.**

I. INTRODUCTION

[Sergii Ma1]Reliable Multi-Destination Transport (RMDT) is a protocol designed to deliver data from one source (sender) to multiple destinations simultaneously, accurately regardless network impairments.

It is implemented as a C++ library and aimed to effectively transmit the same data to multiple recipients. During transmission the only one sending instance is created which initializes a common buffer for all recipients, thus it utilizes less system resources and performs minimum copy operations during the transmission process. Big amount of data can be transmitted to many recipients with less load on a system. The protocol is aimed to send data within whole available bandwidth, what is especially important for Long Fat Pipes – links with high bandwidth and latency. Such links are unsuitable for legacy protocols, primarily Transmission Control Protocol (TCP), which is not able to utilize the full bandwidth within high latency links [1].

To provide its benefits RMDT requires a performance hardware that supports multi-threading and has enough memory to allocate big buffers (up to 1000 MB). Sending and receiving operations are split into 2 threads. One thread is responsible only for reception and another one – only for sending. Sender side application has additionally Event thread, which is responsible for data preparation within buffers and processing of acknowledgments from recipients.

II. BACKGROUND AND RELATED WORK

Multi-destination data delivery is closely related to multicast, but implies a uni-direction transmission. The RMDT protocol is based upon UDP and can be considered, from the network point of view, as the set of multiple unicast streams which transport data in one direction from sender to recipients. Due to this fact, the experience of the legacy transport protocols can be used to design the connection establishment function of RMDT.

The well-know TCP utilizes three-way handshake algorithm [2] to establish a connection, see Figure 1. The server should be switched into LISTEN state, this action is called Passive Open. In the LISTEN state server is able to accept a connection request, Synchronization (SYN) packet.

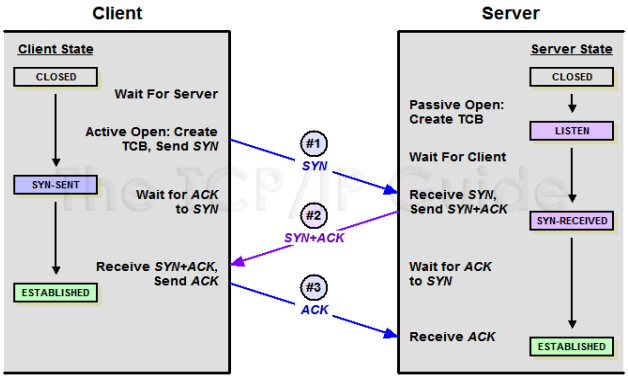


Fig. 1. Three-way handshake process for TCP connection establishment [2]

The connection request must be acknowledged by the server along with negotiation of transport parameters (SYN+ACK packet), primary Initial Sequence Number (ISN) negotiation. The Sequence Number is a number representing a sequence number of the first byte of data in a segment. At the moment of connection establishment, this number is chosen from a special counter [2], that ticks every 4 μs, to eliminate conflicts between different TCP connections. The most basic reason for this is to detect duplicate SYNs and to distinguish the SYN packet belongs to the same connection

or it is a part of a new connection, in the case of lost message during 3-way handshake, for example. The ISN is also used against IP-spoofing technique, but with enhanced degree of randomization [4]:

$$ISN = C(t) + hash(L_{addr}, L_{port}, R_{addr}, R_{port}, key) \quad (1)$$

Where, $C(t)$ – value of the counter, L_{addr}, L_{port} – local address and port, R_{addr}, R_{port} – remote address and port, key – a random value chosen by the host on startup. Any hash function can be used, but as written in [4], MD5 hash function is recommended, as it is well supported by different hardware and has a lot of implementations. Such ISN generation complicates IP-spoofing attacks.

III. 2-WAY HANDSHAKE WITH MULTIPLE PEERS

In the case of point-to-multipoint data delivery protocol all the destination points should be notified and connected. This fact introduces some difficulties, because an error triggered by one peer might fail or, at least hinder the common connection establishment process. The increase of amount of receivers leads to the probability of such a failure increase. That is the major distinction from point-to-point connection establishment. In RMDT the 2-way handshake is performed with each peer. Sender initiate a connection by sending Handshake Request (HS Request) packets to each of the receivers and then waits for their responses. Receiver, in turn, waits for this HS Request from the beginning. Request contains transmission session parameters, such as Initial Sequence Number, Receiver ID, Maximal Segment Size and protocol Version Number, see Figure 2. This

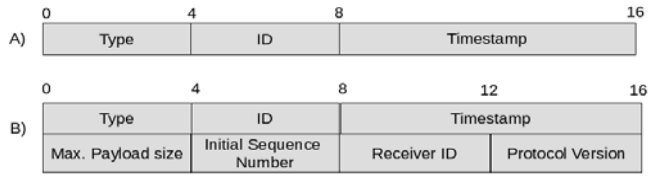


Fig. 2. Packet headers - A) RMDT control packet header; B) Handshake Request packet header.

parameters are checked and set at the receiving application. Then it must send Handshake Response (HS Response) which is expected by the sender for some amount of time. The sender gathers HS Responses during specified time interval from the recipients and then repeats the send operation to those of them which did not responded. Connection is established when all the recipients have responded. It might happen, so the handshakes cannot be exchanged with one or more recipients and the connection would never be established, so the process of handshake will freeze. For such a situation there is a timeout for connection operation, which can be set by the application. In the case of timeout, connection will be also established, if at least one recipient have responded, but an application will be warned about the fact of timeout. This mechanism relieves an application from possible freezing of the connection establishment process.

Because RMDT library runs in 3 threads, there is an Inter-Thread Communication mechanism via queues and notification method of *conition_variable* object provided by the Standard C++ Library and which is used by the handshake process implementation. The implementation is based on two timeouts: timeout for overall process duration and timeout for responses expectation. The first one guarantees that the connection will last no longer then the specified time interval regardless the result of the handshake process, thus protecting against unnecessary hanging inside this process. The second one lets the process to use system resources more effectively and not to flood the network with frequent HS Requests. This timeout has one feature: the process can be waked up before this timeout, when the HS Response comes. This feature allows to save the time on connection establishment in the case, when all the expected responses are come before the timeout. Presence of this feature distinguish two versions of the handshake process implementation – asynchronous and synchronous.

The synchronous handshake is performed in the same way, but its duration is aligned to the specified time period for HS Responses expectation. Owing to the fact that operation status check is performed after timeout at each iteration of this algorithm. The flowchart of the algorithm is shown in the Figure 3.

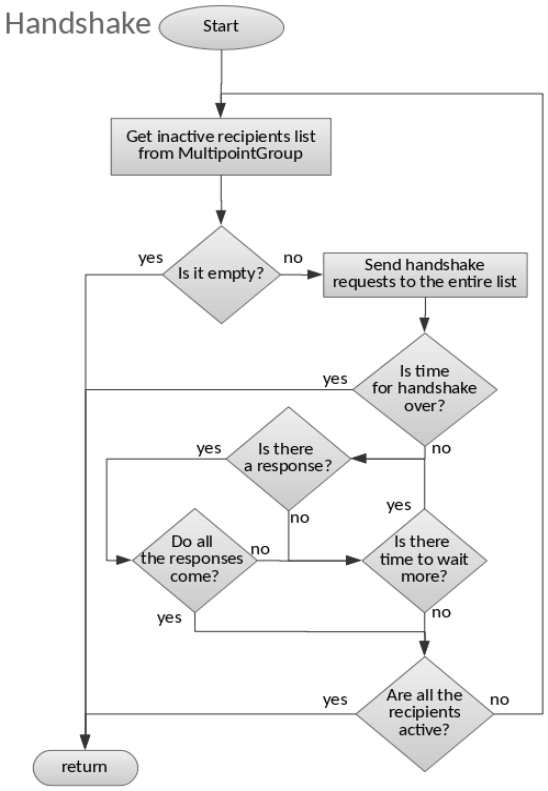


Fig. 3. Flowchart of the asynchronous handshake process.

The asynchronous handshake operation is more responsive, but has a small overhead in form of inter-thread communication and thread synchronization comparing to synchronous one. But it is insignificant for performance systems that the library is designed to.

IV. TEST-BED AND DESCRIPTION OF EXPERIMENTS

To test the protocol a 10 Gbps network, shown in Figure 4, is used. This network is located in the laboratory Future Internet Lab Anhalt (FILA) [5] and is used for experiments and protocol testing. With the help of this network the two implementations of handshake process were tested, namely

measuring the duration of the connection establishment in different conditions of both synchronous and asynchronous implementations. The network interconnects 4 multicore servers with 10 Gbps interfaces, 2 Extreme Networks Summit x650 10 Gbps capable switches and 2 network emulators Apposite 10G. Detailed configuration of the servers is shown in the Table I.

TABLE I
CONFIGURATION OF THE SERVERS

Server name	Linux kernel	CPU	RAM
Germany	4.2.0-23-generic x86_64	2x Intel Xeon X5690 (6-core) 3.5 GHz	40 GB DDR3 1066 MHz
Brazil	3.13.0-37-lowlatency x86_64	2x AMD Opteron 4238s (6-core) 3.3 GHz	32 GB DDR3 1333 MHz
Argentina	3.13.0-35-generic x86_64	2x AMD Opteron 4238s (6-core) 3.3 GHz	32 GB DDR3 1333 MHz
Kazakhstan	3.13.0-45-generic x86_64	2x Intel Xeon E5-2630 (6-core) 2.3 GHz	64 GB DDR3 1333 MHz

The Apposite 10G network emulator allows to introduce in the network different impairments, such as packet delay and packet loss with high accuracy up to nanoseconds [6]. Moreover, it displays and controls link capacity passing through the emulator.

Extreme Network Summit X-650 10Gbps switches perform layer 3 packet switching. All the connections are implemented with optical fiber links and have capacity of 10Gbps.

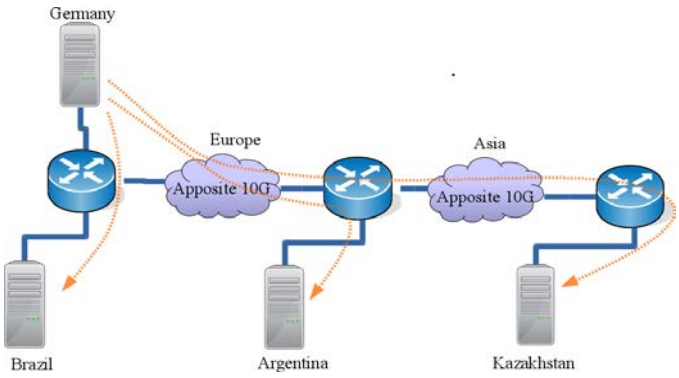


Fig. 4. Network topology

Experiment scenario is to initialize data transmission from source (Germany) to 3 destinations (Brazil, Argentina, Kazakhstan) using the RMDT test application and measure the duration of the handshake process in the network with low impairments: 50 ms of Round-Trip-Time (RTT) and no packet losses. The second scenario introduces more impairments into the links: up to 250 ms of RTT and up to 0.7% of packet losses. These parameters are even more worse than the real links between Germany and USA, for example. In both scenarios the synchronous and asynchronous implementations are compared.

V. EXPERIMENTAL RESULTS

The experiments are divided into 4 groups by allocated send buffer size – 10 MB, 100 MB, 500 MB and 1000 MB, because the buffer allocation time depends on its size. For each of the buffer sizes, 10 measurement iterations were done for both implementations, so 20 in total. Multiple measurements are done to get more precise results, because the system activity affects experiment results.

The results have very slight deviation, so it can be hardly seen in the Figure 5.

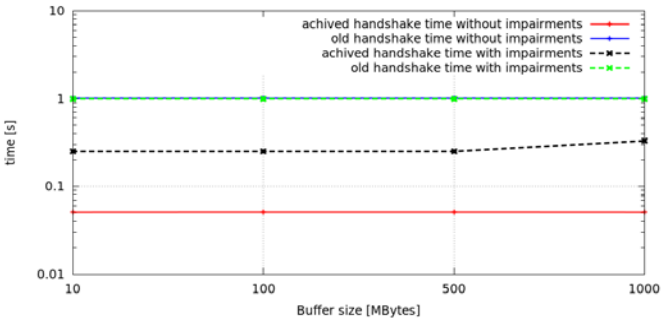


Fig. 5. Experiment results (with std. deviation).

The asynchronous handshake takes almost 20 times less time then the synchronous one within good link (50 ms RTT) and about 4 time faster within link with more impairments (250 ms RTT and 0.7% of packet losses). Impact of packet delay can be easily seen on the asynchronous handshake plot (the black line), 200 ms increase is clear, but it is invisible in synchronous handshake plot. It is pretty clear, that the timeout in synchronous handshake implementation is major factor that defines the duration of the handshake.

TABLE II
EXPERIMENT RESULTS WITHIN GOOD LINK (AVERAGE)

Buffer size, MB	Asynchronous implementation, s	Std. deviation	Synchronous implementation, s	Std. deviation
10	0,050509	0,000031	1,000868	0,000181
100	0,050504	0,000025	1,000756	0,000087
500	0,050497	0,000031	1,000781	0,000059
1000	0,050512	0,000028	1,000806	0,000019

The numbers in the Table II and Table III contains the mean handshake duration obtained from experiments. As it was mentioned above, the duration of the synchronous handshake is aligned to the timeout for HS Responses expectation, which is 1 s in our experiments. There can be smaller value specified for this expectation, in other words sleep of the thread, but in this case the network link will be overloaded by frequent HS Requests send, what is redundant and dangerous especially for links with high latency.

TABLE III
EXPERIMENT RESULTS WITHIN BAD LINK (AVERAGE)

Buffer size, MB	Asynchronous implementation, s	Std. deviation	Synchronous implementation, s	Std. deviation
10	0,250000	0,000033	1,000000	0,000011
100	0,250406	0,000011	1,000512	0,000008
500	0,250415	0,000011	1,000513	0,000006
1000	0,330370	0,252299	1,000515	0,000009

VI. CONCLUSION

There is a need of reliable and fast point-to-multipoint data delivery, especially on the side of huge content distributors. And there is also a lack of new ideas regarding such kind of data transmission. Thus, design of multi-destination protocol, which is aimed to correspond contemporary data delivery requirements and to be able to effectively utilize the available hardware resources is in demand. Connection establishment process of such protocol requires attention as it is the fundamental function of the data transmission.

Designing the handshake operation, two approaches were developed and compared, synchronous and asynchronous one. The asynchronous handshake implementation demonstrated much effective work, especially within the links with good parameters, which are widely used over the world, for example link from [Sergii Ma2]Berlin to Moscow or from Madrid to Tokyo . This approach has the only drawback that it has some overhead on resources utilization due to inter-thread communication, but it is insignificant for contemporary systems which are required for RMDT.

VII. FURTHER WORK

The further work on the protocol will be focused on implementation of the rest necessary features, namely congestion control based on the Available Bandwidth Control (ABC) [7] and finalize the session management to isolate the recipients with different throughput. And much further work will be aimed on inspection of the security issues, because for now the security of the protocol completely relies on the security aspects of underlying UDP. Besides, there is a research can be performed on the security issues of the connection establishment, for which the experience of the TCP with IP-spoofing can be used to prevent similar attacks.

REFERENCES

[1] D. Kachan, E. Siemens, and V. Shuvalov, "Comparison of Contemporary Solutions for High Speed Data Transport on WAN 10 Gbit/s Connections," *J. Commun. Comput.*, vol. 10, no. 6, pp. 783–795, 2013.

[2] "The TCP/IP Guide - TCP Connection Establishment Process: The 'Three-Way Handshake.'" [Online]. Available: http://www.tcpipguide.com/free/t_TCPConnectionEstablishmentProcessTheThreeWayHandsh-3.htm. [Accessed: 01-Mar-2016].

[3] R. Braden, "Requirements for Internet Hosts - Communication Layers." [Online]. Available: http://tools.ietf.org/html/rfc1122.html#page-87. [Accessed: 07-Mar-2016].

[4] S. Bellovin, "Defending Against Sequence Number Attacks." [Online]. Available: http://tools.ietf.org/html/rfc1948.html. [Accessed: 07-Mar-2016].

[5] "F I L A.", [Online]. Available: http://fila-lab.de [Accessed: 07-Mar-2016].

[6] "Apposite Technologies :: Netropy 10G2 Network Emulator." [Online]. Available: http://www.apposite-tech.com/products/netropy-10G2.html. [Accessed: 07-Mar-2016].

[7] D. Kachan, E. Siemens, and V. Shuvalov, "Available bandwidth measurement for 10 Gbps networks," in *2015 International Siberian Conference on Control and Communications (SIBCON)*, 2015, pp. 1–10.

Modeling of Self-similar Traffic

Irina Strelkovskaya, Irina Solovskaya, Nikolay Severin
Education and Research Institute for Infocommunication and software engineering
Odessa National A.S. Popov Academy of Telecommunications
65029, Kuznechnaya Str., Odessa, Ukraine
E-mail: dekanat.ik@onat.edu.ua, i.solovskaya@onat.edu.ua, n_severin@ukr.net

***Abstract*—Modeling of self-similar traffic is performed for the queuing system of G/M/1/K type using Weibull distribution. To study the self-similar traffic the simulation model is developed by using SIMULINK software package in MATLAB environment. Approximation of self-similar traffic on the basis of spline functions. Modeling self-similar traffic is carried out for QS of W/M/1/K type using the Weibull distribution. Initial data are: the value of Hurst parameter *H*=0,65, the shape parameter of the distribution curve $\alpha \approx 0,7$ and distribution parameter $\beta \approx 0,0099$. Considering that the self-similar traffic is characterized by the presence of "splashes" and long-term dependence between the moments of requests arrival in this study under given initial data it is reasonable to use linear interpolation splines.**

***Keywords*: self-similar traffic, Hurst parameter, Weibull distribution, modeling, queuing system, approximation, spline function.**

I. Introduction

Modern telecommunications networks are developing now in the direction of the next generation networks NGN (Next Generation Network), that are based on the application of package technologies to transmit different types of traffic on a single network infrastructure providing quality characteristics QoS (Quality of Service) [1][2].

By its nature, the traffic, serviced in NGN network, is heterogeneous, as it is formed by many but different in their characteristics sources of services and network applications, ensuring the provision of services for voice, data and video TPS (Triple Play Service) [1][2].

It is known [3] that the packet traffic has train character and has the property of self-similarity, the cause of which is a long-term dependence between the moments of package arrival determined by the correlation function at different points in time.

The flow of self-similar traffic is also characterized by the presence of aftereffect, i.e. if the number of requests received by the queuing system after the moment t depends on the number of requests received before the time t [3].

To describe self-similar traffic, considering that the moments of packages arrival have the distribution with "heavy-tailed", often use the distribution of Pareto, lognormal or Weibull [3].

The degree of self-similar traffic is estimated by Hurst parameter. Besides, the self-similarity traffic has the structure being saved in multiple scaling. But real traffic, as a rule, has the number of more “splashes”. It dramatically worsens the value of quality characteristics QoS (values of loss probability, package delay time and jitter) [3].

Taking into account the above said, modeling of the traffic will allow to avoid network overload, exceedance of the standard values of delay time and jitter, considering peakedness of package traffic.

Today, the problem of characteristics evaluation of self-similar traffic characteristics is devoted a considerable amount of works of different authors [5]-[7].

Most of the works are based on experimental data and obtained results of simulation modeling using the R/S analysis, Whittle evaluation, wavelet analysis, and many other methods [5][6].

One of the instruments for investigating the characteristics of self-similar traffic is a simulation modeling that allows obtaining the necessary characteristics. And the choice of the form of approximation of received results is a hot topic of virtually any research [5]-[7].

The application of linear spline to approximate the self-similar traffic is offered in this work.

The aim of this work is to study the characteristics of self-similar traffic using the worked out simulation model in SIMULINK package of MATLAB environment with the following approximation by linear splines.

II. Modeling Of Self-similar Traffic Using The Simulink Package In Matlab Environment

Let we perform modeling of self-similar traffic for queuing system (QS) of G/M/1/K type that serves the requests stream, which intervals are described by arbitrary distribution G, the time of service has exponential distribution M, QS has 1 line and length of requests queue is K [3-4].

In this case, for modeling the process of service requests arrival we use QS of W/M/1/K type, where

W − is a stream of requests with Weibull distribution,

M − is time of requests service distributed according to an exponential law, QS − has 1 line and length of requests queue is K [3-4].

Let we consider the Weibull distribution, given by the differential distribution function [3-4]:

$$f(x)=\begin{cases}\alpha\beta x^{\alpha-1}e^{-\beta x^{\alpha}}, & x\geq 0,\\ 0, & x\leq 0\end{cases} \qquad (1)$$

where α – is a parameter of distribution curve form ($0<\alpha<1$); $\alpha = 2-2H$,

H – Hurst parameter, ($0{,}5 \leq H \leq 1$),

$\beta=\left[\lambda\Gamma\left(1+\frac{1}{\alpha}\right)\right]^{-1}$ – distribution parameter, $\beta>0$,

Γ – gamma function,

λ – intensity of requests arrivals for QS servicing.

