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

Proceedings of the 6th International Conference on Applied Innovations in IT

Volume 6

EDITION
Hochschule Anhalt

Proceedings of the 6th International Conference on Applied Innovations in IT

Volume 6

Koethen , Germany
13 March 2018

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 Innovations in IT (ICAIIT), which took place in Koethen March 13th 2018. 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: 978-3-86011-093-5 (Online)
ISSN: 2199-8876

For permission requests, please contact the publisher:
Anhalt University of Applied Sciences Bernburg / Koethen / Dessau
Email: eduard.siemens@hs-anhalt.de

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: eduard.siemens@hs-anhalt.de
Web: http://icaiit.org

Content

Section 1. Data processing and data analysis

Ivan Luzyanin and Anton Petrochenkov
Detection of Changes in Oil Well Power Consumption Profile on the Basis of Dynamic Time Warping Algorithm 1

Artur Mikhailov and Natalia Gergel
The Use of News Reports to Predict the Values of Macroeconomic Indicators and Indices Represented by Time Series 11

Rustam Fayzrakhmanov, Alexandr Kulikov and Polina Repp
The Difference between Precision-recall and ROC Curves for Evaluating the Performance of Credit Card Fraud Detection Models 17

Section 2. Automation and algorithms

Denis Eltyshev and Ksenia Gnutova
Influence of Fuzzy Clustering on the Accuracy of Electrical Equipment Diagnostic Models 23

Viktor Kubil, Vasily Mokhov and Dmitry Grinchenkov
Modelling the Generalized Multi-objective Vehicle Routing Problem Based on Costs.... 29

Rustam Fayzrakhmanov, Rustam Murzakaev and Anatoly Polyakov
Cutting Time Optimization Using Technology for CNC machines.................... 37

Nikola Jovanovski and Josif Kjosev
Synchronized Control of Four or More Stepper Motors for Computer Numerical Controled Machines and 3D Printers.................... 45

Mikhail Sadiakhmatov and Leonid Mylnikov
Prediction-based Planning in Production System Management through Subsystem Interaction.................... 51

Rustam Fayzrakhmanov, Ivan Polevshchikov and Anatoly Polyakov
Computer-aided Control of Sensorimotor Skills Development in Operators of Manufacturing Installations.................... 59

Leonid Mylnikov, Dmitrii Vershinin and Daniil Fatkhullin
The Use of Optimal Management Tasks for Verification and Adjustment of New Product Release Planning in Discrete Production Systems............ 67

Anatolii Kusakin, Anton Petrochenkov, Dmitrii Leizgold and Ivan Luzyanin
Development of an Algorithm for Transition from an Electric-power System Visual Representation to a Parametric Representation............ 77

Anna Novikova
Direct Machine Translation and Formalization Issues of Language Structures and Their Matches by Automated Machine Translation for the Russian-English Language Pair........85

Section 3. Management in Economic Systems

Julia Dubrovskaya, Elena Kozonogova, Igor Nosov and Yuliya Karpovich
Automation of the Process of Regional Development Management Based on the Use of Smart Benchmarking Technology...93

Stepan Mezhov and Leonid Mylnikov
Specifics of Project Management in Industrial Innovation............. 103

Section 4. Communication technologies

Kostromina Anna, Eduard Siemens and Yurii Babich
A Concept for a High-reliability Meteorological Monitoring System Using AMQP.........109

Jannis Ohms, Olaf Gebauer, Nadiia Kotelnikova, Marina Arikova and Diederich Wermser
Experiences Implementing QoS Aware Routing on Off-the-shelf SDN Switches.............117

Ali Mekhtiyev, Felix Bulatbayev, Yelena Neshina, Aliya. Alkina, Eduard Siemens and Tatyana Shaigarayeva
The External Mechanical Effects on the Value of Additional Losses in the Telecommunicationsfiber Optic Cables under Operating Conditions....... 123

Nikolai Mareev, Dmitry Kachan, Kirill Karpov, Dmytro Syzov, Eduard Siemens and Yurii Babich
Efficiency of a PID-based Congestion Control For High-speed IP-Networks..................129

Detection of Changes in Oil Well Power Consumption Profile on the Basis of Dynamic Time Warping Algorithm

Ivan Luzyanin and Anton Petrochenkov
Electrotechnical Department, Perm National Research Polytechnic University,
Komsomolsky Ave. 29, 614990, Perm, Russia
{lis, pab}@msa.pstu.ru

Keywords: Oil Field, Electric Power System, Statistical Model, Dynamic Time Warping, Distance Measurement, Signal Processing.