The integral function of Weibull distribution has the following form [3]:

ack/s,

μ – request servicing durability, μ= 125 s,

K – the length of requests queue, K=100 requests.

For modeling of self-similar traffic let we set the value of Hurst parameter H=0,65.

Then for distribution of Weibull parameters α and β are equal respectively $\alpha \approx 0{,}7$ and $\beta \approx 0{,}0099$.

$$F(x)=1-e^{-\beta x^{\alpha}}. \qquad (2)$$

Generation of a random value of time interval between requests arrivals in self-similar stream formation in the simulation model is performed by the transition from uniform distribution by the inverse function method according to the expression [3]:

$$x=\beta\left[-Ln(1-R)\right]^{\frac{1}{\alpha}}, \qquad (3)$$

where $R \in [0;1]$ – is uniform distributed random number.

As the initial data of QS W/M/1/K operation, we use its following characteristics:

λ – intensity of requests arrivals for servicing in QS, λ =100

The received results of modeling self-similar traffic for QS of W/M/1/K type for the given initial data using SIMULINK package of MATLAB environment is shown in Fig. 1, where

N – the number of requests,

t – requests arrival time.

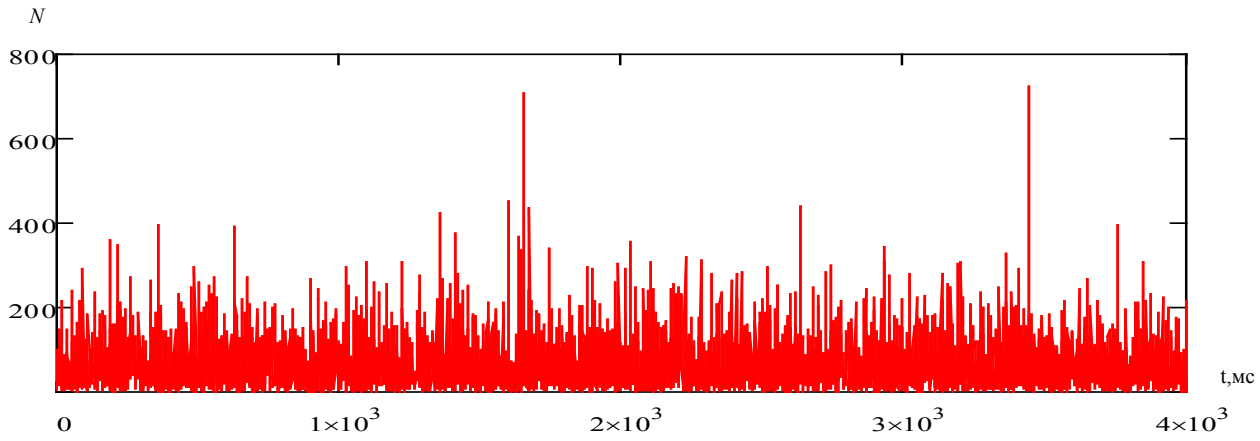


Fig.1. Simulation of self-similar traffic for QS of W/M/1/K type.

According to the received modeling results shown in Fig. 1, we can conclude the following.

The resulting graph shows that the process is no uniform and corresponds to the above described characteristics of self-similarity.

There is scale invariance, the presence of "splashes" of requests and long-term dependence between the moments of request arrivals.

III. Approximation of Self-Similar Traffic by Spline Functions

Simulation of self-similar traffic performed for the QS of W/M/1/K type.

To approximate self-similar traffic we use spline functions, which allow effectively solving the problems of processing various statistics data and experimental dependences having rather complex structure [5].

In this work, we consider the use of linear interpolation splines [8-10].

Let we consider using linear interpolation spline to approximate the results of traffic modeling obtained for the QS of W/M/1/K type, using the SIMULINK package of MATLAB environment, which are shown in Fig. 1.

To approximate by the linear spline the self-similar traffic, shown in Fig. 1, let we perform the selection of values, for example, on the time interval $[3813;3990]$, which is a series of long-term dependency. Fig. 2 shows a selection for a predetermined of time interval.

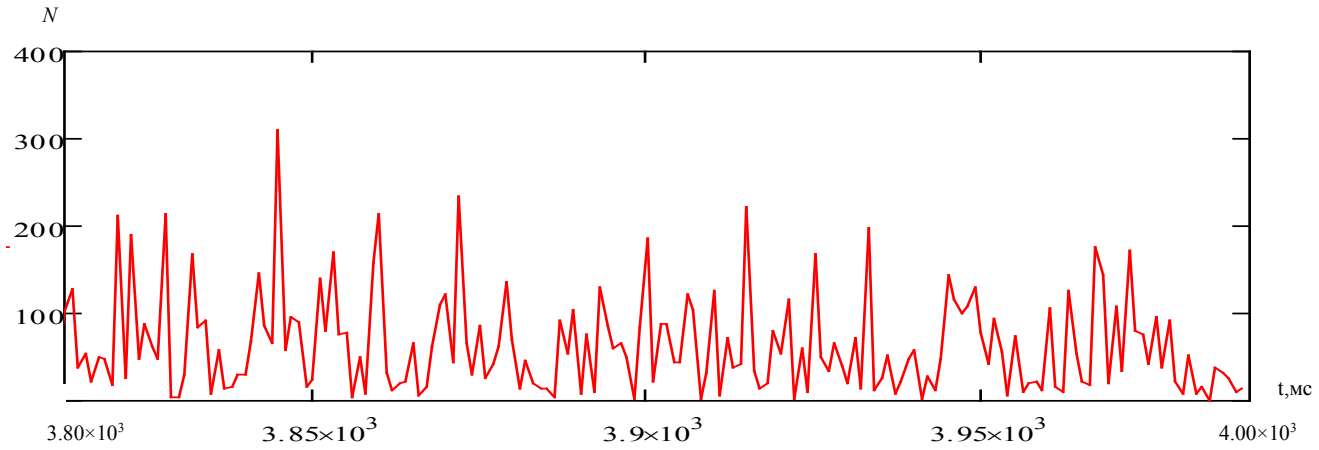


Fig.2. Selection of values for self similar traffic for the interval of values $[3813;3990]$.

Let on the interval $[0;T]$ the values results of traffic modeling are set.

Let we divide this interval $[0;T]$ by points Δ:

$$0 = t_0 < t_1 < \ldots < t_N = T$$

on the interval $[t_i; t_{i+1}]$, $i = \overline{0, N-1}$ on each let we build polynomial of certain degree.

Such polynomial we will use linear spline [7-10].

According to [8-10], interpolation linear spline $S_1(t_i)$ on the interval $[t_i; t_{i+1}]$, $i = \overline{0, N-1}$ – is the continuous piecewise linear function.

Let in the nodes of Δ grid be set values

$$s_i = s(t_i), \quad i = \overline{0, N}$$

some functions $s(t)$, defined on the interval $[t_i; t_{i+1}]$.

Interpolation spline $S_1(t_i)$ is defined by the conditions [8-10]:

$$S_1(t_i) = S_i, \; i = \overline{0, N}. \qquad (4)$$

If we define $h_i = t_{i+1} - t_i$, so with $t \in [t_i, t_{i+1}]$ the equation of linear spline will be [6]:

$$S_1(t) = f_i \frac{t_{i+1} - t}{h_i} + f_{i+1} \frac{t - t_1}{h_i}, \qquad (5)$$

or

$$S_1(t) = f_i + \frac{t - t_1}{h_i}(f_{i+1} - f_i). \qquad (6)$$

Let we consider the self-similar traffic on the selected interval $[3813;3990]$ by setting a uniform grid of decomposition with intervals h=0,01.

For each point of decomposition the values of requests numbers are known in each interpolation node.

Using linear interpolation spline $S_1(t_i)$ and expressions (4-6), we get an approximation of the traffic, shown in Fig. 3.

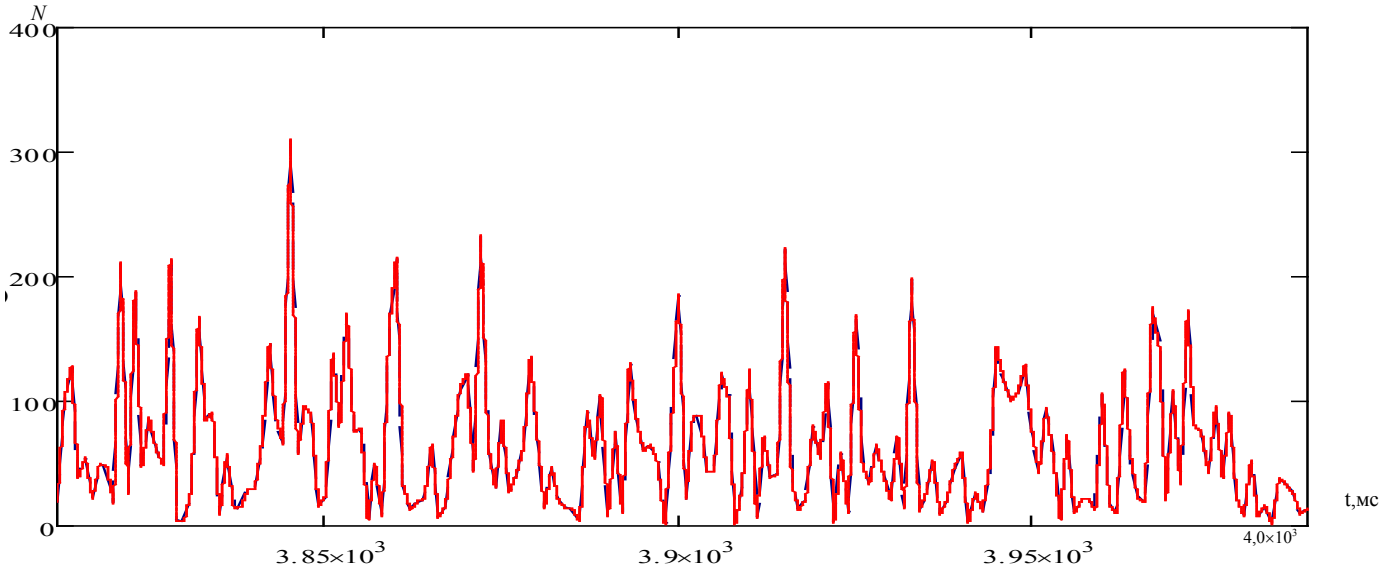


Fig.3. Approximation of self similar traffic by linear spline

The received results suggest the possibility of using linear interpolation splines for approximation of self-similar traffic obtained by the worked out simulation model by using SIMULINK package in MATLAB environment.

IV. Conclusions

1. Modeling of self-similar traffic for QS of W/M/1/K type:

– incoming stream of requests distributed under the law of Weibull,

– distribution of service time – exponential,

– single-line system,

– the length of the queue of requests is K is performed.

2. With the help of SIMULINK software package in MATLAB environment the simulation model is developed to study the characteristics of self-similar traffic.

3. Modeling self-similar traffic is carried out for QS of W/M/1/K type using the Weibull distribution. Initial data are:

– the value of Hurst parameter H=0,65,

– the shape parameter of the distribution curve $\alpha \approx 0{,}7$ and distribution parameter $\beta \approx 0{,}0099$.

4. Approximation of self-similar traffic by linear splines is considered.

5. Considering that the self-similar traffic is characterized by the presence of "splashes" and long-term dependence between the moments of requests arrival in this study under given initial data it is reasonable to use linear interpolation splines.

References

[1] P.P. Vorobienko, L.A. Nikityuk, P.I. Reznichenko *Telecommunication and Information Networks.* Kiev, SMMIT-KNIGA: 2010, 640 p.

[2] A.V. Roslyakov, S.V. Vanyashin, M.Yu. Samsonov and oth. *Networks of the next generation NGN.* Moscow. Eko-Trends, 2008, 424 p.

[3] V.V. Krylov, S.S. Samohvalova, *Theory of telegraphic and its applications.* St. Petersburg, BXV-Petersburg, 2005, 288 p.

[4] Kleynrok L. *Theory of queuing network*. Moscow. Engineering, 1979, 432 p.

[5] O.I. Sheluhin, A.M. Tenyakshev, A.V. Osin, *Fractal processes in telecommunications.* Moscow, Radioengineering, 2003, 480 p.

[6] O.I. Sheluhin, A.V. Osin, S.M. Smolskii, *Self-similarity and fractals. Telecommunication applications.* Moscow, PHISMATLIT, 2008, 368 p.

[7] I.V. Strelkovskaya, V.V. Popovskii, D.Y. Buhan, Comparative methods of approximation in the results of recursive evaluation of the state of network elements and their modes. Telecommunication systems and technologies. *Applied radio engineering. State and perspectives of development: 3-d International. Radio engineering forum:* coll. of scient. works, P. 2, pp. 15–16, 2008.

[8] Yu.S. Zavyalov, B.I. Kvasov, V.L. Miroshnichenko, *Methods of spline functions.* Moscow, Science, 1980, 352 p.

[9] V.V. Popovskiy, S.O. Saburova, V.F. Oliynik and oth. *Mathematical bases of telecommunications systems theory*. Kharkiv, TOV «Company SMIT", 2006, 564 p.

[10] I.V. Strelkovskaya, I. Solovskaya, N. Severin, S. Paskalenko, Approximation of self-similar traffic by spline-functions. *Modern Problems of Radio Engineering, Telecommunications and Computer Science: proceedings of the XIII International Conference (TSET'2016),* Slavske, Ukraine, February 23 – 26, pp. 132-135, 2016.

Similarity Measurement of Biological Signals Using Dynamic Time Warping Algorithm

Ivan Luzianin[1], Bernd Krause[2]
[1,2]Anhalt University of Applied Sciences – Computer Science and Languages Department
Lohmannstr. 23, 06366, Koethen, Germany
[1]Perm National Research Polytechnic University - Electrotechnical Department
Komsomolsky Ave. 29, 614990, Perm, Russia
E-mail: [1]lis@msa.pstu.ru, [2]bernd.krause@inf.hs-anhalt.de

***Abstract*—The problem of similarity measurement of biological signals is considered on this article. The dynamic time warping algorithm is used as a possible solution. A short overview of this algorithm and its modifications are given. Testing procedure for different modifications of DTW, which are based on artificial test signals, are presented.**

***Keywords:* biological signal, dynamic time warping, ECG, artificial signals, testing methods.**

I. Introduction

Modern methods of functional diagnostics of human body provide wide range of opportunities for leading and recording large amount of biological signals. These signals are indirect indices of human body changes. Sometimes, estimation of changes in these signals are only way of making diagnosis. Thereby, the task of fast and correct interpretation of biological signals becomes actual. For a long time, analysis of biological signals was carried out with the help of human experts. The correctness of results interpretation in this approach depends on experience and qualification of an expert. Expert also can make a mistake during analysis, which may cause wrong interpretation of analysis results.

Nowadays, using modern mathematical methods and high performed computers, statistical data about changes of biological signals and their aftermaths for human body based on multiple measures were obtained [1][2]. This data help experts with interpreting of biological signals but the task of precise analysis of biological signals is still actual.

One approach to the analysis of biological signals is a similarity measurement between tested signal and some reference signal, which parameters are known [3]. For solving this task dynamic time warping algorithm (DTW) are widely used. This method works well in many different fields, e.g. in tasks of speech recognition, and analysis of complex time series. [4]-[10]. However, algorithm has high sensitivity to input signals and their changes. Therefore, there are many different modifications of DTW, allowing getting additional advantages when analyzing real data. Besides that, if input signals are very complex, the task of interpreting of algorithm results becomes complicated.

Due to problems described above, it is necessary to test this method and its modifications using simple signals with known parameters. Experiments with different signals will give information about sensitivity of algorithm to concrete changes in concrete types of signals.

The main goal of the research is to find an optimal modification of DTW algorithm, which is the most appropriate for analyzing specific changes in biological signals. Testing procedure of DTW algorithm based on artificial testing signals is described in this article. The questions of extracting and interpreting information from output parameters of the algorithm are considered.

In the second part of the article a comparison analysis of classical DTW algorithm and one of its modifications are presented.

II. Features of Biological Signals

Most biological signals are quasiperiodic. This term has strict mathematical definition for deterministic systems. However, for biological signals this means that they can change their period in time due to stretching, shrinking and shifting of single patterns relative to each other's. Besides that, biological signals have variable amplitude, which can changes from one pattern to another within certain limits.

One of the examples of biological signals is electrocardiographic (ECG) signal. In general, ECG signal consists of sequence of three-dimensional cycles of electric vector of the heart (heart vector). Analysis of these cycles are complicated task. To simplify it ECG signal is divided into three projections on coordinate axes. Fragment of X-projection of ECG signal is shown on Fig. 1.

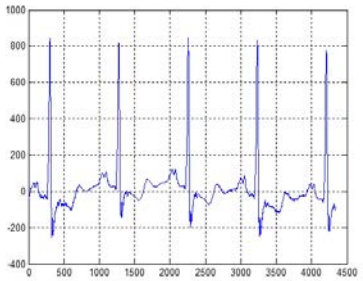


Fig. 1. Fragment of X-projection of ECG signal.

This signal is a continuous sequence of heartbeat cycles. For similarity analysis, separation of these beats is needed. This separation might be done, for instance, by standard algorithms, which are based on determining fiducial points of every heartbeat [1],[11]. Fig. 2 shows a single beat, separated using the described technology.

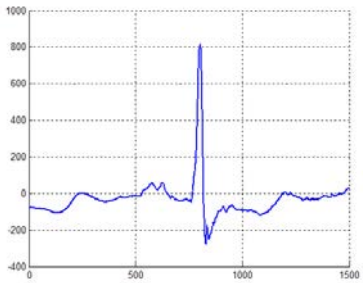


Fig. 2. Single heartbeat extracted from the ECG signal.

In practice biological signals are often divided into different segments and complexes. Regularities described above are typical for both a whole signal and its segments. As an example a part of a special segment of single heartbeat is considered (Fig. 3). This signal is obtained with the help of high pass filtering because the most interesting changes of this segment waveform appears in high-frequency component and these changes are invisible in standard ECG.

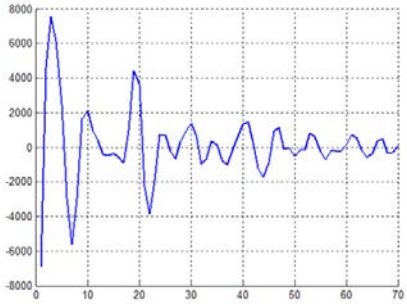


Fig. 3. Part of ST-segment of single heartbeat.

III. Test Signals Development

As described above, before analyzing real biological signals it is needed to find the most appropriate algorithm for this task. To do this, it is necessary to understand relationship between output parameters of algorithm and features of input signals. For solving this task special testing procedure based on artificial test signals are developed.

Test signals must have the same features as real signals. In ideal case, test signals must precisely repeat behavior of real signals and have known parameters available for changing. One of the examples of such test signals for simulating ST-segment shown above are signals based on sequences of impulses filtered with second-order band pass Butterworth filter (Fig. 4).

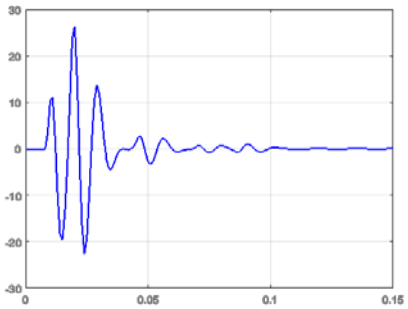


Fig. 4. Example of test signal based on sequence of impulses.

Such test signals are described in [12].

This approach allows precisely simulate different complexes of single heartbeats but the correct interpretation of testing results becomes very difficult due to complexity of a signal.

In this case, at the first step it is necessary to use signals with simple waveform (e.g., segments of sine wave and their combinations). Such signals can have the following changeable parameters:

- width of segment
- amplitude of segment
- initial displacement
- length of final interval.

Examples of such testing signals are described in [13].

In this study single segments of sine function are used as test signals.

DTW algorithm needs two signals: reference and test. Signal shown on Fig. 5 are used as a reference signal. This signal is not changeable during testing.

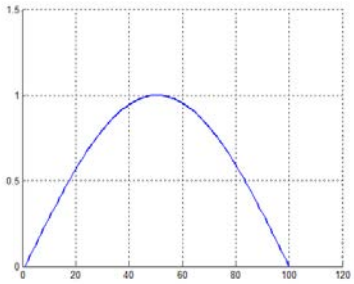


Fig. 5. Reference signal.

As a testing signal arbitrarily changeable signal is used (Fig. 6).

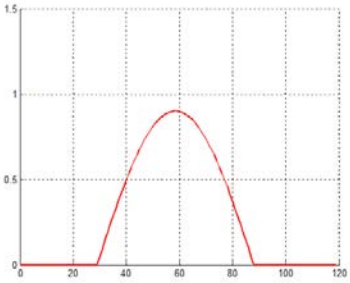


Fig. 6. Testing signal.

All test signals are a combination of the initial displacement, half-wave segment of sine function with variable amplitude and width and the end part of zero-values.

Changing parameters of test signal is made with respect of reference signal parameters.

Both test and reference signals are shown on Fig. 7 (reference signal is blue; test signal is red).

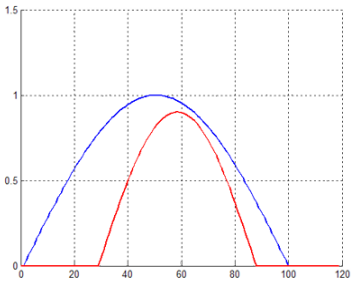


Fig. 7. Both reference and test signals.

IV. Classical DTW Algorithm

General description of DTW is given in [3]. As input parameters, two signals: reference and test are used. Both signals are sampled at equidistant moments t_i, e.g. $\Delta t = t_{i+1} - t_i$. The aim of algorithm is a warping of time axis using stretching and shrinking in such way that the test signal matches the reference one as good as possible.

We will denote the reference signal by $R = (R(t_1) \dots R(t_M))$ and the test signal by $T = (T(t_1) \dots T(t_N))$.

At the first step of DTW so called local distance matrix d are calculated. Matrix has the following form:

$$d = \begin{pmatrix} d(1,1) & \cdots & d(1,N) \\ \vdots & \ddots & \vdots \\ d(M,1) & \cdots & d(M,N) \end{pmatrix}$$

We denote the distance between $R(t_m)$ and $T(t_n)$ by $d(m,n)$, where $m \in [1 \dots M]$ and $n \in [1 \dots N]$ are indices of reference and test signal respectively. Shortly we say that d is the local distance matrix includes of pairwise distances between each pair of points (m,n).

At the second step accumulated distance matrix D is calculated based on matrix d. Matrix D has the following form:

$$D = \begin{pmatrix} D(1,1) & \cdots & D(1,N) \\ \vdots & \ddots & \vdots \\ D(M,1) & \cdots & D(M,N) \end{pmatrix}$$

In general, the element $D(m,n)$ of the accumulated distance matrix defines the minimal total distance of the two signal segments $R_m = (R(t_1) \dots R(t_m))$ and $T_n = (T(t_1) \dots T(t_n))$.

The calculation of D needs the initial steps:

$$D(1,1) = d(1,1) \quad (1)$$

and

$$D(m,1) = d(m,1) + D(m-1,1), \quad (2)$$

$$D(1,n) = d(1,n) + D(1,n-1), \quad (3)$$

which give the elements of the first row and the first column of D. Then the other elements can be calculated iteratively by

$$D(m,n) = d(m,n) + \\ + \min(D(m-1,n), D(m,n-1), \\ D(m-1,n-1), \quad (4)$$

here $m \in [2 \dots M]$, $n \in [2 \dots N]$ are indices of points of reference and test signals respectively.

Formula (4) allows finding minimal transition to the next point from all admissible ones. Admissible transitions are defined with the help of so-called weighting matrix, which includes all possible ways of transition from one point to another. This matrix is sometimes also called step-pattern [14].

It should be noted, that this formula is used only in algorithms without modifications of weighting matrix.

At the third step, the optimal warping path w are calculated based on accumulated distances matrix. This path is the shortest admissible way from the first element of D to its last element. Calculation of w starts from $D(M,N)$ and goes backward until finding element $D(1,1)$. In this way we get the connection of both time series with the shortest total length. This is only a short explanation of mathematical background of DTW. Detailed mathematical description of this is given in [3].

There are two types of constraints of classical DTW algorithm: global constraints and local constraints.

Global constraints mean that both first and end points of reference signal must match starting and ending points of warped test signal, i.e. the following conditions are satisfied

$$M1 = w(N1); \quad (5)$$

$$M2 = w(N2), \quad (6)$$

where $M1, M2$ are starting and ending points of reference signal respectively and $N1, N2$ are starting and ending points of test one [9].

Local constraints mean that warping path is able to go only from current point to its closest neighbors and is not able to go backward. In this case weighting matrix has the following form

$$((m, n-1); (m-1, n-1); (m-1; n)), \quad (7)$$

where $m \in [1 \dots M]$ and $n \in [1 \dots N]$ are numbers of points of reference and test signals respectively. Visual representation of classical weighting matrix is shown in Fig. 8. More detailed information about local constraints is given in [3].

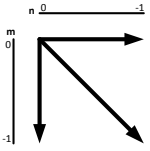


Fig. 8. Weighting matrix for classical DTW algorithm.

V. Modifications of DTW Algorithm

The idea of DTW modification is in relaxing or even eliminating global or/and local constraints.

Local constraints modification is realized with changing of classical weighting matrix. There are two ways of this changing. First is including weighting coefficients to all admissible ways of transition in classical weighting matrix. Second way is adding new admissible ways of transition or modifying existing ones [10][15].

Global constraints modifications are realized with use of open-beginning and open-end approaches [15]. Open

beginning approach assume that condition (5) presented in section IV not to be satisfied and open-end approach allows condition (6) not to be satisfied.

In this work modification with relaxing of local constraints using approach of adding new admissible transitions to the weighting matrix are described.

The modified weighting matrix has the following form:

$$((m-1,n);\ (m,n-1);\ (m-1,n-1);\ (m-2,n-1);\ (m-1,n-2)) \quad (8)$$

Visual representation of the matrix is shown on Fig. 9.

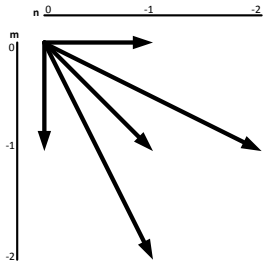


Fig. 9. Weighting matrix for modified DTW algorithm.

VI. Testing Procedure and Functions, Created During Testing

The first goal of testing is sensitivity estimation of classical DTW algorithm and its modifications to changes of input signals and differences between reference and test signals.

The second goal is comparison of classical and modified algorithm to find the most appropriate variant for solving the research task.

In the testing process so-called subsequent matching of test and reference signal are used. In this case, reference signal is a template, imposed on the test signal to find fragments of reference signal in test one. This approach is suitable for testing both simple signals and ones that are more complex.

Testing procedure includes two steps: testing of classical DTW algorithm and testing of its modification with changed weighting matrix.

The same signals are used for testing both DTW variations.

Length of reference and test signals are used in this work is not the same.

Reference signal has constant parameters during testing and parameters of test signal are changed with respect to parameters of reference one.

During testing the following diagrams and graphs are created:

- Diagram of local distances matrix
- Diagram of accumulated distances matrix with the graph of an optimal warping path
- Graphs of signals before and after time warping procedure
- Graph of matching function, which illustrated matching between single points of signals.

VII. Testing Classical and Modified DTW Algorithm

Series of experiments with different test signals was carried out on the research. The most significant results are presented in this article. In this section testing of signals with the following parameters are presented.

Reference signal parameters:

- amplitude – 1
- width of impulse – 100
- no initial displacement
- final zero-valued segment is not existent.

Parameters of test signal corresponding to parameters of reference signal:

- signal amplitude – 0.8 of amplitude of reference signal
- signal width – 0.7 of reference signal width
- initial displacement – 0.4 of reference signal width (starts with the starting point of reference signal)
- length of final zero-valued segment – 0.45 of reference signal width.

Both reference and test signals are shown on Fig. 10.

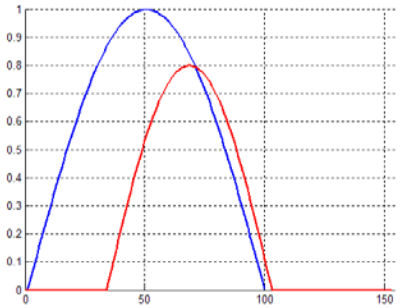


Fig. 10. Both test and reference signals (reference are blue, test are red).

Diagrams and graphs obtained during testing this pair of signals are described below.

At the first step, classical DTW algorithm was tested.

Local distance matrix (Fig. 11) is shown just for visualization of relationships between reference and test signals. This matrix does not contain any significant information except pairwise distances between each points of reference and test signals. At the diagram, lighter regions correspond to shorter distances between points; darker regions correspond to longer distances.

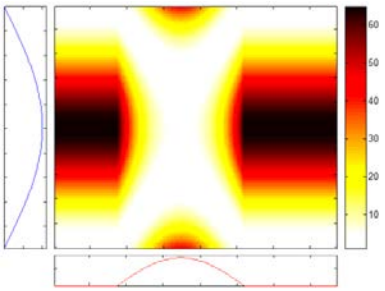


Fig. 11 Local distances matrix.

Accumulated distances matrix with the optimal warping path is shown in Fig. 12. As in diagram above, lighter regions correspond to shorter distances between points; darker regions correspond to longer distances. Optimal warping path (black line) always lies in the region of the shortest accumulated distances.

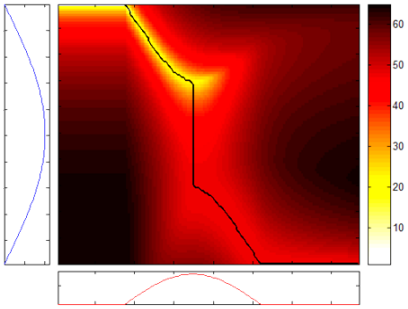


Fig. 12. Accumulated distances matrix and optimal warping path (black line) for classical algorithm.

Experiments showed that decreasing of test signal amplitude with respect to amplitude of reference signal caused appearing of vertical linear segment in the optimal warping path. Length of this segment corresponds to amount of points of reference signal, which are placed above test signal. This length increases linearly with linear increasing the difference between amplitudes of reference and test signals.

Segments of simultaneous increasing and decreasing both reference and test signals corresponds to polygonal lines placed close to diagonal before and after linear segment. When shrinking test signal with respect to base signal, these segments go left with increasing of width difference between test and reference signals; when stretching test signal, segments go right.

Initial displacement and final zero-valued segment correspond to horizontal lines in optimal warping path. Length of these lines depends on length of corresponding parts of test signal.

Signals before (left graph) and after (right graph) time warping are shown in Fig. 13.

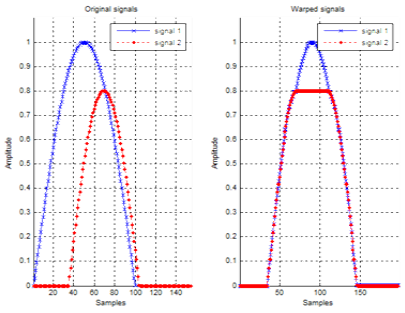


Fig. 13. Both reference and test signals before (left graph) and after (right graph) warping for classical algorithm.

Experiments showed that classical DTW algorithm well compensates differences of width between reference and test signals. The algorithm is also good for compensating the time shifts because it compensates both initial displacement and final zero-valued segment. However, this algorithm is not able to compensate differences in amplitude. The upper horizontal segment corresponding to this difference appears on the graph after time warping.

Matching function (Fig. 14) shows matching of single points of reference and test signals with blue lines. In addition, this function contains information about matching distances for each single point of signals, i.e. length of points shifting during time warping.

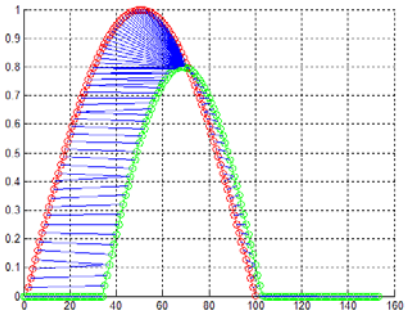


Fig. 14. Matching function for classical algorithm (reference signal is red, test signal is green, blue lines show matching of each point of reference and test signals).

One can see that points are matching irregularly. So-called multiple matching problem appears. This means that after warping of time axis, matching between multiple points of one signal and a single point of another signal appears. Multiple matching problem causes vertical or horizontal segments in the graph of optimal warping path.

On the graph showed above at segments of increasing and decreasing both reference and test signals this problem is not significant because amount of multiple points connected to a single one is not big. However, in the segment corresponding to amplitude difference this problem is very significant because all points of reference signal above test signal are matched with a maximum point of test signal that causes long vertical line in the middle of warping path. Points of the test signal corresponding to initial displacement and final zero-valued segment are matched sequentially to each other.

At the second step, modified DTW algorithm described in the section 5 was tested.

The same signals as at previous step were used for testing.

When testing any modifications of DTW there was no changes in the local distances matrix.

However, accumulated distance matrix (Fig. 15) for modified DTW algorithm is modified.

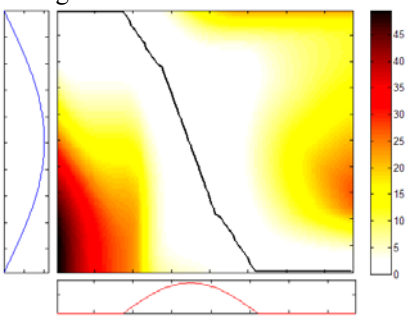


Fig. 15. Accumulated distances matrix and optimal warping path (black line) for modified algorithm.

Since the weighting matrix is changed from (7) to (8), the corresponding formula to (4) gives in general smaller values of $D(m,n)$.

Weighting matrix of this modification has more admissible transitions then in classical one, thus vertical segment of optimal warping path is absent. Instead of this, a linear segment corresponding to amplitude difference appears. As in previous case, “step segments” of optimal path correspond to segments of simultaneous increasing and decreasing both reference and test signals. Horizontal lines of optimal path correspond to initial displacement segment and final zero-valued segment of testing signal.

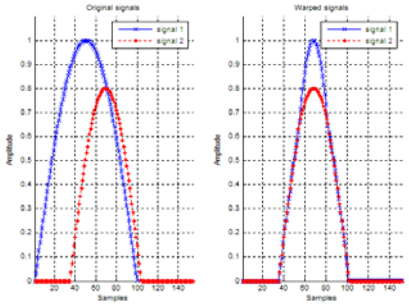


Fig. 16. Signals before (left graph) and after warping (right graph) for modified algorithm.

Studying of warped signal (Fig. 16) shows that when using modified algorithm, it better deals with differences between amplitudes of reference and test signals even if this difference is very big (instead of linear segment in previous case, here the segment, which repeats waveform of signal with lower amplitude). However, this modification is not able to compensate completely the amplitude difference between signals. Algorithm also well enough deals with width differences and shifting of signals between each other.

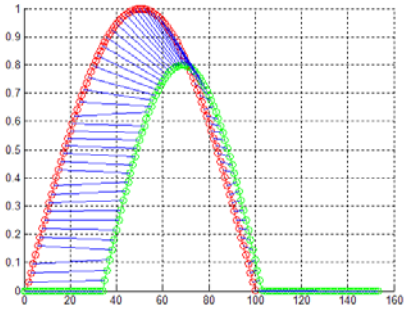


Fig. 17. Matching function for modified algorithm.

Analysis of matching function (Fig. 17) shows that modified algorithm could better match points of reference and test signals. Besides that, multiple matching situations is completely absent. As in classical algorithm, points of the parts with zero-valued signal are sequentially connected to each other.

VIII. Summarizing of Test Results

Comparative analysis of two DTW algorithms showed that classical and modified algorithm well enough compensate the difference in width between reference and test signals if this difference is big (this is normal for real biological signals). Both algorithms also can completely compensate shifting of signals relative to each other.

Modified DTW algorithm better deals with amplitude differences of signals then classical one. It should be noted however, that neither classical algorithm nor modified one is able to completely compensate difference in amplitudes. To solve this problem so-called two-dimensional DTW might be used [16]. This algorithm allows warping both time and amplitude axes.

During testing, problem of multiple matching was studied as well because it can have make significant changes in optimal warping path causing horizontal or vertical segments in the optimal warping path. There are different ways of dealing with this problem: one of them is trying to completely avoid such situations, another way is to avoid these situations but restrict the maximal number of points connected to a single one. The question about amount of points connected to a single point decides based on information about input signals and research task. It should be noted, that maximal amount of multiple points connected to a single one needs to be small because large amounts of points connected to one point do not allow warped function completely match to reference one.

When studying this problem during testing it was found that amount of points matched to a single one depends on two factors: density distribution of signal's points, and amplitude difference between signals.

Experiments showed that influence of amplitude difference might be eliminated using modifications of DTW algorithm. In particular, modification of DTW considered in this article minimizes amount of multiple matching situations and amount of points, which are able to match single point.

For eliminating second factor it is necessary to provide the same density distribution of points along reference and test signals. Besides that, the difference between densities of reference and test signals has to be not very big.

In practice, it is often impossible to provide these conditions because density of points in real biological signals depends on their length, waveform and other factors. In this case, it is necessary to make additional experiments to find the optimal amount of points, which are able to match a single one.

IX. Conclusion

The problem of using DTW algorithm for similarity measurement of biological signals was considered in this work. A short description of classical approach for DTW and some of its possible modifications was given.

Procedure for testing different modifications of DTW algorithm based on artificial signals with changeable parameters was presented. Main requirements for artificial signals, which simulated real biological signals, were described. To present abilities of testing procedure comparative analysis of classical DTW approach and one of its modifications was carried out.

Results of analysis showed some advantages of considered modification in contrast to classical DTW procedure. However, for more detailed study of DTW algorithms features and their behavior when testing real biological signals more tests with more complex signals is needed.

At the next steps of research comparing of large amount of different modifications of DTW procedure will be carried out based on presented testing procedure. It is planned to increase amount and complexity of test signals and then use different examples of real data to check the results of experiments with artificial data.

Another direction of further work is developing of methods of effective extraction of information from the output parameters of DTW algorithm. During studying of output data of classical DTW algorithm it was found that some of them are not informative for analyzing of input signals waveform changes. For this reason, it is necessary to develop additional methods of retrieving information from output parameters of DTW algorithm.

REFERENCES

[1] L. Sörnmo, P. Laguna, "Bioelectrical signal processing in cardiac and neurological applications, 1st Edition", Academic Press, 2005.

[2] G. D. Clifford, F. Azuaje, P. E. McSharry, "Advanced Methods and Tools for ECG Data Analysis", Norwood: Artech House Inc., 2006.

[3] C. Cassisi, P. Montalto et al. (2012). "Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining", in: A. Karahoca (Ed.), "Advances in Data Mining Knowledge Discovery and Applications", InTech, 2012, pp. 71-94.

[4] Y. Jeong, M. K. Jeong, O. A. Omitaomu, "Weighted Dynamic Time Warping for Time Series Classification", Pattern Recognition, No. 44, 2011, pp. 2231-2240.

[5] T.S. Han, S.K. Ko, J. Kang, "Efficient Subsequence Matching Using the Longest Common Subsequence with a Dual Match Index", Proceeding MLDM of the 5th International Conference on Machine Learning and Data Mining in Pattern Recognition, Berlin Heidelberg: Springer, 2007, pp. 585-600.

[6] Y. Zhang, T. F. Edgar, "A Robust Dynamic Time Warping Algorithm for Batch Trajectory Synchronization", American Control Conference, Seattle, June 2008, pp. 2864-2869.

[7] P. Tormene, T. Giorgio et al., "Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation", Artificial Intelligence in Medicine, No. 45, 2009, pp. 11-34.

[8] B. Huang, W. Kinsner, "ECG Frame Classification Using Dynamic Time Warping", Proceedings of the IEEE Canadian conference on Electrical & Computer Engineering, 2002, pp. 1105-1110.

[9] L. R. Rabiner, A. E. Rosenberg and S. E. Levinson "Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-26, Dec. 1978, pp. 575-582.

[10] M. Müller, "Dynamic Time Warping", Information Retrieval for Music and Motion, Berlin Heidelberg: Springer 2007, pp. 69-84.

[11] L. N. Sörnmo, M. E. Nygårds, "Deliniation of the QRS complex using the envelope of the e.c.g.", Medical & Biological Engineering & Computing, No. 21, 1993, pp. 538-547.

[12] R. Atarius, L. Sörnmo, "Detection of Cardiac Late Potentials in nonstationary noise", Med Eng Phys, 1997, pp. 291-298.

[13] P. Lander, E.J. Berbari "Time-Frequency Plane Wiener Filtering of the High-Resolution ECG: Background and Time-Frequency Representations", IEEE Trans Biomed Eng., Apr. 1997, pp. 247-255.