Abstract: At present oil companies are forced to continually decrease electric power inputs. However, energy efficiency of oil well equipment decreases in time. Well re-equipment enables to stop energy efficiency loss but it requires large additional inputs. The possible solution of this problem is development of the energy efficiency growth strategy that does not include equipment replacement. To do this the oil well model that is able to precisely estimate energy efficiency of every element in electric power system needs to be constructed. Oil well technological and mechanical parameters, determining production efficiency, are strongly connected to the electric parameters of equipment. Therefore, they need to be included in the model. Models used in oil companies for energy efficiency estimation reflect dependencies between described parameters but they do not consider instant changes of electric parameters caused by changing of electric power system regime. Mathematical models of electric power systems that consider instant changes of electrical parameters are based on differential equations which have complicated solutions. The paper considers a method for instant changes analysis in power consumption profiles of oil well equipment that is based on dynamic time warping algorithm. It is demonstrated that instant changes of electrical parameters at the short time period caused only by electric power system regime changes and are independent from well production conditions. Based on this thesis it is proposed to study instant changes of electrical parameters in wells with similar production conditions. The comparison of two modifications of dynamic time warping algorithm is presented. Investigation of the properties of given modifications when applying to power consumption profiles exposes limitations of using the method. However, the study of other algorithm modifications allows to find possible ways of overcoming the restrictions.

1 INTRODUCTION

When operating oil field, two processes occur: on the one hand, depleting of the oil reserves causes changing of extraction conditions, on the other hand, ageing of well equipment causes increase of electrical energy loss in elements of electric power systems (EPS) and hydraulic loss in tubing strings. These factors lead to increase of the operation and maintenance expenses of the oil wells.

To ensure a stable profit, oil companies are forced to yearly increase the oil extraction while reducing the cost of operation and maintenance of oil fields. Unreasonable selection and misuse of well electrical equipment causes inefficient EPS operating regimes (e.g. underload and overload) and also leads to an increase of the electrical power inputs due to losses.

When operating the well, it is necessary to maintain parameters of the technological process that ensure maximal flow rate to the well under given geological, climatic and technological conditions.

The technological parameters restrict operating of mechanical and electrical equipment of the well.

Since full-scale experiments in oil fields are not allowed, the methods of studying the well are based on mathematical modeling [1], simulation [2][3] and time series analysis and prediction [4][5].

At present, different technological, hydraulic and electrical models of oil well are developed. These models allow to make decisions on rational choice and effective operation of well equipment. However, they are based on theoretical equations and do not completely meet real operational conditions.

Moreover, to increase precision of these models the object parameters identification needs to be done. It requires obtaining of internal parameters of equipment (e.g. motor flux linkage, rotor and stator resistances and others) that is impossible in real conditions.

In these conditions, the task of evaluating the equipment parameters subject to its operation features under the conditions of uncertainty and data incompleteness becomes important. To solve this, oil field statistical model can be constructed. This model makes possible to analyze the object by indirect method based on statistical data representing changes in electrical, mechanical, and technological parameters.

Since the well production conditions and the operating practice vary depending on many factors, it is necessary to understand the nature of these changes and the leverage of various factors on them. Based on this information it is possible to make a data clustering for identification of typical regimes.

The main indicator that determines the economic efficiency of the oil well is the specific power consumption. This is the ratio of the amount of electrical power, consumed by oil extraction equipment, to the mass of the produced oil or the volume of the liquid produced. Consumption level depends on electrical equipment type, EPS regime parameters and control algorithm applied to the pump electrical drive. The volume of extracted liquid depends on geological, climatic and technological parameters. The mass of produced petroleum is determined by chemical composition of the formation fluid and the content of water and gas in it.