[14] 6 E.J. Keogh, M.J. Pazzani, "Derivative Dynamic Time warping", Proceeding of the First SIAM of International Conference on Data Mining, 2007, pp. 1-11.

[15] T. Giorgino, "Computing and Visualizing Dynamic Time Warping Algorithms in R: The DTW Package", Journal of Statistical Software, Vol. 31, Issue 7, 2009, pp. 1-25.

[16] M. Schmidt, M. Baumert et al., "Two-Dimensional Warping for One-Dimensional Signals — Conceptual Framework and Application to ECG Processing", IEEE Transactions on Signal Processing, Vol. 62, No. 21, Nov. 2014, pp. 5577-5588.

Methods of Visualization and Analysis of Cardiac Depolarization in the Three Dimensional Space

Sergei Artemov[1], Bernd Krause[2]
[1]Perm National Research Polytechnic University - Electrotechnical Department
Prof. Pozdeev str. 7, 614013, Perm, Russia
[2]Anhalt University of Applied Sciences - Computer Science and Languages Department
Lohmannstr. 23, 06366, Koethen, Germany
E-mail: [1]art-sa@bk.ru, [2]bernd.krause@inf.hs-anhalt.de

***Abstract*—Research analysis of electrocardiograms (ECG) today is carried out mostly using time depending signals of different leads shown in the graphs. Definition of ECG parameters is performed by qualified personnel, and requiring particular skills. To support decoding the cardiac depolarization phase of ECG there are methods to analyze space-time convolution charts in three dimensions where the heartbeat is described by the trajectory of its electrical vector. Based on this, it can be assumed that all available options of the classical ECG analysis of this time segment can be obtained using this technique. Investigated ECG visualization techniques in three dimensions combined with quantitative methods giving additional features of cardiac depolarization and allow a better exploitation of the information content of the given ECG signals.**

***Keywords:* ECG, PCA, cardiac depolarization.**

I. Introduction

The representation of heart activity is formed as a time sequence of potential difference with respect to different points on the human body. There a two types of ECG: session and continuous – use as source data here, it need for keeping track abnormality in heart behavior.

If we imagine the body as a three-dimensional object it is possible to measure the summary vector of potentials changes in space and time. Here we refer to a measurement of this electrical heart vector in a orthogonal X,Y,Z-coordinate system with a given sampling rate in time (e.g. 1 kHz). Based on these three dimensional signals and a background knowledge of heart muscle structure - including knowledge about pulses excitation, formation and propagation - it is possible to analyze the heart beats and eventually identify pathologies. It is a challenging task to visualize ECG and analyze them in three dimensional space in a way that would be most understandable to medical staff and to get as much data as possible for further analysis and data mining tasks.

From three dimensional signals clearly visible additional ECG features are derived, for example, the rotation of the electrical heart angle, the dynamic characteristics of angle changes and unforeseen changes n heart position relative to the body which may be a result of respiratory movements and so on.

II. Classical ECG Representation

The standard approach to visualize ECG signals is time charts of single leads. Parameters such as the rhythm frequency, the amplitude of the peaks, the beats regularity, the average heart electrical angle, as well as various kinds of deviations, and other can be evaluated. Mostly evaluation is carried out through a review of the peaks and the parts of the signal.

The most interesting part is the QRS complex because here cardiac depolarization of the heart muscle occurs. This is the most dynamic activity of the heart. Figure 1 shows graphs [1] of two beats with similar characteristics as beat amplitudes and morphology, where morphology is characterize by the peaks Q, R, S.

First beat is used as reference, but second one comes from a beat which is followed directly by tachycardia.

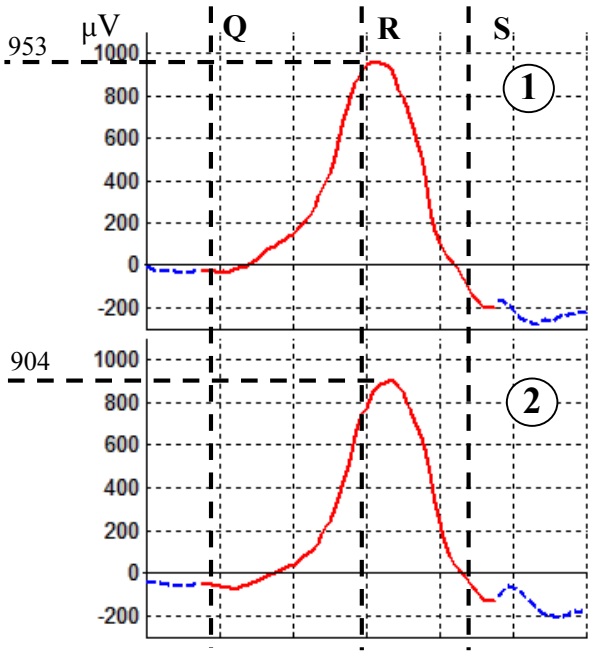


Fig. 1. Comparison of the QRS graphs of beats.

At a first view the difference between the graphs are invisible, they have a similar shape, amplitude, pulse length. But it is difficult to distinguish morphological differences in their shape. We have a similar situation with other leads. To reveal differences in behavior we need joint information of all three leads.

III. Three Dimensional Representation

To improve the ability to detect differences in behavior of the heart a three dimensional consideration of ECG is necessary.

The initial data are the three leads of X,Y and Z components of a heart vector in three-dimensional space which describe time-space trajectory of this vector. The new approach allows estimate the interdependency of the leads and determine the invisible changes in the time diagram [2][3]. Figure 2 shows a graph of a QRS loop of a beat in the three-dimensional space with marked QRS peaks.

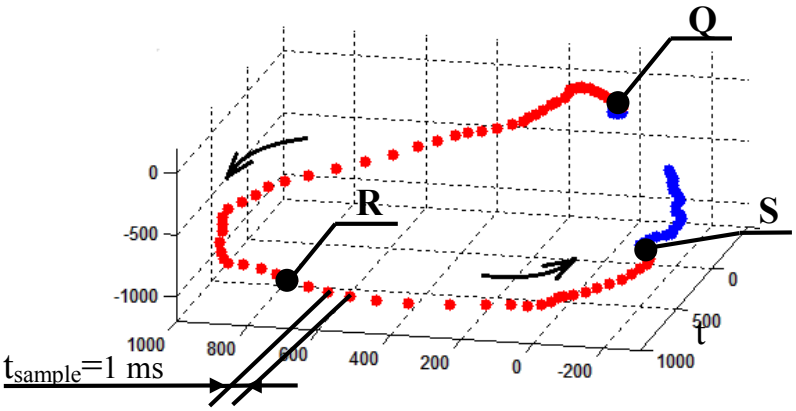


Fig. 2. A three-dimensional representation of QRS part of a beat.

The graph in figure 2 represents a loop where the heart vector runs through its trajectory with a specific kind of bending, twisting, compression and expansion - the dynamic characteristics of the signal.

Figure 3 shows images of the previous beats (from Fig. 1) in three dimensions. The loops are presented from different perspectives. The beats have similar morphologies which are rotated in space.

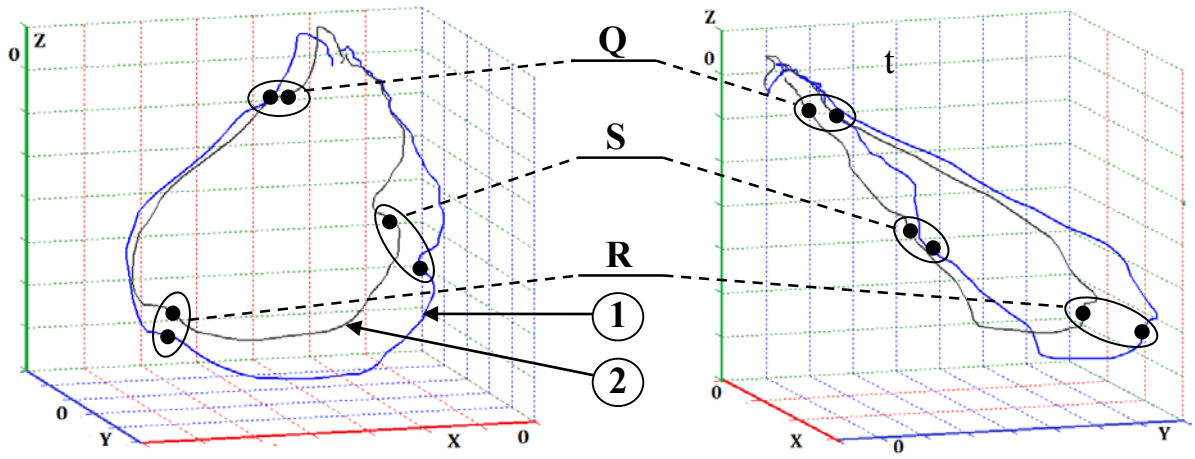


Fig. 3. Beats graphics with different views.

IV. Approximation of the whole QRS Complex by Plane

Three dimensional visualization of ECG is not sufficient for further quantitative data analysis. For detailed data analysis and data mining tasks additional features are necessary witch give additional information.

The simplest way to represent QRS loop is the mean vector of heart electric potentials.

The classical generalization of this dimensions reduction technique is the principal component analysis (PCA). This technique allows the transformation of a set of points into a space with a smaller dimension k>0. For our purpose a two dimensional plane is adequate.

Figure 4 shows the rotated approximating planes of the two beats from figure 3. Obviously the trajectories of both beats are more similar in their individual planes than in other representations. That's why, it is important to identify these planes and their mutual relation, including the rotation which is necessary to match both trajectories as good as possible.

PCA method description [4][5]:

Given a finite set of vectors (points):

$$x_1, x_2, \dots, x_m \in \mathbb{R}^n \quad (1)$$

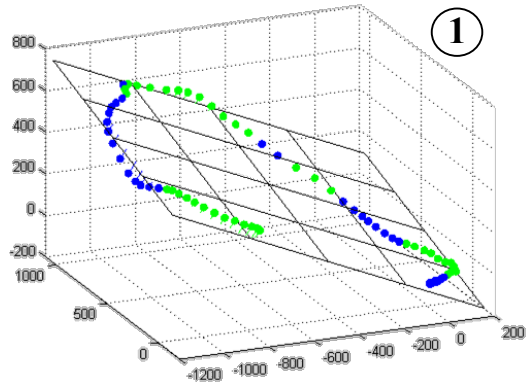
where in our case we have $n = 3$ and $x_i = (X(t_i), Y(t_i), Z(t_i))$ with equidistant $t_1 < t_2 < \cdots < t_m$. Generally the task is to find (for $k = 0,1, \ldots, n-1$) linear subspaces L_k,

$$L_k \in \mathbb{R}^n \qquad (2)$$

such that the sum of the squared distances from the x_i to L_k was minimal

$$\sum_{i=1}^{m} dist^2(x_i, L_k) \rightarrow min \qquad (3)$$

where $dist(x_i, L_k)$ is the Euclidean distance from the point to the linear manifold. To solve this problem one has to subtract the mean of the vectors $x_1, x_2, \ldots, x_m$ from each point to center the data around the origin.



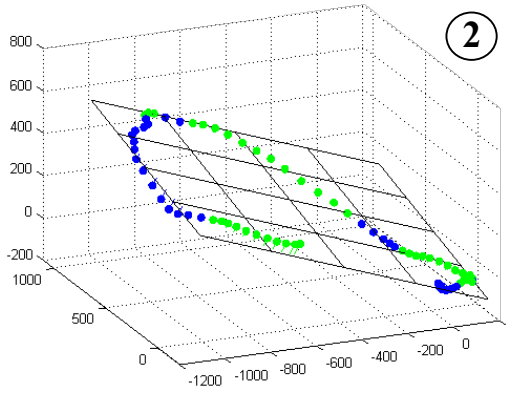


Fig. 4. Approximating planes for two beats.

Then the Hesse normal form of the plane (k=2) with normal vector n allows to express the sum of these distances by

$$\sum_{i=1}^{m} (n^T x_i)^2 = n^T \left(\sum_{i=1}^{m} x_i x_i^T\right) n \rightarrow min \qquad (4)$$

where the centered data are also denoted by $x_1, x_2, \ldots, x_m$. Since the Matrix $A = \sum_{i=1}^{m} x_i x_i^T$ is symmetric and positive definite it has three strictly positive eigenvalues and an orthogonal system of three (normalized) eigenvectors. These vectors are the principal component vectors.

The approximating plane is spanned by the first two eigenvectors of A and its normal vector n is orthogonal to them, that is, it is the eigenvector to the minimal eigenvalue which minimizes (4).

Figure 5 shows again the approximating planes of the previous two beats (Figure 3, 4) in a common coordinate system. The graph displays the normal vectors of both

planes, called global orientation, and the exact rotation expressed by the angle α between these normal vectors.

On the basis of medical arguments – this rotation can be explained by breathing, movement or specific orientation of heart for each person.

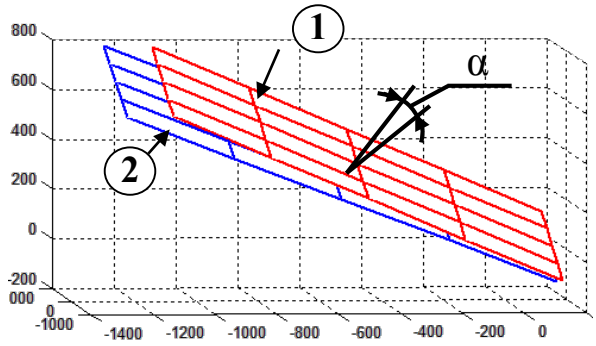
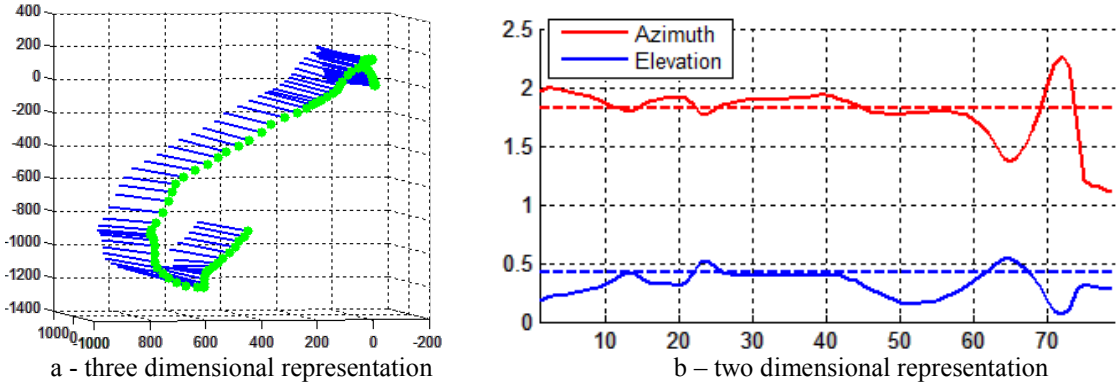


Fig. 5. Approximating planes of two beats.

V. Local Approximation of QRS Segments

The previous example describes how to find approximating plane and global orientation n of QRS complex based on all point of the loop. For finding a corresponding local orientation n_i the approximating plane is calculated only for the part of the loop given by a sliding time window Δt. Thus by shifting time window for each point t_i (the midpoint of the window) we get the local orientation vectors n_i.

In figure 6-a local orientation vectors are shown for Δt=25ms. In figure 6-b the local orientation vectors are displayed. Additionally, the dashed lines indicate the global orientation vector.



a - three dimensional representation

b – two dimensional representation

Fig. 6. Graphs of local orientation vectors.

VI. Conclusion

Representation of ECG depolarization part in three-dimensional form is one of the new ways of presenting information. For further clear data representation such visualization methods are useful. They give directions for further investigations and analysis. Based on this additional data are available that show features and relations between the individual output leads as new signals.

These data allow consider additional data for learning and data mining tasks and increase accuracy of methods and diagnostic abnormalities in heart behavior.

Furthermore, this approach can be applied to other parts of ECG which also allow predicting the behavior of heart beating such as late potentials parts ST complex etc.

References

[1] M.-E. Nygards, L. Sornmo,"Delineation of the QRS complex using the envelope of the ECG" - Medicine and biology, 1983, No 21 pp.538-547.

[2] M. Kotas, "Projective Filtering of Time-Aligned ECG Beats" – IEEE transactions on biomedical engineering vol.51, No.7, 2004, pp. 1129 – 1139.

[3] L.G. Tereshchenko, L. Han, A. Cheng, "Beat-to-beat three-dimensional ECG variability predicts ventricular arrhythmia in ICD recipients" – Heart Rhythm Society, 2010, pp. 1606 – 1613.

[4] W Müller, T. Nocke, H. Schumann , "Enhancing the visualization process with principal component analysis to support the exploration of trends" - Asia-Pacific Symposium on Information Visualization, 2006.

[5] I.T. Jolliffe, "Principal Component Analysis", Springer, New York, 2002.

Handling the Problem of Unbalanced Data Sets in the Classification of Technical Equipment States

Obukhov Egor
Perm National Research Polytechnic University - Electrotechnical Department Komsomolsky Ave. 29, 614990, Perm, Russia
E-Mail: obuhov2014@bk.ru

***Abstract*—Questions of handling unbalanced data considered in this article. As models for classification, PNN and MLP are used. Problem of estimation of model performance in case of unbalanced training set is solved. Several methods (clustering approach and boosting approach) considered as useful to deal with the problem of input data.**

***Keywords:* unbalanced data, probabilistic neural net, multilayer perceptron, classification, evaluation of performance, preparation of data.**

I. Introduction

This article deals with comprehensive estimation of the technical state of complex technical systems, based on a structure-approach and further analysis and dynamic monitoring of the structure elements. Most of all a result of technical state estimation of the equipment is determined by selecting the most informative parameters of monitoring: vibration, pressure, temperature etc. Obviously, it is important to make a correct diagnostic model to develop good methods for the recognition of different states of the technical equipment [1].

Thereby, to develop a high-quality model it is necessary to use the most significant parameters of monitoring and to have a representative data set that means the data set have to give comprehensive information about possible states of the monitored equipment.

II. About Data and Problem of It

The technical equipment has special detectors to track and monitor parameters. The values of the parameters are taken from detectors at a certain frequency and are transmitted to a database. Therefore, the values transmitted in the database are used as dataset to train model of prediction future state of the equipment.

The data set has around 2000 instances and a very large dimension. It has advantages such as absence of missing values, but also disadvantages. The main disadvantage is that the data set is unbalanced. This means there is a big bias in the amount of instances of the classes. For example if we have two classes, the first one contains 100 instances and the second one 1000 instances, so the first class is a minority class of rare instances, the other is a majority class.

Most machine learning algorithms work better when both classes of the training set has an equal number of elements. If the number of instances of one class is far different from the other, then problems appear. This is the best illustrated with following example.

If consider the current data set (parameter monitoring) and train a machine learning algorithm on the data, suppose there are two possible outputs as follows:

1) 10 instances of minority class and 20 of the majority class are misclassified.
2) 2 instances of minority class and 60 of the majority class are misclassified.

If we calculate the performance of the model by the amount of misclassified examples, then obviously that first case is better. However in terms of correct classification of minority class (especially when the class has priority), then second case is the best choice. Thereby it is important to have a correct metric for efficient estimation of the model performance.

Unbalanced data sets [2] are a special case for classification problems. This type of sets supposes a new challenging problem for Data Mining, since standard classification algorithms usually consider a balanced training set. So the question is how to handle with it?

The goal of the article is to find a way of handling with unbalanced data sets and improve the performance of the unbalanced data sets classification.

For that goal, it is necessary to create a representative and high-quality training set.

III. Approaches to Handle Unbalanced Data

There are different ways to handle unbalanced data. Let's consider several of those, which could be useful for the specific task:

1) In the first approach, it is necessary to divide the major class into L distinct clusters, then train L classifiers, where each classifier is trained on only one of the distinct clusters, but on all of the data from the minority class. To be clear, the data from the minority class are used in the training of all L classifiers. Finally, use ensemble of the L learned classifiers as a final classifier.

2) This is similar to number (1), but a little different. Let N be the number of samples in the minority class. Cluster the majority class into N clusters (agglomerative, K-means

clustering etc.), and use the resulting cluster mediods as the training data for the majority class. To be clear, you throw out the original training data from the majority class, and use the mediods instead. Finally, the classes are balanced.

3) The third one is based on boosting. The algorithm trains the first learner, L1, on the original data set. The second learner, L2, is trained on a set on which L1 has around 50% chance to be correct. The third learner, L3, is trained on the cases on which L1 and L2 disagree. As output, return the majority of the classifiers. Why it improves, the classification can be found in [3].

In case of application this method on the unbalanced data. Let L1 return always true (majority class). L2, is being trained, where L1 inconvenient. L3 trained, when L1 and L2 disagree, that is L2 predicts false (minority class). Therefore, false (minority class) prediction can be only when both L2 and L3 predicts false.

Whatever method one can use will help in some ways, but hurt in others. To improve the efficiency, one can train separate models using all of the methods listed above, and then perform model averaging over all of them.

The methods are not applied yet, and on the stage of research, the task was to find out the ideas and approaches, which could be useful to the task. In future, it is necessary to apply the methods to prove the efficiency and suitability of it.

IV. Data preparation (Normalization and feature selection)

The data are recorded from detectors of the technical equipment and has a large dimensionality around 80 attributes. In order to reduce the dimensionality of the data and find out the effective number of variables, we applied the following approaches:

- Correlation analysis

Take our 80 attributes, one of them is label, so we have around 80 pairs (label + one, every time different attribute). A correlation is a number between -1 and +1 that measures the degree of association between two attributes, which allows us to estimate the degree of association between label and one certain attribute. The table 1 shows the interpretation of correlation values.

TABLE I
Interpretation of correlation values

Negative Value (V)	Positive Value (V)	Interpretation
-0,2 <V< 0,2		Very low correlation
-0,5<V<-0,2	0,2<V<0,5	Low correlation
-0,7<V<-0,5	0,5<V<0,7	Middle value of correlation
-0,9<V<-0,7	0,7<V<0,9	High correlation
V < -0,9	V > 0,9	Very high correlation

By calculating the correlation coefficient for each of the pairs, it is possible to make some filtering and delete less correlated attributes.

- Principal component analysis (PCA)

PCA is method to reduce the dimensionality of the variable (attribute) set, by using a new coordinate system that is lesser in dimension than the number of original variables. This transformation will usually be accompanied by a loss of information. The goal of PCA is to preserve as much information contained in the data as possible.

- Global sensitivity analysis

It is a tool from Statsoft "Statistica" [4], which gives information about the relative importance of the variables used in a neural network. In sensitivity analysis, one determines how the neural net will response (increasing or decreasing error rates) to some changes of its input variables. During the analysis, this tool exclude an attribute and make training of classifier without the attribute. If an important attribute excluded then error increase significantly. If an unimportant attribute excluded, the error will not increase very much.

- Normalization

As a classifier to solve, the task of determination of states of technical equipment was chosen a neural net. In this case, it is necessary to make some normalization of the training data, because in such type of classifier small values of a variable (like 0.5) and big values (like 100) have different influence on the final prediction. It means one has to reduce the range of data to a limit between 0 to 1. The limit depends of activation function type.

The method we applied to normalize the data: Statistical normalization.

The main idea of the method is to convert the data into a set with normal distribution with mean=0. The formula of statistical normalization is $Z=(x-u)/s$ (x-current value, u-mean value of the variable, s- standard deviation).

V. Choosing classification algorithms (modeling)

As models for classification we have chosen two neural nets: probabilistic neural net and multilayer perceptron. We will consider their distinctive features and find out why they are well suited for this kind of task [5].

MLP:

A multilayer perceptron (MLP) is a typical common kind of neural nets and a good fit to almost all types of tasks. It is feedforward neural net, so we have several successive connected layers and each current neuron gets and processes signals from neurons of the previous layer. The input layer is used just to transfer information into a hidden (computing) layer. The amount of neurons in the input and output layers is determined by the dataset parameters (number of attributes and type of target attribute). More information about working of this type of neural nets available in [6].

PNN:

A probabilistic neural net (PNN) is some kind of a neural net which similar to a radial basis function (RBF) network. The main difference is that a PNN has one neuron (in hidden layer) for each point in training set, that means the kind of net requires more memory space to store the model then MLP. It is not a good match to work with large training sets. PNN is quite accurate with small to medium size data sets and it has modifications, such as matrix of loss, which add an additional layer with rates (price) of classification errors. The modification allows to make some accent on one class, the possibility fits well to the current task of determination state of an engine. More information about the structure of hidden layers and activation functions available in [7].

If we consider the question of tuning the model's parameters [8], then compared to PNN (where is only one parameter to choose), a MLP has the following parameters:

- Amount of neurons of hidden layer

A way used to calculate amount of neurons in the layer: (number of attributes + number of classes) / 2 + 1

- Training cycles
- Learning rate
- Momentum

For automatically tuning the parameters and finding the best combination of them, we used the statistica program tools. It allows to set some limits for parameters and to train specified number of nets. As a result, we have nets ranked by performance.

VI. Metric to Estimate the Performance of Models

As was said before, such typical metric as classification accuracy is not a good metric, because if a model correctly classify just instances of majority class, then the model have high accuracy by using the metric of estimation.

In the case, when the minority class is a class which represents fault states of the technical equipment and it is more important to accurately classify the vectors of the class, than vectors of the other class.

That is why, the confusion matrix was chosen as a metric to determine the quality of the model. The matrix allows to estimate the recall of a specific class and getting a clearer representation of the model efficiency.

To calculate the recall of the minority class in the data set (monitoring parameters of technical equipment) it is necessary to split the dataset into 3 parts: training set, validation set, testing set. The first step is to train the classifier with data from the major part, which is more than 50 % of whole data set. The second step is to make some intermediate validation of model efficiency following by tuning the model. The last step is a final one-time testing. It allows to make more accurate estimation of the model performance, because testing on a validation set is some kind of training (changing parameters of a model and retraining the model). The results of the final testing are shown in table 2.

Calculation of classification accuracy:

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} = \frac{323+2130}{323+2130+132+57} = 0{,}92$$

Therefore, the result is an accuracy equal to 92%, which is a good result for classification. However, the method of estimation does not represent the whole efficiency of the model.

Calculation of the recall of class True:

$$Recall = \frac{tp}{tp+fn} = \frac{323}{323+57} = 0{,}85$$

Comparison of classification results PNN and MLP:

The result is 85 percent, that means class True is recognized with probability of 85 percent. So with propability of 15 percent a failure or failts will not be recognized, that's not enough for effective operation of a system.

Training the models (MLP and PNN) on unbalanced data/ unrepresentative data showed, performance/accuracy of the models are equal, however it is not enough for effective performing of the current task of technical state of technological equipment determination.

TABLE II
Confusion Matrix

		Real values / classification	
		True	False
Predicted values / classification	True	tp = 323	fp = 132
	False	fn = 57	tn = 2130

VII. Conclusion

In the paper it was explained what unbalanced data are and what influence it has on the performance of a system to classify different states of a technical equipment. Principal approaches of handling unbalanced data are discussed. Making a representative training set by using k-means clustering is shown. Several machine-learning algorithms such as probabilistic neural net and multilayer perceptron have been chosen as models to deal with the task of determination of different states of the technical equipment.

The traditional methods of diagnosis and control of parameters of the technological equipment are time-consuming and it doesn't present a possibility for express analysis. The results of the approaches presented in the article can be useful in some difficult situations as additional information to make a good decision.

Future research will be concentrated on applying all of that approaches and methods of handling with unbalanced data. The main issue is to figure out how much does this affect the final performance of the model. It is important to know which type of neural net fits to the case most.

References

[1] Basmanov M., Menshikov S., Morozov I., Strebkov A.," System of parameter's diagnostic of GTU: modern approach" Delovaya Rossia N7, 2011.-42-43 p. (In Russian)

[2] Icamaan B. Viegas da Silva; Paulo J. L. Adeodato, "PCA and Gaussian noise in MLP neural network training improve generalization in problems with small and unbalanced data sets" Proceedings of International Joint Conference on Neural Networks, San Jose,

[3] Robert E. Schapire, "The Strength of Weak Learnability" Machine Learning, 5(2):197–227, 1990

[4] Statsoft program "Statistica". [Online]. Available: http://www.statsoft.ru/

[5] Dolina O.N. and Kyzmin A.K. Particularities of developing expert systems based on neural net modeling" Journal of Saratov technic state university, 2009.-266-272 p. (In Russian)

[6] S. Sathyanarayana, "A gentle introduction to backpropagation", July 22, 2014

[7] Specht, D. F. "Probabilistic neural networks". Neural Networks 3:1990.- 109–118 p.

[8] A.I. Gavrilov and P.V. Evdokimov, "Neural Network optimum parameters determining under industrial process mathematical model construction" Vestnik of ISPU N4, 2007(In Russian)

Possible Approaches to Commercializing Applied University Research: a Russian Case Study

Leonid Mylnikov
Perm National Research Polytechnic University – Innovation department
Komsomolskiy Avenue 29, 614990, Perm, Russia
E-mail: leonid.mylnikov@pstu.ru

***Abstract*—This paper discusses the environment around universities in Russia and possible approaches to analyzing and choosing the method for applied research results commercialization as well as selecting promising applied research areas in that environment. Conceptual foundations for decision making during the commercialization and roadmap/action plan creation processes are outlined. These can be useful to both universities for planning their activities as well as for organizations that plan to cooperate with universities or that are interested in university generated research. This being said, obtained models and used evaluation parameters may be unique and may depend upon the particular project, university, region, and personal preferences of decision makers. Thus, consideration of these parameters and characteristics only has merit when making decisions in the dynamics of change of these parameters. For this purpose statistical information is needed that characterizes the competencies of the research organization (university) in question, needs of partner organizations, governmental and societal requirements, and science and technology prospects. After determining the promising research areas it's time to look at particular projects, which in turn are also characterized by various parameters dependent upon their objectives. Considering the values of these parameters in their dynamics allows control of project parameters in the course of its execution. This in turn allows prediction of negative situations and alleviation of such by setting the target values of parameters and using best practices and standardization of management processes to achieve those values.**

***Keywords:* innovation, infrastructure, commercialization, strategy.**

I. Introduction

New research results suitable for use in products constitute a resource, the existence of which is a strategic advantage. Implementation of such results in the form of innovations requires key competencies, which the developers usually lack. Creating innovations in market economy conditions is one of the factors of accelerated and sustainable development. The process of creating and managing innovations has in itself become an innovation [3], which indicates insufficient development of innovation creation and implementation methods.

Development of innovations management theory and practice focused on solving local problems due to difficulties of managing innovations as a system [4], which created a deficit of formal methods for management and foundation [5] for implementing innovations as a system. As a result, regardless of all the advantages of comprehensive innovations management there are no systematic models for implementing innovations as complex systems in previous research, which doesn't allow to effectively solve innovations management, expertise, and validation problems.

New developments are created in R&D organizations. This creates new opportunities for analysis of methods of their commercialization, promising research avenues, and production and economy systems based on statistical bibliographical data, patent data, and data of statistical analysis of joint science and industry projects and their common requirements.

II. Description of Environment and Conditions in which Innovation Projects Based on Research Results Appear

Results of scientific research in the form of innovation projects are implemented in open systems, creating a lot of uncertainties in the form of links to external systems. Under conditions of uncertainty the effectiveness of projects depends on the effectiveness of resolving these uncertainties and consequently on information expenses. Information expenses as well as lost profit depend on time. Time plays a special role in innovation projects because innovation projects' lifetime has a tendency to shorten, while the number of product modifications increases. The amount of uncertainty and the complexity of the management problem are also high due to the increasing number of components and subsystems of the production and economic system that are involved in product implementation. Structure of economic relations has a direct influence on production [1] (industrial production index) and added value created when implementing scientific development into an innovation project.

Appearance of such projects is impossible without considering the environment in which they are created and implemented. Projects originate in scientific and educational institutions that exist in the environment represented schematically in fig. 1.

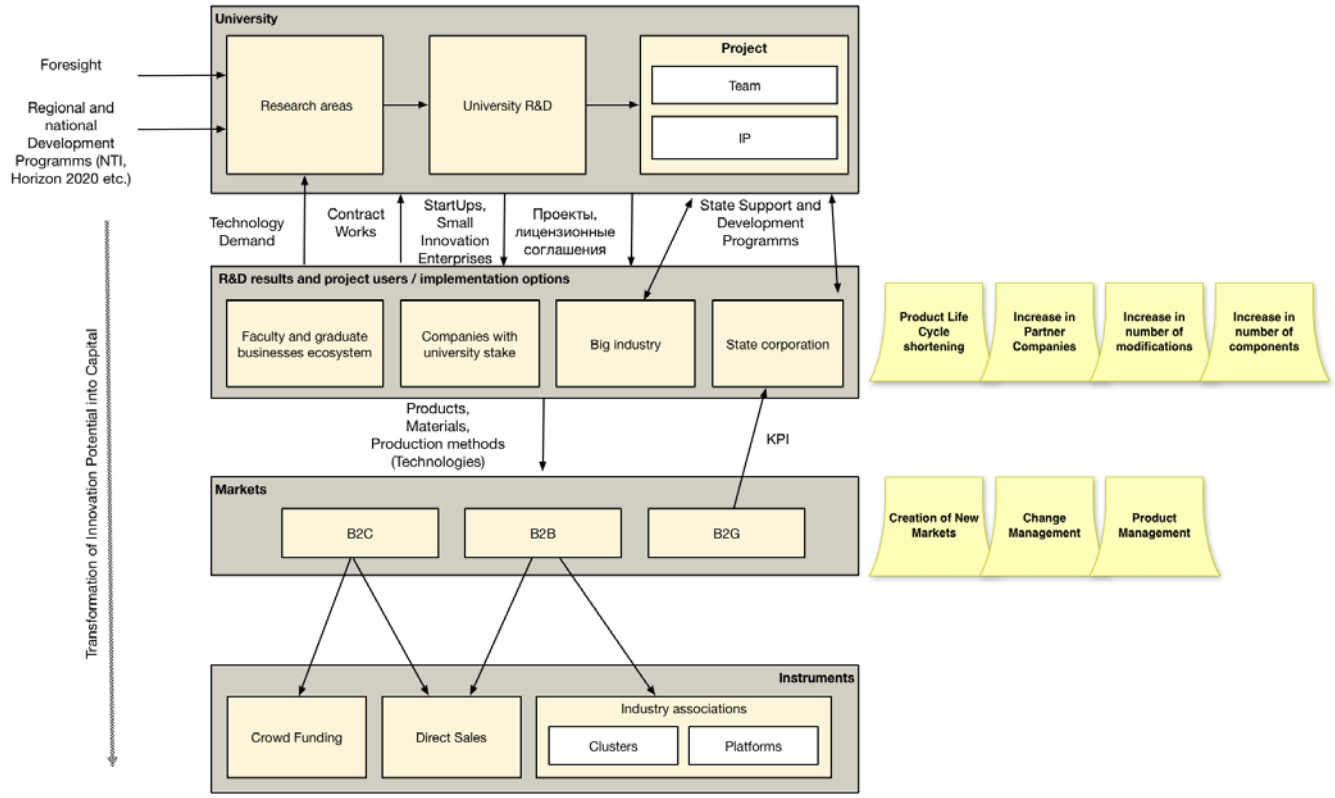


Fig. 1. Innovation commercialization structure in Russian universities.

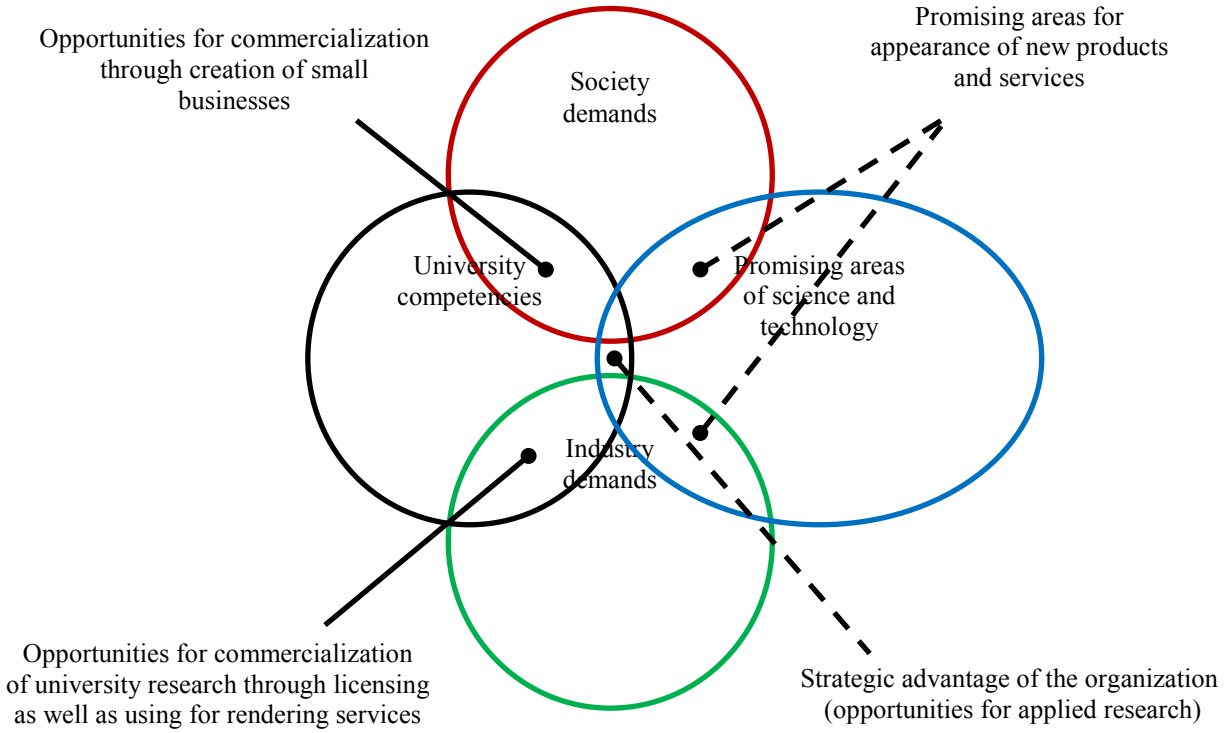


Fig. 2. Opportunities and promising areas for development in university research.

III. Determining promising commercialization avenues for university environment

Choosing the best method for implementing research results depends on project specifics, which can be determined based on information about the project and statistical data about the environment in which the project originated (see fig.2) considered in their dynamics at the moment of appearance of the underlying research results.

To determine university competencies and promising science and technology areas citation dynamics in citation databases and received patent dynamics can be used. Industry demands can be determined by studying the dynamics of requests submitted to research organizations as well as topics of joint research with industry (actual figures can be seen in Table 1).

Analysis of publication activity dynamics and that of the topics of industry demands will be described by the innovations curve. Thus, it is possible to determine the development stage of a technology or scientific area as well as its future growth potential.

These data are measurable and can therefore be forecast [2] using trends in the form of functional descriptions. Technology lifetime, its scientific and economic potential can be evaluated.

Description of the situation around research organizations allows targeted data collection and comparison. Comparison of these data allows determining promising areas of science and technology for appearance of new products (see Table 2, for example).

Regardless of limited statistics it can be concluded that university competencies lie in the areas of material science, mechanical engineering, and information technology, while the demands besides these areas lie in management systems and life sciences. This indicates that commercializing university's own research results through small innovation enterprises in robotics (mechanical engineering and IT), performing contract work in management systems for the industry, and interdisciplinary research in cooperation with specialized organizations in life sciences can be effective for the university.

When developing a commercialization strategy for a university or research organization it must be considered that not every university possesses competencies in all areas. As a rule, each organization has a specialization. This is supported by development programs, which determine the specialization. Thus, when evaluating commercialization opportunities and choosing research areas one should consider research organization's specialties and associated areas as priorities. Associated areas are chosen based on product requirements. This approach allows construction of new products using existing technologies and according to demands based on morphosythesis (morphology machine) or Koller operations.

TABLE 1
Factors and indicators for determining promising avenues for innovation projects in certain areas

Factor	Indicator
1.Research organization (university) competencies	1.1. Patents and software licenses as well as databases for a certain research area belonging to the research organization in question.
	1.2. Number of small companies with university stake in a certain area of science and technology.
	1.3. Number of independent R&D performed for big industry.
	1.4. Supported applications within the university's strartup support programs (for example START program).
2. Demands of industry partners	2.1. Number of joint applications for government programs in a certain area.
	2.2. Number of independent requests from industry for contractual R&D work in a certain area of science and technology.
3. Government and society demands	3.1. Promising areas determined by foresight.
	3.2. Promising avenues for development of science and technology by industry (national technology initiative).
	3.3. Amount of innovation goods and services (Rosstat).
4. Promising areas of science and technology	4.1. Number of publications in a certain area of science (in one of the international citation databases) with a geographical filter.
	4.2. Number of publications in a certain area in industry journals (in one of the international citation databases, given in percentage of total number of publications or percentage of leader's publications) with a geographical filter
	4.3. Number of patents with a geographical filter

When analyzing specific projects one must consider macroeconomic parameters of the region in which the project is to be implemented and the specifics of the particular form of implementation. To this end one must consider such statistical data as new business density dynamics for 1K population (World Bank data), business survival and development rate, (entrepreneurship support foundations data, for instance START program data for 2nd and 3rd round participants), data about the research organization that initiated project implementation (number of staff in small innovation enterprises with university stake that are already operating, their working capital, number of such companies, total attracted investments), data about similar projects (product price and sales dynamics) etc.