The paper considers questions of analysis of changes in well parameters under different operational conditions.

The aim of the research is studying of power consumption changes when changing electrical parameters of well equipment and pump control parameters under different fixed values of technological and mechanical parameters.

The degree of relationship between available electrical, mechanical and technological parameters and the intensity of their changes is studied in the research. Based on these data, the analysis of changes in power consumption profiles will be carried out.

2 OIL FIELD PARAMETERS

Oil well is a vertical, inclined or horizontal bore connecting surface with reservoir.

When productive formation keeps shut in, the reservoir pressure is equal at every point and liquid does not flow. When formation exposing the pressure at the wellbore becomes less than reservoir pressure and liquid starts flowing to the well [6]. Flowing continues until the difference between reservoir and wellbore pressures becomes less than the sum of hydraulic resistances in a tubing string.

The main operational characteristics of oil well are production condition and lifting type.

Production conditions define energy sources that provide maintenance of reservoir pressure sufficient for lifting liquid to the surface. The study assumes production conditions are given by oil field operating practice.

Lift type defines tools used for lifting the liquid. This study considers wells with pumping based on electrical submersible pumps (ESP) with induction motors (IM) placed inside the well.

Production rate (measured in barrels per day (BPD)) determines volume of liquid potentially being extracted from the well at a given time period. BPD depends on well inflow and determines total company profit obtaining from the well.

Well operational expenses depend on different parameters. When pumping, the most expenses are electrical energy costs (up to 40% of total costs). Therefore, well operational efficiency is measured by specific power consumption described above. In these conditions, increase of operational efficiency can be obtained either by BPD increasing or by decreasing energy consumption. This paper considers abilities of energy consumption decreasing when fixed BPD values.

The next subsections describe main well parameters and their dependencies.

2.1 Technological parameters

The main technological parameter that determines liquid extraction efficiency is well inflow. It can be obtained using Darcy equation [1, 6]. The general solution of this is complex; therefore in practice the specific solution is used. It holds when the following assumptions:

- flowing is radial around the well;
- reservoir characteristics and liquid composition do not change in sufficiently long time period.

When following above assumptions, geological parameters of the productive formation do not significantly change in a short time period. In these conditions, the pressure drawdown (calculated as the difference between reservoir and bottomhole pressures) determines well inflow. The value of it has to be maintained a constant in accordance with technological process.

2.2 Mechanical Parameters

The subject of this study is well equipped with electrical submersible pumps with induction motors (ESP). The ESP provides lifting of the reservoir liquid to the wellhead and maintaining wellhead pressure sufficient for moving liquid to the booster pumps.

Main parameters of ESP are head (h) and flow rate (q). Head is the height of vertical column of liquid generating at the discharge of the pump.

Flow rate defines liquid volume that pump is able to lift to the height equal to h under given hydrodynamic parameters of tubing string. It depends on pipe diameter, flow velocity and pipe hydraulic resistance.

When head is given, flow rate can be obtained using H-Q curve. This curve is presented in ESP manuals.

ESP converts kinetic energy of shaft rotation into pressure energy. The following equations describe connections between pump parameters and rotational speed of the motor:

$$q_2 = q_1 \left(\frac{n_2}{n_1} \right) \quad , \tag{1}$$

$$h_2 = h_1 \left(\frac{n_2}{n_1} \right)^2 \quad , \tag{2}$$

$$BHP_2 = BHP_1 \left(\frac{N_2}{N_1} \right)^3 \quad . \tag{3}$$

In the above formulas h_1, h_2 are pump heads, q_1, q_2 are flow rates, BHP_1, BHP_2 are pump break horsepowers, and n_1, n_2 are rotational speeds in two different operational conditions respectively.

When substituting nominal values of corresponding parameters to the (1-3), pump characteristics for any given rotational speed can be obtained.

2.2 Electrical Parameters

ESP is driven by induction motor that installed in one shaft with a pump stages. Therefore, it can be assumed that motor torque is equal to the pump torque:

$$\tau = \tau_r \quad , \tag{4}$$

where τ is a motor torque, τ_r is a pump torque.