To evaluate dynamics one should determine the rate and acceleration of the rate of change of curve values at a given moment (value of the first and second derivatives). If the acceleration rate is positive, a market creating technology can be expected to appear. At zero acceleration and positive rate of change positive innovations can be expected. At the same time it can be a signal to the fact that a new technological solution is necessary to solve the problem as the existing one has exhausted itself[1].

For detailed analysis of the project with a large volume of statistical data on similar projects (in cases when statistics about the project is inaccessible, similar project data can be used) other data can be used that has the greatest impact on project evaluation with regards to its objectives. These parameters can be chosen based on algorithms of the SlopOne family using all the collectible data.

Considering the values of parameters in their dynamics allows managing project parameters in the process of project implementation, thereby predicting negative situations as they occur and alleviation them by setting the target values of parameters and using best practices and standardization of management processes to achieve those values. These methods are connected with training and use of information systems and are aimed at detecting conflicts when implementing projects and choosing the most adequate technical and managerial solutions.

TABLE 2
EXAMPLE OF STATISTICAL INFORMATION FOR DETERMINING PROMISING AREAS OF SCIENCE AND TECHNOLOGY AND OPPORTUNITIES FOR COMMERCIALIZATION OF R&D RESULTS (PERM NATIONAL RESEARCH POLYTECHNIC UNIVERSITY (RUSSIA) 2013-2015)

	Factor according to Table 1.	**New materials**	**Mechanical engineering, materials processing**	**Information technology (development and use)**	**Energy and resource saving, geology, recycling**	**Robotics, mechatronics**	**Technological chemistry**	**Management systems**	**Consulting**	**Electrical power generation and**	**Life sciences**	**Nanotechnologies**	**Electronics**
1.Organization competencies	**1.1.**	**40**	**80**	**134**	**1**	**5**	**6**	**2**	**3**	**12**	**9**	**56**	**2**
	1.2.	**1**	**4**	**2**	**2**	**-**	**-**	**2**	**2**	**1**	**-**	**3**	**-**
	1.3.	**4**	**32**	**15**	**4**	**3**	**1**	**2**	-	-	-	-	-
	2013	-	-	8	1	1	-	1	-	-	-	-	-
	2014	2	21	6	-	-	-	1	-	-	-	-	-
	2015	2	11	1	3	2	1	-	-	-	-	-	-
	1.4.	**-**	**2**	**-**	**-**	**1**	**-**	**-**	**-**	**1**	**-**	**-**	**-**
2. Industry partner demands	**2.1. (поддержано)**	**-**	**6**	**-**	**-**	**-**	**1**	**-**	**-**	**-**	**-**	**-**	**-**
	Подготовлено	1	11	-	2	-	1	-	1	-	-	-	1
	2.2.	**7**	**8**	**9**	**9**	**3**		**10**		**1**	**5**	**1**	**-**
3. Government and society demands	**3.1.**	+			+	+					+	+	
	3.2.	**1**	**1**	**1**	**1**	**1**	**-**	**2**	**-**	**1**	**2**	**1**	**-**
	FoodNet	-	-	-	-	-	-	-	-	-	+	-	-
	EnergyNet	-	-	+	-	-	-	+	-	+	-	-	-
	AeroNet	+	+	-	+	+	-	+	-	-	-	+	-
	NeuroNet	-	-	-	-	-	-	-	-	-	+	-	-
	3.3. for 2014	n/d	1062025,8	33686863,3	n/d	n/d	15918116,2	n/d	n/d	n/d	n/d	n/d	31313132,2

[1] A filament light bulb can serve as an example, where in order to achieve more brightness temperature of the filament must be increased. However increasing temperature leads to melting of the filament. The solution is to use alternative technologies based on inert gases or LEDs.

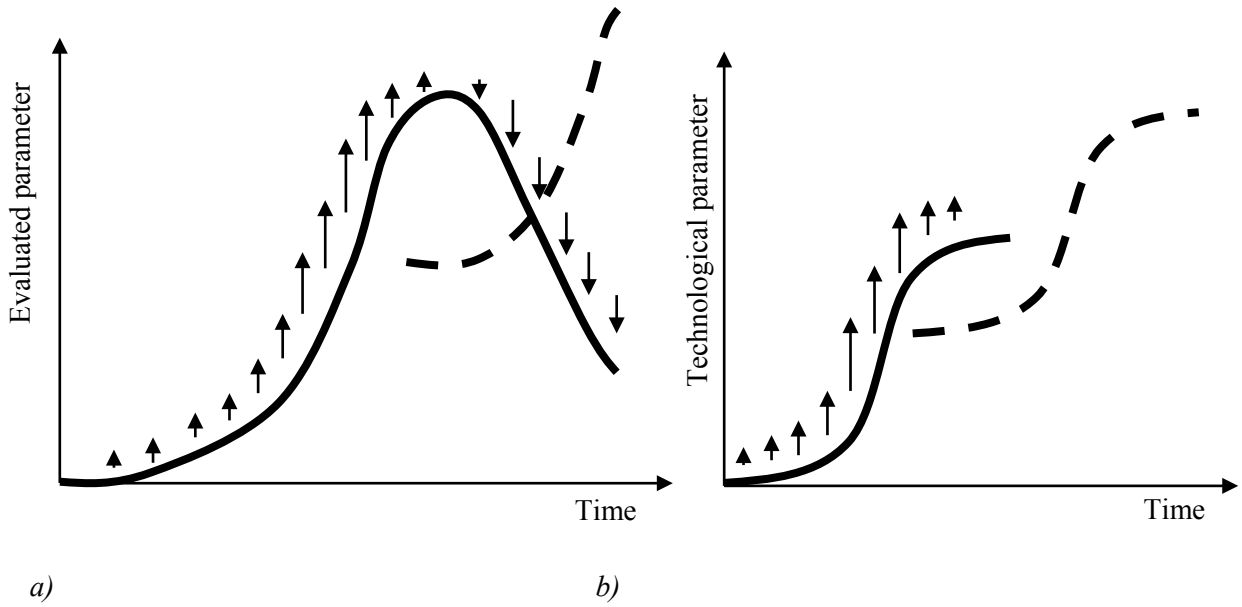


Fig. 3. Conceptual representation of innovation curve *a)* and S-shaped curve *b)* with indicated rate of change values (first derivative).

IV. CONCLUSION

In the process of choosing projects one must work with various parameters. The solution to the problem, which will allow making managerial decisions when managing the chosen projects, is to create a model that establishes a connection between the parameters that characterize the project and the environment in which the project is being implemented. This solution will allow to take the process from a formal sequence of actions to obtaining numerical indicators, as well as to consideration of an inverse problem of search for the parameter values that describe the project and based on those values the problem of search of a suitable environment for project implementation (in which desirable values of the target project function are obtainable (revenue, for instance)).

Research results implemented into products and technologies as a rule are inbuilt into product chains or production processes of big industry or complement products orientated towards future development of material existence, determined by the plans of corporations or governments using the foresight technology. Determining the form and direction of project implementation is only possible in cooperation between developers, big industry representatives, and forecasting experts.

REFERENCES

[1] Бессонов В.А. Проблемы анализа российской макроэкономической динамики переходного периода / В.А. Бессонов, М.: Институт экономики переходного периода, 2005..

[2] Amberg M., Mylnikov, L. Innovation project lifecycle prolongation method 2009. № Innovation and knowledge management in twin track economies: challenges & solutions (1-3). pp. 491–495.

[3] Murray A.S. The Wall Street journal essential guide to management: lasting lessons from the best leadership minds of our time / A.S. Murray, 1st ed-e изд., New York: Harper Business, 2010. 207 p.

[4] Mylnikov L.A. A system view of the problem of the modeling and control of production innovations // Scientific and Technical Information Processing. 2012. № 2 (39). pp. 93–106.

[5] Young H.P. The Evolution of Conventions // Econometrica. 1993. № 1 (61). C. 57.

Search of Method for Analyzing "Viability" of Innovative Projects

Aleksandr Rashidov[1], Werner Loch[2], Igor Shmidt[3]
[1,2]Hochschule Anhalt – Fachbereich Informatik und Sprachen,
Lohmannstraße 23, 06366, Köthen, Deutschland
[1,3]Perm National Research Polytechnic University - Electrotechnical Department,
Komsomolsky Ave. 29, 614990, Perm, Russia
E-mail: [1]santos.93@bk.ru, [2]w.loch@inf.hs-anhalt.de, [3]shmidt@msa.pstu.ac.ru

***Abstract*—Questions of "viability" evaluation of innovation projects are considered in this article. As a method of evaluation Hidden Markov Models are used. Problem of determining model parameters, which reproduce test data with highest accuracy are solving. For training the model statistical data on the implementation of innovative projects are used. Baum-Welch algorithm is used as a training algorithm.**

***Keywords:* Innovative project, Hidden Markov model, Baum-Welch algorithm, project viability.**

I. Introduction

Nowadays, there are many innovative projects in different fields of science. Some of them do not find the application and implementation.

Innovative projects is a complex systems of interdependent and interrelated resources, time limitations and persons aimed at achieving specific goals and objectives for the priority areas of science and technology [1].

The aim of this research is to determine the "viability" of innovative projects.

"Viability" in this study refers to the probability of innovative projects will be implemented.

The relevance of the study due to the following factors:

- A large number of innovative projects presented in the various programs;
- The need to determine parameters of the "ideal" innovative project, i.e. such a project, which will be implemented in the future for sure.

The definition of "viability" will give probabilistic predictions about the prospects for the implementation of this innovative project; it will assess the financial risks and other factors related to the innovative project.

The objectives of the study are:

- Description of the innovation project lifecycle;
- Search and analysis of necessary baseline data;
- Search of method for determining the "viability" of the innovation project;
- Assessing of possibility of using this method for the evaluation of "viability".

To evaluate the "viability" of these projects, data of the projects, which participate in programs "UMNIK" and "START", implemented by the Foundation for Assistance to Small Innovative Companies in the scientific and technical sphere are used. Thus, the "viability" can be defined as a likelihood that projects will pass all stages in "UMNIK" and "START" programs and subsequently will be successfully implemented.

II. Innovative Project as an Object of Study

The general concept of the innovation project and its life cycle is given in [2].

Main phases of the innovation project lifecycle include:

1) Pre-investment phase - performing of various studies for planning innovation project;
2) Investment phase - preparation of documentation for the implementation of an innovative project;
3) The phase of tenders and contracts - contracts for the supply of various equipment and performance of works are making;
4) Phase of the project implementation – plans and schedules of works on the project are developing;
5) The final phase of the project - innovative project closes, the launch of the facility and the start of implementation of the product, analyzes the results of the work.

As for the analysis are only available data on the participation of innovative projects in the Russian programs "UMNIK", "START" and "UMNIK to START", in this study under the lifecycle of the innovative project to be understood passage of the project cycle stages of participation in these programs and further implementation.

The "UMNIK" is a federal Russian program to promote the development of youth innovation projects [3]. Only individuals (program focuses mainly on young professionals with no experience in business) participate in the program. The program consists of several stages, presented in Fig. 1.

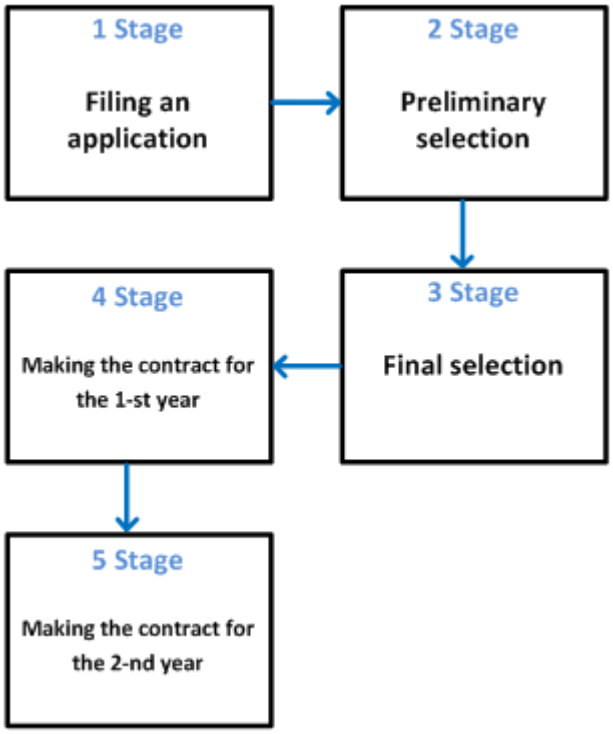


Fig. 1. Structure of the program "UMNIK".

To participate in the program one shall submit a request to the organizing committee, which is considered in several stages. Organizing Committee chooses from among the most promising applications and forwards them to the pre-selection. As part of the pre-selection of the applicant is a project in the internal report. As a result of reports of the commission shall decide on the admission of the participant to the next stage or reject the application. The next stage is the final selection, in which the reports are heard preliminary selection of the winners. According to the results of the final selection addressed the issue of providing funding. The winner is awarded a contract with a small innovative company for one year, during which the project is implemented. In the first, the Party shall provide a report on the results of which addressed the issue of the extension of funding for the second year.

The program "START" is a federal Russian program to promote the development of small innovative companies [4]. Participate in the program only to legal entities. The structure of the "START" program is presented in Fig. 2.

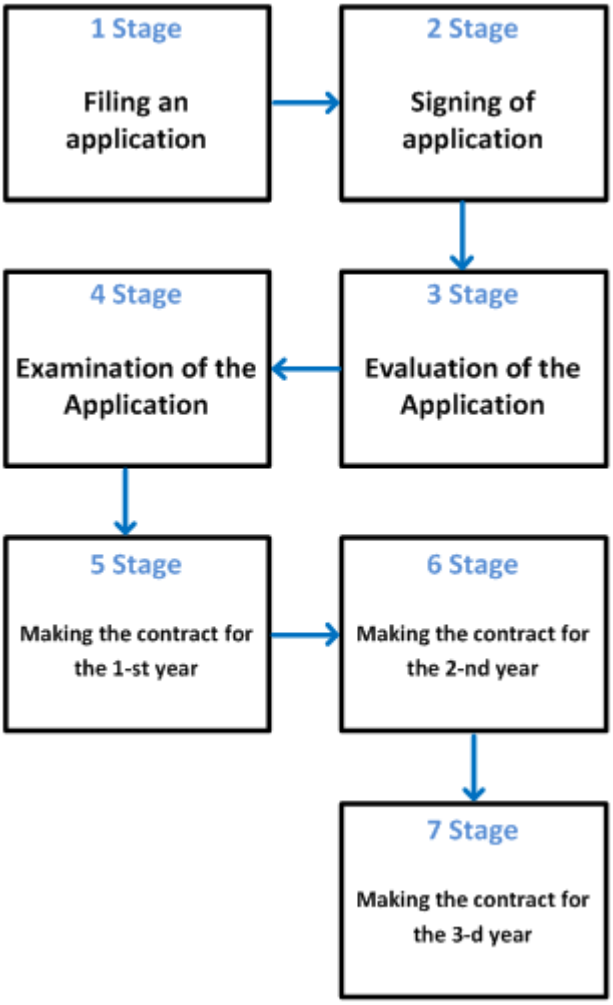


Fig. 2. Structure of the program "START".

To participate in the program one shall submit a request to the organizing committee. After that committee considers this request in several stages. Organizing Committee chooses from among the most promising applications and takes them into consideration. This application is no longer available for editing. The next step is the evaluation of the application according to different criteria, application requirements deviate inappropriate. As part of the examination of the application, the Committee shall select the most promising projects for financing, as well as the possibility of holding a full-time hearing, for a more detailed examination of the application. With the project, successfully passed examination of the application is a contract for one year, during which will be further development of the project. In the first year of funding decision is made to continue funding the project, as well as the party attracted extra-budgetary investment for the development of its project for the second year. The result of the second year of the project, a decision to continue or refuse funding for a third year.

The "UMNIK to START" is a youth innovation projects program, which allows the winners to participate in the "START" program [5]. Participate in the program winners of the program "UMNIK" who are willing to create a small innovative company to participate in the "START" program. Structure "UMNIK to START" program is presented in Fig. 3.

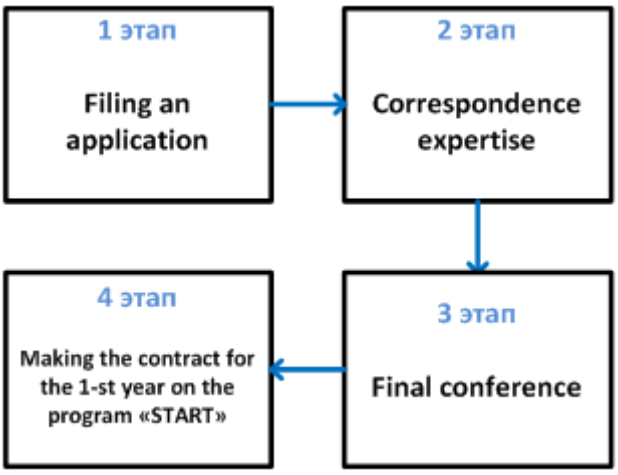


Fig. 3. Structure of the program "UMNIK to START".

To participate in the program "UMNIK to START" winners "UMNIK" programs submit an application to the Organizing Committee, considered at several stages. At the stage of the examination of correspondence to select the most promising innovative projects. The decisive step in this program is the final conference, which decision is the conclusion of a contract for the first year of the program "START".

III. Search of Method for Determining the "Viability" of the Innovative Project

Search and analysis of methods for assessing the probability of realization of innovative projects was conducted in the study. A result of analyzing method of hidden Markov models was chosen.

Markov process - random process in some system called a Markov, if the probability of transition of the system to a new state depends only on the state of the system at the moment and does not depend on when and how the system has passed in this state. More details Markov process is considered in [3][4].

Hidden Markov model - a model of the process in which the process is considered to Markov, the unknown, in which state S_i of the system (status hidden), but each state S_i can with some probability b_{ioj} produce an event o_j that can be seen. Hidden Markov model is described in [5][6].

Based on these determinations, it can be concluded that the process described above can be considered as a Markov process, because it contains a finite number of discrete states, the transitions between them take place at regular intervals, and for each possible transition to determine the probability that a constant transition.

Initially, it is possible to describe only the structure of the Markov model, transition probabilities are unknown. The transition probabilities can be found indirectly by using data on the implementation of the projects. Thus, can use a hidden Markov model to solve this problem.

IV. Search and Analysis of Training Data

It is necessary to have input data for training a model in this research. The input data were selected reports the Foundation for assistance development of small companies in the scientific and technical sphere for the years 2010-2013.

Data about total amount of projects participated in the program "UMNIK", amount of projects, which are gone out to the program START and amount of contracts are made on the program UMNIK to START are presented in the Table I.

Data about total amount of projects participated in the program "START" and contracts are made every year on this program are presented in the Table II.

Analysis of input data takes place on various parameters, such as:

- The number of submitted and accepted applications for participation in the programs;
- Funding for each program ("UMNIK", "UMNIK to START" and "START");
- The number of applications submitted for each specific topic;
- Features the most successful projects.

The selected data is suited to train the model, but they are not enough for obtaining high-quality results and answer the question about the "viability" of specific projects.

For obtaining of more adequate results need to be introduce additional parameters, such as:

- The implementation of cost;
- The complexity of implementation (Number of people and the number of hours to be spent on project);
- Significance (Evaluation the significance and necessity of the project for the real work).

In finished form, such data cannot be obtained since a part of them is intellectual property, and the other part is not calculated on a straight statistical authorities. In this regard, it is necessary to calculate this data based on indirect indicators (e.g., macro parameters).

TABLE I
Statistical data about projects and contracts of program "UMNIK" during the period 2007-2013

Program UMNIK	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Amount of participants	804	1037	1334	1533	1910	1841	1986
Project, which are gone over to the program START	43	48	50	65	61	68	63
Amount of contracts made on the program UMNIK to START	-	-	-	-	-	52	68

TABLE II
Statistical data about projects and contracts of program "START" during the period 2007-2013

Program "START"	2007	2008	2009	2010	2011	2012	2013
Amount of applications	1430	1480	1533	2068	1812	1938	1639
Making the contract for the first year	380	400	450	488	537	495	499
Making the contract for the second year	82	77	63	143	165	177	181
Making the contract for the third year	17	31	27	33	46	71	70

V. Assessing of the possibility of applying the method of hidden Markov models for the evaluation of "viability"

Usually, when working with a hidden Markov model, there are three tasks.