IM consumes power of two types: active power (P) that is spent on the shaft rotation and reactive power (Q) that is spent on electric field generation. Reactive power is usually compensated by special equipment, therefore this study considers only active power consumption.

The frequency converters are usually used to control IM in oil wells. They change the rotational speed of the model by changing both mains frequency and voltage. The equation (5) describes dependency between synchronous speed of the motor and AC power frequency.

$$n_{synch} = \frac{120 \cdot f}{p_{poles}} \quad , \tag{5}$$

where n_{synch} is synchronous speed, f is frequency of AC power, p_{poles} is number of poles in stator.

Active power of the IM is calculated by the following formula:

$$P = \tau_r \cdot \omega \cdot \eta_m \cdot \eta_p \quad , \tag{6}$$

where ω is angular velocity, τ_r is pump torque, η_m is motor efficiency, η_p is pump efficiency.

Synchronous speed and angular velocity of IM is connected by the following expression:

$$\omega_{synch} = \frac{\pi \cdot n_{synch}}{30} \tag{7}$$

The shaft rotation speed of the induction motor is less than synchronous rotation speed of magnetic field by the value of $\Delta\omega$ depending on the slip value:

$$slip = \frac{\omega_{synch} - \omega}{\omega} \quad , \tag{8}$$

where ω_{synch} is synchronous IM speed, ω is IM shaft speed.

The slip value depends on the torque developed by the engine, however, under given conditions, the slip can be assumed constant and equal to the nominal value. The bases of this assumption are given below.

Dependencies between the IM torque and the pump torque and induction motor rotational speed when different values of voltage and frequency are shown in Figure 1.

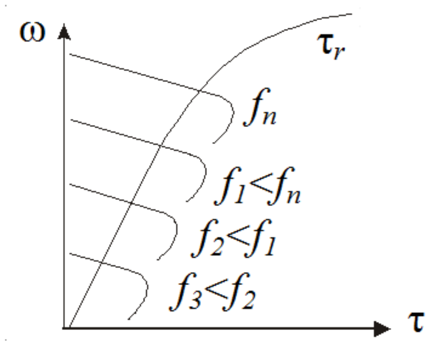


Figure 1: Motor and pump torque curves under different frequency values.

The figure shows that for the given load type, the IM torque required to rotate the pump shaft is significantly less than the critical torque at the entire frequency range. Under operating conditions, the frequency control range is sufficiently small (30-60 Hz), and the torque developed by the motor does not change without changing the frequency. The intersections of the IM and pump torque curves are at the segment where the torque curve has sufficiently slight slope. Thus, the slip in the whole control range has insignificant changes in comparison with the nominal value. In this case, it is possible to not consider slip changes when change the motor torque.

Formulas considered in the section, describe dependencies between main parameters of oil well but the typical models based on them are robust and do not allow to study tiny changes in power consumption profiles of wells [2][6][7].

3 STATISTICAL MODEL OF OIL FIELD

For estimating power consumption changes in oil fields the average consumption values are used (in energy units, kWh).

Standard averaging intervals are day, month, quarter, and year. Energy efficiency is estimated in a whole field and is determined by average annual integrated consumption index. For increasing energy efficiency, energy consumption of a field is yearly decreasing on fixed value.

The expected values of the well energy consumption are calculated based on the formulas described in the section above. Calculated parameters are averaging then by the whole field.

Current and expected power consumption are calculated based on actual volume of produced oil and expected values of production rates. The annual electricity consumption reduction is determined by these parameters.

This technology has the following shortcomings: the potential optimizing abilities of a single well are not considered; averaging over long periods does not allow to determine the cause of ineffective operating regimes of the well electric equipment. Ineffective regimes are both regimes with high power consumption and emergency regimes.

The average energy consumption is used as the main parameter for energy efficiency estimation. To determine the optimal averaging interval of this parameter, 32 wells placed in two fields with different geological characteristics and lifecycle stages were analyzed. The profiles of well energy consumption were built with the averaging intervals of month, day, hour, minute and second. When analyzing profiles with shorter averaging intervals, the average values of power consumption over the previous interval were used as a template for comparing. The study showed that the optimal averaging interval is an hour.