1) Definition of probability of occurrence of a sequence of observations $O = o_1, o_2, ..., o_T$ model $\lambda = (A, B, \Pi)$.

2) Selection of chain state $Q = q_1, q_2, ..., q_T$ which best fits the available observation sequence O, for given λ and O.

3) Determination of the model parameters $\lambda = (A, B, \Pi)$, ensuring maximum probability $P[O|\lambda]$.

More detailed given tasks described in [6][7].

In this research, solved the third problem. To solve this problem requires a sequence of input data. Structure of Markov model must be known. It is necessary to determine the initial probability of each state and transition probabilities between states in this model.

In this research use the following structure Markov model based on the input data structure (Fig. 4).

The states described in the model, based on the stage of the program, described in Section 1.

The program "UMNIK" - states 1-8;

The program "UMNIK to START" - states 9-12;

The program "START" - states 13-22.

States of program "UMNIK":

1 – Filling an application and appeal to the regional government;

2 – The deviation of the application by the regional government;

3 – Preliminary selection of applications (presentation);

4 – Final selection;

5 – Conclusion of an employment contract with a small innovative company for 1 year (financing);

6 – Termination of R & D funding;

7 – Continuation of R & D funding for 2 years;

8 – Continuation of work on the project.

States of program "UMNIK to START":

9 – Filing an application;

10 – Extramural expertise;

11 – Deviation of application;

12 – Final conference.

States of program "START":

13 – Filing an application;

14 – Evaluation of application;

15 – Deviation of application;

16 – Examination of application;

17 – Making the contract for the first year;

18 – Presenting a project;

19 – Making the contract for the second year;

20 – Termination of the contract;

21 – Making the contract for the third year;

22 – Continuation of the work small innovative companies.

To solve this problem use the Baum-Welch algorithm, similar to the EM-algorithm. Description EM-algorithm presented in [8][9]. Detailed description of Baum-Welch algorithm is given in [6][10].

Brief overview of the Baum-Welch algorithm:

For given a set of states: $S=\{S_1,S_2,\ldots,S_N\}$ and a set of observation symbols: $V=\{v_1,v_2,\ldots,v_M\}$, model $\lambda=(P,O,\pi_1)$ is defined, where

$P=[p_{i,j}]$ - Probabilities of state transition;

$O=[o_j(m)]$ - Probabilities of observation;

$\pi_1=(\pi_{1,i})$ - Initial state probabilities.

Algorithm consists of two steps:

At the first step from the known parameters of the model probability of being at time t in state S_i ($\gamma_t(i)$) and probability of being at time t in state S_i and at time t+1 in state S_j given the observation sequence and the model ($\xi_t(i,j)$) are calculated using following formulas:

$$\xi_t(i,j)=P(q_t=S_i,q_t+1=S_j|O,\lambda)$$

$$\gamma_t(i)=\Sigma\ \xi_t(i,j)$$

At the second step model λ are compute based on values $\gamma_t(i)$, $\xi_t(i,j)$ then re-computing of similarity given and calculated λ are performed.

Then the first step is repeated using the calculated. For multiple repetition of this procedure, probability of coincidence the calculation of function with the original one increases. The algorithm converges in the moment when the probability stops changing.

In this research:

Set of states: $S=\{S_1,S_2,\ldots,S_N\}$ – the state of the program shown in Fig. 4;

Set of observation symbols: $V=\{v_1,v_2,\ldots,v_M\}$ – known data on innovative projects (the number of projects, contracts are made, etc.);

Model $\lambda=(P,O,\pi_1)$.

It is necessary to determine: probabilities of state transition, probabilities of observation and initial state probabilities.

These parameters will be calculated in the future.

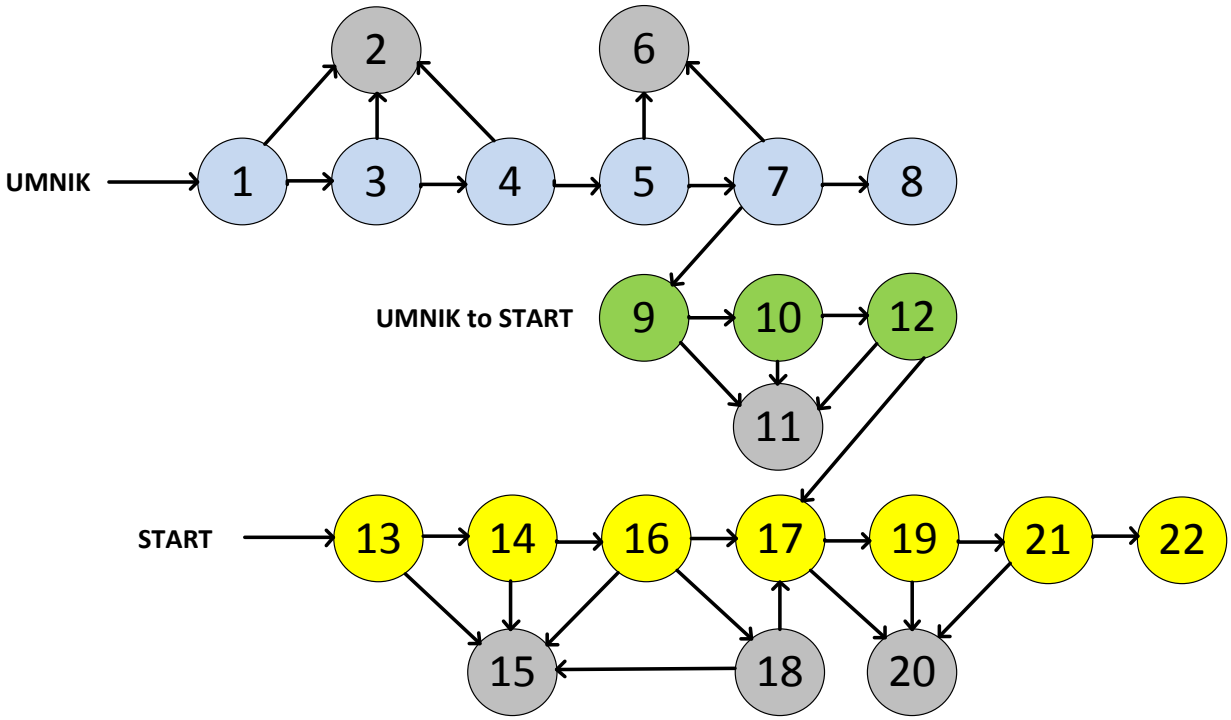


Fig. 4 - The structure of the Markov model.

VI. Conclusion

Analysis of method of innovative projects "viability" evaluation was carried out in the research. It was found that the method of hidden Markov models is suitable for solving this problem, in particular, Baum-Welch algorithm can be used for this task. This algorithm has been briefly described, the assumptions and constraints was considered.

In future it is planned to prepare of test data based on information about the functioning of the federal program to promote the development of innovative projects. Estimation of the transition probability between the phases of the programs presented in the Markov model will be implemented based on these data. Statistics for several years will be used for determining the average value of the transition probability between states described in the model. Based on this, transition probability deviation corridors for the following measurements will be found. The corridor, in which the minimum and maximum values of the transition probability for the projects, moving through stages of implementation, are laid, will be defined using the Baum-Welch algorithm.

References

[1] Basics of Innovation Management: Theory and practice: tutorial / Edited by. P.N. Zavalina and al. – M.: JSC «NPO «Publishing house « Economy », 2000. – 475p. (In Russian)

[2] Adizes I. Corporate lifecycles: how and why corporations grow and die and what to do about it / I. Adizes, Englewood Cliffs, N.J.: Prentice Hall, 1988

[3] Russian federal program "UMNIK". [Online]. Available: http://www.fasie.ru/programs/programma-umnik/

[4] Russian federal program "START". [Online]. Available: http://www.fasie.ru/programs/programma-start/

[5] Russian federal program "UMNIK to START". [Online]. Available: http://www.nauka-nov.ru/fsrmpnts/665/

[6] Portenko N.I., Skorohod A.V., Shurenkov V.M. Markov processes / Portenko N.I., Skorohod A.V., Shurenkov V.M. - : , 1989. - 765 p. (In Russian)

[7] Dynkin E.B. Foundations of the theory of Markov processes / Dynkin E.B. - Moscow: FIZMALIT, 1992. - 228 p. (In Russian)

[8] Elms A.J., Procter S., Illingworth J. The advantage of using an HMM-based approach for faxed word recognition// International Journal on Document Analysis and Recognition (1998) 1: 18–36

[9] Rabiner L.P. Hidden Markov models and their use in favorites applications in speech recognition: Overview//TIIER, т.77, N2, february 1989 — p. 86–120.

[10] Christopher M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics). 2006

[11] Geoffrey McLachlan and Thriyambakam Krishnan. The EM Algorithm and Extensions. John Wiley & Sons, New York, 1996.

[12] Jeff A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical Report TR-97-021, U. C. Berkeley, April 1998.

[13] Welch L. Hidden Markov Models and the Baum-Welch algorithm // IEEE Information Theory Society Newsletter, 2003.

System Analysis of Innovative Projects in Mechanical Engineering

Stepan Mezhov
Altai State University−Management and Information Systems
Lenin Ave., 61, PO 656049, Barnaul, Russia
E-mail: megoff@mail.ru

***Abstract.*—Modern production and innovation processes in industrial corporations differ by complex nature of modern technologies and organizational solutions. Adequate methodological approaches to their analysis and evaluation of efficiency are required. The present study contains the concept and the model of operational and innovation program for which formulated theoretical and methodological principles and systematic statement of the general problem of planning, designed to optimize the process of resource usage under different criteria, including: investment, financial, technological, industrial and information capabilities of the corporation.**

Keywords: **Operational and innovation processes of the corporation, the production program of the industrial corporation, innovations, product life cycle, modeling.**

I. Introduction

It is not a secret that the issues of building an innovative economy are still difficult to study as at the present time observed a violation of systemic organizational development principles, proportions and balance between the internal and external environment, there is virtually no theoretical and methodological framework for solving such problems, while the successful experience of western corporations is not available for Russian enterprises.

Innovations, as a basis for organizational development and competitiveness of the enterprise, require constant investments, the sources of which are in the area of operating activity and certainly have a significant impact on its effectiveness. The reduction in current financing worsens profitability and efficiency ratiosin the reporting period, the decrease in financing of innovationsharms strategic effectiveness of the corporation.

In varying degrees of completeness, this problem is solved in most western corporations which successfully combine the tactical and strategic aspects of the activity.

It must be emphasized that at the present time there is a certain contradiction between the scope of theoretical and applied research on the formation of the methods and mechanisms to improve the management of modern corporations and the lack of innovative corporations in the Russian economy.

On the basis of above-mentioned it can be concluded that the scientific problem, connected with the development of methodology for planning the operational and innovation activities of modern industrial corporations, is considered extremely important. It appears that the solution of this problem shall provide a set of methods and tools for better planning of operational and innovation processes and evaluate the effectiveness of investments in innovation projects with due regard to innovative capacity of the enterprise.

Methodological aspect of knowledge of the system or the process involves the construction of a more or less adequate model allowing to significantly expand the range of research and get the structural, historical, logical and functional data, which for various reasons are not available during the direct study of the object. In addition, these data are necessary for the constructive solution to the organizational and managerial problems of corporate governance, in particular related to innovations.

In order to solve scientific problems indicated in the present study used a systemic methodology, which allows to rely on different approaches and concepts in forming the models of the object. The economic-mathematical methods and models, including methods of correlation, regression, factor analysis, methods of expert assessments and others are also used.

II. Results

The main new scientific results are the following programs significant for the development of science and practice:

1. The methodological approach to corporation planning that develops existing approaches on the basis of a set of optimization and simulation models, taking into account the balance of operational and innovation processes based on formal criteria. In contrast to existing approaches (Pleshinsky, 2004) (Sobolev, 2000) (Titov, &Mezhov, 2013) (Titov, 2007), (Danilin, 2006) the production program is complemented by the optimal strategy for implementation of a new product, as well as evaluation of investment strategies and changes of innovative capacity of the enterprise.

2. The model of planning of the operational and innovation program iteratively generates many optimization options, implementing the demand forecast based on the life cycles using the simulation model. It also complements and develops existing approaches to the planning of the production program (Sobolev, 2000) (Titov, &Mezhov,

2013), (Danilin, 2006) by way that this model is implemented in a dynamic form, explicitly takes into account the time factor and the set of transformation strategies of innovation potential.

III. DISCUSSION

1. The formulation of tasks of the planning process of the operational and innovation program of the industrial corporation.

Any strong corporation can only be viewed from the standpoint of the unity of the operational and innovation activity.

The enterprise finances the production activity for each business process (E_i□BP), requires the necessary resources (R_1, …,R_n), processing some of them, produces the output equal to(X_1, …, X_n). Then the enterprise sells it and receives an income, the part of which is intended for the formation of investment sources. From these sources formed a capital for financing the operational activity (I_o), i.e. for restoring E_1, ...$E_n$and financing the innovations (I_n). Raising debt capital in the event of the owned capital is not excluded. Thus, the total invested capital can be determined as $I=I_o+ I_n$.

According to our research, there are several approaches and tasks, interdependent by these approaches, that are most consistently studied by many authors.

Separate economic and mathematical models or their groups (systems) for the development of advanced, current and operational schedules and production programs, are currently developed, experimentally tested and partially implemented into planning practice of operational and innovation processes in companies.

Most of the considered models of prospective and current planning in the mathematical sense are considered as problems of mathematical programming. In planning practice making decisions is usually based on the need to achieve multiple targets, i.e. the solution is to be evaluated simultaneously according to several criteria, but the canonical formulation of tasks of mathematical programming requires a one-criterion approach, that certainly limits its extensive use in the planning practice.

A rather constructive approach to forecasting the production programs of the corporation development on the basis of economic-mathematical tools is a system of optimization, partly integer models, widely described in the works (Pleschinsky, 2004), (Sobolev, 2000), (Titov, &Mezhov, 2013), (Titov, 2007). In these models considered only part of the product life cycle related to the production and sale of products, business and financial processesare widely represented, but there is no planning process of research and development of new products, technologies (R&D).

One more constructive approach to the planning process of production programs of enterprises in terms of innovations on the basis of economic and mathematical models is an approach developed in the works of (Danilin, 2006). As is known, the frequency of stages of development, pre-production, release and use of the product since the stage of theoretical studies and prior to removing the product from the production is called the product lifecycle. In other words this is the period from the design of a product to the termination of its sales. In the product life cycle it can be indicated the following consolidated stages: research scientific work, design and experimental preparation, process-layout preparation, the launch of a preproduction run, the production growth, stabilization of production volumes, the decline in the product production and phase-out.

According to these approaches, modeling allows to determine an expected year of the launch of a new product, taking into account the economic conditions of production for the company as a whole. A demand for products of each type, sales volume, net total cash flow with due regard to savings for each year, the complexity of the program and other factors are used as restrictions; maximizing the net present (current) value and other factors are used as a criterion. The forecast based on the optimization models can be represented as a single-point discrete process in which the parameters of the environment require the special evaluation.

The present studies (Mezhov, 2012), (Mezhov, 2011), (Mezhov, 2014) show that such a formulation of the problem is noturgent for today's most advanced corporations. The works (Bogoviz, &Mezhov, 2014), (Kondratyev, 2003), (Prahalad, 1996), (Chesbrough, 2006), (Boutellier, R., 2008), (Gassmann, 2006) show that the replacement of products in the production program is entirely dictated by the market demand and the principles of business competition. However, the model scheme of product life cycles remains an important methodological promise to build a comprehensive operational and innovation program. We shall discuss it further.

In an instrumental relation the task of formation of such a program shall be based on optimization models, that is natural when optimizing the resource utilization, on forecasting models, as occur the cash flow forecast and a number of characteristics for investment appraisal, and on imitation models, which are necessary for the assessment of options for planning of production, sales and investment policies.

Assumptions and basic conceptual distinctions of the proposed formulation of the task from the conventional approaches to planning of production and sales plans:

1. A market enterprise independently decides on all matters of innovation policy on the basis of the entrepreneurial initiative.

2. An enterprise has the ability to carry out R&D and introduction of new products on the basis of its own innovative capacity.

3. An enterprise carries out research and development of new products on a regular basis, rather than on the basis of discrete projects adopted at the time of the product start-up, i.e. provides the whole range of search and research works, including the scheme of open innovations (it integrates with partners), optimizing investments between the production and innovations.

4. The units being planned have all the necessary competencies and knowledge of the main trends of development of the industry, therefore, well ahead of the product launch determined the perspective concepts that are further brought to production. Concepts of the products

brought to the innovation process are considered as accepted projects that consistently brought to production and sales.

5. The total investment volume aimed at complete development of a new product shall be recovered through sale and bring the industry average profit and rent in accordance with the strategic plan of the corporation.

6. It is obvious that the requirement to the specified payback is entirely due to the innovative potential of enterprises and suggests its account in the formal procedures for evaluating the effectiveness of investments allocated in innovations.

7. Since the concept of evaluating the effectiveness and payback of the investments (NPV) requires forecasting the cash flows on a sufficiently long time interval, then itshall necessarily be reflected in the general model of the program formation.

8. The innovative potential of an enterprise affects the effectiveness of in-vestments, therefore, it is necessary to take this fact into account in the traditional procedures for evaluating the investments.

9. Since the main trends of the production of each product are specified by the product life cycles, while the output also depends on the demand, actually the life cycle curve defines an average statistical trajectory, around which varies the value of the random variable of the "demand" for the product for each year, statistically obeying the life cycle curve.

10. Taking into account the fact that the market power of any corporation is determined by the market share, the sales plan shall be guided by the maximum output of products with limited resources.

We propose to solve these tasks using models, system methods, including imitation and optimization modeling that shall allow to reveal the quantitative interaction of investments, innovations, sources of funding, that can lead to making well-founded strategic decisions (Mezhov, et al., 2014).

The purpose and nature of the model of planning the production of products in conditions of constant investment spent on innovations can be stated as follows.

Knowing the forecast production cycle of each product: the dynamics of output and sales volume in volume terms, production costs, labor intensity, etc. on intervals of the pro forma cycle, the volume and pace of investment and possible intervals of timing of the products launch, current and strategic development indicators, it is necessary to define such terms of the launch of new products in order to perform the tasks in terms of growth of the value of the corporation, provide the main economic and financial indicators of effectiveness and sustainability with the resources allocated to the corporate budget, including borrowed capital. At the same time it is necessary to determine the optimal volume of innovation investment, observing, on the one hand, the balance between profitability and liquidity, on the other hand, the optimal duration of the start of sales of new products, in order to maximize rents from possessing unique expertise.

Optimal volumes of innovation financing are defined by strategies for R & D implementation. These strategies are considered as logically related, consistent volumes of investment of each stage of the innovation process, as well as the rate of work production and release of new products onto the market. At the same time there occurs the following dilemma: the acceleration of the innovation process requires large amounts of funding, that definitely reduces the current performance of the corporation, but early marketing of new products shall provide an additional income in the form of Schumpeterian rents. In the framework of **operational and innovation program** (OIP) the task of balancing the volume, rate and payback of innovation investment shall be solved.

2. Description of the subject area.

An industrial corporation has a high potential for innovations, on a regular basis it manufactures, sells products with known product life cycles (projected market demand) and carries out the research to develop new advanced products and technologies.

The integral numerical value of innovation potential can be determined by the expert-analytical methods, for example, by the method of analysis of hierarchies. From our point of view, an expected potential is due to the high level of organization and management.

In the framework of the strategic plan for solution of various problems the corporation allocates a certain investment volume I, which is specified in the tactical plan created for solution of the current tasks. t indicates the number of periods which a tactical plan is formed for, containing production and innovation component, I_t indicates the investment volume allocated to this plan. Then:

$$I = \sum_{t=1}^{T} I_t ,$$

where T is the number of the planning periods (years).

In turn, the volume of the current investment may consist of internal and borrowed funds. In this case, we assume that the borrowed funds are used only to finance current operations, financing of innovation is carried out at the expense of internal funds, major investments in upgrading technology for new products can be carried out at the expense of internal funds and long-term loans.

Let the corporation produce J kinds of products, then $x_j, j = 1,2 \ldots J$ shall be the output of the j^{th} product.

Let us denote by φ_{ni} the set of indices of the products i, $i = 1, 2, \ldots N$, the set of potential products (development on new products) located in the innovative development, then I_{nit} investments shall be spent on each designed product, so that:

$$I_{nt} = \sum_{i=1}^{N} I_{nit}$$

After the start of mass production the investment in development of a new product shall be terminated and have only production costs.

In this context, the formation of the model of the production program shall be based on the concept of the product lifecycles. Based on research of retrospective statistics on the economic product life cycles of separate industrial products, and after their appropriate processing with the use of mathematical models appears the possibility to forecast the duration of the production of a new product with an acceptable accuracy, taking into account the time from the time the product is launched till its withdrawal from the market.

We can get the curve of the product life cycle, reflecting the stage of production, if using the method of statistical analysis of forecasting the various technical and economic indicators of production depending on the prevailing trends of output volume by years of the accounting period. In the framework of the present study it should be considered that the product life cycles are determined by the marketing department and are exogenous variables specified by an external source.

We denote the lifecycle of the *j* product as G_j, j =1,2… J, then J products can have various product life cycles as a continuous time function,

$$G_j = G_j(t) \quad (1)$$

Or as a discrete, tabular form of a task, in the form of compliance:

$G_j^i - t_j^i$; j = 1,2… J; i = 1,2…π, i – numbers of time periods,

π – a number of discrete partitions of the table (a number of years).

In fact it is possible to have both the analytical dependence, for instance, sales volume depends on the time, and a tabular dependence with the corresponding derived indicators, such as income, expenses, etc. All these indicators can be represented both in value and natural forms.

If for the construction of the analytic dependence used regression analysis, the curve $M(G_j(t))$ shall represent an average value of the fluctuations in sales over time, while in the framework of regression analysis determined the standard deviation of sales $-\sigma_j(t_k)$using Monte Carlo modeling method, where t_k means the time of forecasting the volume of demand. Thus, $\sigma_j(t_k)$ shall be different for different time intervals, which is natural, since this standard deviation reflects the average spread of demand in different time intervals of forecasting. Obviously, the acceptable value of sales in the planning, according to the law of three sigma, may lie with high probability in the following interval:

$$M\left(G_j(t)\right) - 2\sigma_j(t_k) \le G_j\ (t) \le M\left(G_j(t)\right) + 2\sigma_j(t_k). \quad (2)$$

This interval shows that during modeling it shall be considered the probable deviation of sales volume from its average value (short-term forecast), which is defined by the product life cycle curve (long-term forecast). With a steady drop in demand for the product, the separate increase or accelerated decline in sales volume, the management shall not be misleading, since the main trend shall be determined by the curve $M(G_j(t))$. Such a shape of the curve can be represented by the power polynomial (parabola) in procedures for regression analysis, or in a tabular form, which can be more convenient for practical use in the planning of the program.

In fact, it is possible to form a production program with the use of linear models. However, the features of the subject area, described above, require a few other approaches to the planning.

Recall that, in fact, in the formation of production volume for each product using the model it is necessary to optimize not the volume that we know roughly in the projected life cycle, but the number of important indicators conditioned by the efficient use of available resources with certain interval estimates (equation (1)) of the production volume in volume terms. In this case the optimization makes sense because only the upper limits of volumes are set. There are other problems, namely, within the structure of demand (rolling short-term forecast) and long-term forecasting of product life cycles for each product it should be formed a production program and an innovation investment program, which shall provide the replacement of products, derived from production. Then we can formulate the problem of the OIP formation as follows:

In the presence of available life cycles of all products: $G_j = G_j(t), j = 1,2,\ldots, J$; planning time, or at discrete consideration, specifying the demand forecast (2), it is necessary to find an optimal production schedule in physical indicators at the time t:

$$x_j^t \le G_j^*(t);\ x_j^t \in X_t;\ t = 1,2,\ldots,T;\ j = 1,2,\ldots,J, \quad (3)$$

$$x_{j*}^t = (1-\alpha)\, x_{j*}^{t1};\ x_{nj*}^t = \alpha x_{nj*}^{t1}. \quad (4)$$

where X_t means an optimal plan, x_j^t means nomenclature plan components, α means a control parameter, the share of new products to replace the old ones, with α =0, a new product is not included in the program; when α = 1, an old product is completely excluded from the program, with 0 <α<1 in the program included both products in the proportions shown in the equation (4); x_{j*}^{t1} means a product being excluded from the program; x_{nj*}^{t1} means a new product launched into production; t_1 means the time of the product replacement.

The complexity of the procedure for determining the optimal plan X_t with due regard to the mechanism of replacement of an old product onto a new product (4) is conditioned by the need to taken into account the following:

- lifecycles on each product defining the upper limits of the output;
- production capacity that determines the limiting boundaries of the output: if the demand forecast for a product is less than the output capacity on the same product, i.e. $G_j(t) < X_j^m$, then $x_j^t \le Gj,(t)$, and if $Gj,(t) > X_j^m$, then $x_j^t \le X_j^m$, it means that when demand exceeds the capacity, the value of an expected volume does not exceed the capacity, if it does not provide an additional product launch;
- options of the strategy used for replacement of an old product onto new ones, for instance, replacement can be "slow", "fast", "compromise", when for some time both products are included in the plan;
- options of investment with or without an improvement of the innovative capacity, etc.

The sales volume for the period t is determined as follows:

$$U_t = \sum_{j=1}^{J}[c_j\ x_j + c_{j*}(1-\alpha)x_{j*}^{t1} + c_{nj*}\alpha x_{nj*}^{t1}] \quad (5)$$

where c_{nj*} means the price for a new product, c_{j*} means the price for a product excluded from the program, *nj** means the index of a new product.

The operational and innovation program shall reflect the growth of the key performance indicators that determine the long-term development of the enterprise, and at the same time its financial stability in the short term. Provision of financial equilibrium in the process of enterprise development is a system of measures aimed at maintaining the mechanism of the balanced growth on basic economic indicators in the long run.

The development model of the enterprise leads us to the issue of financial support of the enterprise`s growth. For the growth to be sustainable, the operational and financial policies and growth targets shall be agreed. Such growth is usually expressed in terms of an increase in internal funds.

The work (Higgins, 1998) clearly shows that the achievement of an expected growth rate depends on the applicable financial policies. According to (Higgins, 1998) the growth rate:

$$g = r_{pr} \times d_p \times o_a \times k_{eq} \quad (6)$$

where r_{pr}– profitability of sales, d_p – the share of undistributed net profits of the enterprise spent on re-investment, o_a – asset turnover ratio, k_{eq} – asset/equity ratio (financial leverage).

With the imbalance in periods of the growth of internal funds and the sales growth, the model of J. Van Horn (Van Horne, 2009) is proposed. In this case, it is proposed to introduce into the model the initial sales volumes U_0, the initial size of the internal funds K_c, the absolute amount of dividends *Div*, which the company is planning to pay, as well as the volume of issue of ordinary shares:

$$sq = \frac{(K_c+\Delta K_c-Div)(1+\frac{K_3}{K_c})\times\frac{U}{A}\times\frac{1}{U_0}}{1-[r_u\times(1+\frac{K_3}{K_c})\times\frac{U}{A}]} - 1,$$

where *sq* –the level of attainable growth, ΔK_c – capital increase for the next period.

The Higgins model is focused on the growth, mainly due to its internal funds, while the J. Van Horn modelis focused on the growth due toborrowed funds. The problems of innovation development of Russian corporations involve high risk and complexity of capital raising, both in the institutional and financial aspects. Therefore, the Higgins model is more suitable for modeling the growth of Russian corporations.

3. The forecasting model of the production program.

For the formation of a functional and a system of limitations of the production program in the OIP structure an approach described in the scientific work (Mezhov, 2012) can be used as an instrumental basis.

This approach is based on the use of linear or non-linear functions as models of sales and costs, these functions can be constructed on the basis of the regression analysis on retrospective data of the company for a certain period of time. If such functions are constructed for each product, their accumulation can serve as a forecasting model for the production program of the enterprise. At the correct construction of the functions of sales and costs, when the fixed and variable costs are clearly indicated, the interaction of these functions adequately reflects the economic model of the enterprise.

The nature of an actual behavior of the functions of production and costs allows to accurately describe the curves f1 and f2 using the second order polynomials by the regression analysis in the following form:

$$Y=\alpha_1x^2+\alpha_2x+\alpha_3.$$

Then, using an appropriate definition of the coefficients, occurs the following equation:

$$f_1=a_1x^2+b_1x+q_1, \quad (7)$$

$$f_2=a_2x^2+b_2x+q_2. \quad (8)$$

In this case the functions of profitability can be found:

$$f_4=(a_1-a_2)x^2+(b_1-b_2)x+(q_1-q_2). \quad (9)$$

If we define $a_4=a_2-a_1$; $b_4=b_2-b_1$; $q_4=(q_1-q_2)$, as$q_1\to0$, while q_2 is significantly greater than 0, as it shows the share of fixed costs, then $q_4<0$:

then:

$$f_4=a_4x^2+b_4x-q_4. \quad (10)$$

Thus an informative (econometric) meaning of the parameters a_k, b_k and q_k areas follows: a_1–adjusting factor with dimension equal to$\frac{1}{pcs^2}$; b_1– price of a product; c_1–regressionerror, at the correct presenting of $f_1, q_1 \to 0$; a_2–adjusting factor of dimension costs $\frac{1}{pcs^2}$; b_2–variable costs per unit; q_2–fixed costs per unit, at the same time$\sum_i^J q_{2i} = s_{cnst}$, here it is important to emphasize that s_{cnst}means fixed costs for the whole OIP; a_4 – means adjusting factor of profitability from the production volume with dimension equal to $\frac{1}{pcs^2}$; b_4–account of the fixed costs in the profit calculation.

Then using a non-linear model of the economic production processes, the information obtained using this model, it can be constructed the optimization model of the production program.

The function, objectives, sales volume:

$$U_t = \sum_{j=1}^{J}[c_j\ x_j + c_{j*}(1-\alpha)x_{j*}^{t1} + c_{nj*}\alpha x_{nj*}^{t1}] \to max \quad (11)$$

$$U_t = \sum_{j=1}^{J}[c_j\ (\varphi_j, t)\, x_j + c_{j*}(\varphi_{j*}, t)(1-\alpha)x_{j*}^{t1} + c_{nj*}(\varphi_{nj*}, t)\alpha x_{nj*}^{t1}] \to max,$$

where φ_j, φ_{j*}, φ_{nj*} are the elements of the set of controls developed by the planning of the corporation group, for improvement of the effectiveness of products sales on the market. The use of controls allows, in particular in the automatic mode, to adjust the prices of products in the objective function. The *F* set contains the options to adjust the system of constraints of the optimization model and the objective functions in the automatic mode (Mezhov, 2012). The controls are considered as possible sets of prices, technological coefficients of the matrix of restrictions, all possible combinations of aggregation (for example, 12 types of nomenclature positions, the types No. 2, 5, 7 form a restriction on the joint production output, as well as the right parts of the restrictions can change due to changes in separate types of resources or their prices, etc.).

All possible controlsf_1 and f_2, or deformations, in general, lead to either an increase or a decrease in the area of the break-even work, in particular, the function$f_i$can have the following controls:

- an increase in sales volume due to an increase in output at constant prices;
- an increase in sales volume due to the price increase at a constant volume of production;
- a decrease in sales volume by lowering the prices;
- an increase in sales volume by reducing the price and increasing the demand;
- changes in sales volume due to minimization of receivables.

Formally all these controls lead to the fact that the initial curve f_1 is subjected to the deformation by the control operator, i.e.

$$f_1\mathrm{d} = \varphi f_1,$$

where φ is a control operator, which can be defined by a constant, function, procedure.

For example, it is necessary to reflect in the model an increase in sales volume by increasing the original price at 1.2 times, therefore, we shall obtain $f_1 d = 1.2 f_1$. Or we can suppose there is a situation when an increase in the volume of production changes the sales volume in proportion to the linear function F = φ1, then f_1d= *(*φ1*)*f_1.

Similarly as the cost function $f_2$is concerned. The controls can implement the following real-life situations related to the actual events in the deformation process of the set of technological choice:

- increase or decrease in the total costs: $f_2 d = w_1 f_2$;
- increase or decrease in the fixed costs: $f_2 d = f_2 p + w_2 f_3$;
- increase or decrease in the variable costs: $f_2 d = w_3 f_2 p + f_3$;
- various combinations of changes in costs.

TABLE 1
PARAMETERS OF FUNCTIONS OF THE MODEL OF THE PRODUCTION PROGRAM

No.	**Product description**	***a***	***b***	***q***	
1	Product 1	$f_1 x_1$	a_{11}	b_{11}	q_{11}
		$f_2 x_1$	a_{21}	b_{21}	q_{21}
		$f_4 x_1$	a_{41}	b_{41}	q_{41}
J	Product *j*	$f_1 x_j$	a_{1J}	b_{1J}	q_{1J}
		$f_2 x_j$	a_{2j}	b_{2j}	q_{2j}
		$f_4 x_J$	a_{4j}	b_{4j}	q_{4j}

After the formation of Table 1, using derived relations of functions of the model, the initial parameters of the production program have been calculated.

Limitation on profits:

$$Pv = \sum_j^J f_4 Xj = \sum_j^J (a_{4j} x_j^2 + b_{4j} x_j - q_{4j}) + (a^n_{4j*} x^2_{j*} + b^n_{4j*} x_{j*} - q^n_{4j*}) \geq P^n_v, \quad (12)$$

where P^n_v –means normative value of profits, a^n_{4j*}, b^n_{4j*}, q^n_{4j*}- coefficients of the profit functions for a new product.

Intheequation$f_4$we deliberately indicate the share of the profits generated from the production and sales of a new product, since it is necessary to show the impact on profit of excluded and implemented products, as well as the interaction between the total investment volume and the rate of the launch of new products.

Limitation on investments in OIP:

$$\sum_j^J (a_{2j} x_j^2 + b_{2j} x_j + q_{2j}) \leq I_o, \quad (13)$$

where $I_o$means a normative (planned) value investments in the production program, the left part of the equation (13) shows the total estimated investments in the production program.

The regulatory or given value of the total investment volume, as it has previously been shown, consists of two parts: investment allocated on production and innovations:

$$I = I_o + I_n , \quad (14)$$

The equation (14) can be written as follows:

$$I = \lambda I + (1 - \lambda) I, \quad (15)$$

λ indicates the share of investments in production, while (1-λ) indicates the share of investment in innovation, i.e.

$$I_o = \lambda I, I_n = (1 - \lambda) I \quad (16)$$

Accounting for λ in the structure of investment allows to use it as a control parameter in the OIP optimization.

4. The general algorithm for the modeling of operational and innovation program

1. On the basis of retrospective and analytical data of the enterprise we can form a file of life cycles of production of main products, if necessary the procedure for the aggregation of similar nomenclature items can be carried out.

2. Dependencies on the planning time of sales (demand) for each product, based on marketing data, can be combined in the tabular form, additionally including in it the standard deviation, also for each product. Obviously, the forecasting of the random variable of "demand" based on the life cycle curve is due to the presence of this value in the interval, for example, according to the equation (2).

3. For each product *j* we form the functions of costs and sales in accordance with Table 1. If we present *f1, f2, f4* by linear functions, then for sales volume it is necessary to know the price of each product, for example on the basis of management accounting data, variable costs per output unitand fixed costs: in total for the program and for each product separately.

4. Demand forecast $G_j^w$for products*j* based on product life cycles shall be carried out on the basis of the Monte Carlo algorithm, and then for each product formedthe restrictions of the following type:

$$X_j \leq G_j^w \leq X_j^m$$

for the optimization model of the production program. This intermediate restriction is necessary to account for the production capacity for each product: it suggests that the upper random forecasting limit G_j^wshould not exceed the production capacity for this product.

5. Carry out the calculation of the optimization model and determine the optimal program.

6. For the product being replaced *j** should be formed the options for strategies of investment and replacement of *j* on *j**. The options for strategies are conditioned by the investment volume and rate, the nature of the investment in stages of the innovation process, the rate of production growth up to the project level and the payback period, for example, uniform investment at all stages; investing in large quantities in the early stages; investments in new technologies and organization of production, i.e. improving the innovation capacity, etc. The replacement strategy consists of two vectors, the elements of the first vector show the volume of investments for each year up to the year of the start of mass production and further for each year the projected production volume of a new product; second vector shows a mechanism for replacement of an *"old" products* on a *"new"* one in accordance with equations (3), (4).

7. For each option of the replacement of a product on a new one (similar to the investment project) we calculate *NPV*. If *NPV* = 0, we define the proposed the period of expected project payback on the launch of a new product, which is included in this option of the strategy for the evaluation of the initial investment I_n.

8. The source of financing (covering) I_n is a part of fixed costs of the enterprise. The total amount of investment in innovation is distributed for the years in accordance with an

option of the replacement strategy (Clause 5) of this algorithm, so that

$$I_n = \sum_{t=1}^{T}(1-\lambda_t)I,$$

at the same time we have previously shown that I - is the amount of investment in OIP, in the optimization model it serves as the right part of restrictions on the volume of investment in the entire program of current production and innovations. The mechanism of the ratio is as follows: the volume of innovation investment for the following yearI_{nt+1} is formed due to the undistributed profit for preceding year t and general fixed costs.

9. Then turn to the calculation of the financial parameters and indicators, as well as the parameters of sustainable growth.

10. Paragraphs 6 and 8 shall be repeated for all options of replacement strategies.

11. Select the version of the decision that is the most satisfactory in cost and time criteria.

IV. Conclusion

Thus, as a result of the study implemented a systematic formulation of the general task of the formation of operational and innovation program, designed to optimize under different criteria the process of resource usage, including: investment, financial, technological, production and information capabilities of the corporation. There have also been solved the problem of the construction of the model used for planning innovations from the point of view of investment strategies and optimization of launch of new products on the market and maximize the rents.

The optimization task allows to evaluate the effectiveness of not only the investment projects, innovations, but also any other organizational and technical measures aimed at changing the technical, economic and financial indicators, achievement of their expected values.

The solution of the given problem of planning the activities of the enterprise for the future can be regarded as the optimization of planning the operational and innovation activities of the enterprise, technical, economic and financial planning, the analysis of the financial and economic state of the enterprise over time, based on all the main indicators in and outside this model.

Such a problem is important for research on the development of a methodology for sustainable development of the enterprise, as well as for the practice of its management. Consistency in the technical, economic and financial management of the company allows to justify the prospect of its development, economically evaluate the options for such a development, justify the value of the stochastic reserve of sales volume, profits and other strategic indicators in order to ensure the performance of the key planned indicators.

References

[1] Bogoviz, A., Mezhov, S. Innovation Processes in the Russian Economy: problems and development trends. Economics of Sustainable Development, 2014, 2(18), 21-28.

[2] Boutellier, R. Managing Global Innovation, Uncovering the Secrets of Future Competitiveness, 3rd revised ed. Berlin: Springer, 2008.

[3] Chesbrough, H. Open Innovation: Researching a New Paradigm. Oxford: Oxford University Press, 2006.

[4] Cooper, R., Edgett, S. Stage-Gate and the critical success factors for new product development. BP Trends, July, 2006.

[5] Danilin, V. Operating and financial planning in corporations (methods and models).Moscow: Science, 2006.

[6] Gassmann, O. Opening up the innovation process: towards an agenda. R&D Management, 2006, 36(3), 223-226.

[7] Higgins, R. Analysis for Financial Management, 5th ed. Irwin, 1998.

[8] Kondratiev, V. Corporate governance and investment process. Moscow: Science, 2003.

[9] Mezhov, S. Economy of the innovation corporation: Theory and problems of efficiency. Barnaul: Publishing House AAEP, 2012.

[10] Mezhov, S. Operational and innovative processes of corporations: the paradigm of planning and efficiency assessment. Bulletin "AAEP", Barnaul, 2014б 2(34), 54-57.

[11] Mezhov, S. Some approaches to solution of the problems of organization of mass innovative production. Innovations, 2011, 6(152), 118-123.

[12] Mezhov, S. Theoretical and Methodological Bases of Forming the Operational and Innovation Program of the Company Advances.Natural and Applied Sciences, 2014, 8(5), 451-457.

[13] Pleschinsky, A. Optimization of inter-firm interactions and intra-firm management decisions. Moscow: Science, Series "Economics of the modern Russia", 2004.

[14] Prahalad, C. Competing for the future. Harvard Business School Press, April 1996.

[15] Ragulina, Y., Malyshkov, V. The Entrepreneurial Climate in Russia: The Present and the Future. Life Science Journal, 2014, 11(6s).

[16] Sobolev, V. Modeling and optimization in managing the development of large economic systems (the full product life cycle). Novosibirsk: NGAEU, 2000.

[17] Titov, V. Optimization of industrial corporation management: methodology and modeling issues. Novosibirsk: IEOPP SB RAS, 2007.

[18] Titov, V., Mezhov, S. A model of operational and innovative program of industrial enterprise. Improving the institutional management mechanisms in industrial corporations, edited by Titov V.V. & Markova V.D.Novosibirsk: PublishingHouse "IEOPPSBRAS", 2013.

[19] Van Horne, J. Fundamentals of Financial Management, 13th ed. Prentice Hall, 2009.