Table 1: Oil Well Parameters.

Parameter (Symbol)	Units
Active power (P)	kW
Frequency (f)	Hz
Motor Rotational Speed (n)	RPM
Motor temperature (T)	˚C
Intake pressure (p_{in})	bar
Wellhead pressure (p_{wh})	bar
Head (h)	ft.
Liquid production rate (q)	BPH

During preliminary study changes intensity analysis as well as correlation analysis of given parameters were carried out. It was found that electrical parameters are changed intensively while other parameters are subject to weak changes. Correlation analysis showed that all electrical parameters and motor temperature have strong relations with correlation coefficients of 0.7 (Spearman correlation) and 0.9 (Pearson correlation). Technological and electrical parameters as well as mechanical parameters have weak correlation (with correlation coefficients of about 0.02 for both methods). Based on obtained data it was concluded that the volume of information is insufficient for determining statistical dependencies between electrical and non-electrical parameters of oil well. The preliminary study showed that energy consumption changes in a single well are not caused by changes of geological, technological and mechanical processes in a short time period. The changes come from internal electrical and thermal processes in equipment and external parameters of EPS regimes.

On the ground of above analysis it was suggested to divide full energy consumption profile into two parts. The first part is caused by technological process. It is relatively stable for a single well in a short time period. The second part is caused by external changes of EPS regimes and internal changes in equipment. This part has significant changes even in a short time period. Minimizing of energy consumption can be obtained by reducing them. To do this the variable part of the signal is to be extracted from the whole profile. The extraction is based on comparative analysis of test energy consumption profile and reference one. Profiles with known production conditions and EPS regimes were selected as reference signals.

At the main research phase the deviation analysis of power consumption profiles from the template profile was carried out. The profiles that demonstrate consumption of oil well in the known production regimes and under the same operating conditions as the investigated wells were selected as template profiles.

To study the changes in electricity consumption, eight identical samples were generated for four wells (two samples per well on October and March, respectively). This choice was made based on the results of enterprise inspection that showed the most unfavorable changes in electricity consumption in the autumn and spring. Selected wells were in operation during the given time intervals and the biggest amount of data was obtained from them.

Each sample consists of 30 columns corresponding to the day of month. Each column has 24 rows where average hour consumption values are placed. The day consumption change graphs were built using the samples. These graphs were used both as reference and as test signals in the analysis procedure described in the next section.

4 ENERGY CONSUMPTION ANALYSIS

Dynamic time warping algorithm (DTW) was originally introduced as a tool for similarity measurement of complex signals [8] but it is also possible to use it for measuring differences between test signal and given template [9]. Algorithm transforms test signal into template by stretching and shrinking different segments of time axis. Algorithm accuracy depends on similarity of test signal and template after warping. When warping, optimal warping path is constructed. Optimal path is a matrix that contains minimal amount of transformations providing maximal similarity of warped signals. Full description of the DTW and its features is given in [8] – [13].

Warping path is defining points of signals being shifted when warping and the shifting distances. To do this optimally, the weighting matrix is used. Weighting matrix constrains possible ways of points shifting and maximal shifting distances. Weighting matrix influences the accuracy of algorithm and its ability to give right similarity measures. Different types of weighting matrices are considered in [14] – [16].

In the study two different weighting matrices are used (9) and (10). Below expressions describe possible shifting ways and distances for classical and modified DTW respectively.

$$\begin{bmatrix} m & n-1 \\ m-1 & n-1 \\ m-1 & n \end{bmatrix}, \quad (9)$$

$$\begin{bmatrix} m & n-1 \\ m-1 & n-2 \\ m-1 & n-1 \\ m-2 & n-1 \\ m-1 & n \end{bmatrix}, \quad (10)$$

To estimate the deviation of signals by the shape of the optimal path curve, the method of analyzing the deviations of the path from the diagonal, proposed in [9], was used.

During the analysis, the following parameters of the algorithm were evaluated: the matching accuracy of the test reference signals after warping, the distribution of the matched points of the test and reference signals, distribution of distances between warping path curve and diagonal. The following graphs were built: signals before and after warping, matching diagrams, warping path diagrams along with diagonal and lines showing distances between path and diagonal.

The study results for the classical DTW when the test power consumption profile corresponding to the stationary regime with small deviations are showed in Figures 2-5.

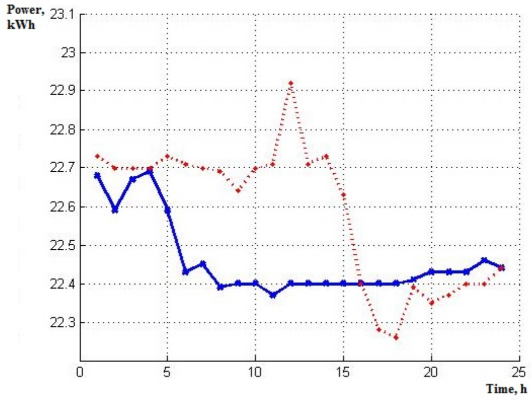


Figure 2: Power consumption profile of oil well. Dotted line - reference signal, solid line - test one.

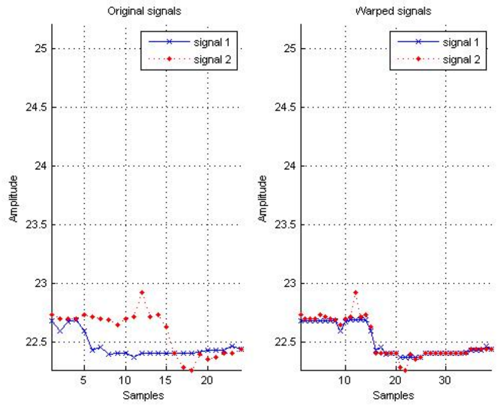


Figure 3: Warping diagrams for classical DTW. Left graph – signals before warping, right graph – signals after warping.

The graph of signals after warping shows that the curves of the test and reference signals are close to each other. It indicates sufficiently high accuracy of the algorithm for this type of curves.

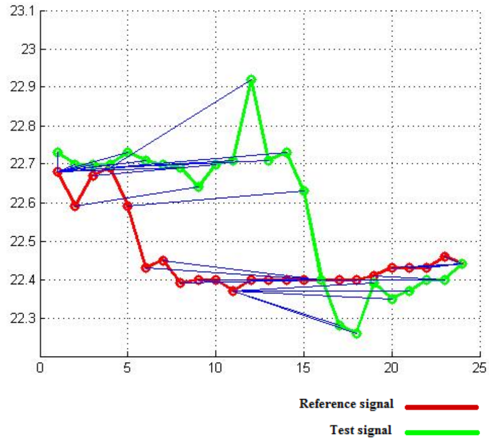


Figure 4: Matching diagram for classical DTW. Reference signal is green, test signal is red. Blue lines show shifting distances of points.

Matching diagram has multiple matching points. There are the points of a signal where more than one matching lines come (in the figure matching lines are blue). The multiple matching points reduce the accuracy of the algorithm and the information capability of the warping path curve. They produce long straight sections on the warping path.

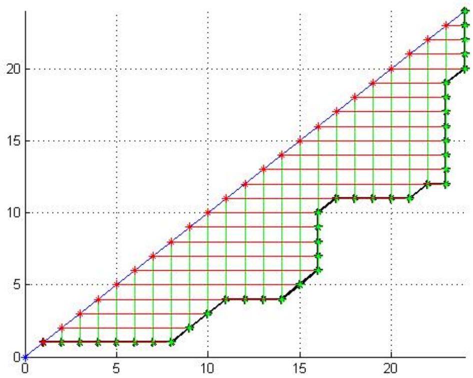


Figure 5: Warping patch along with diagonal for classical DTW. Lines show distances between path and diagonal.

Deviations of the warping path from the diagonal show the discrepancy between the reference and test signals at each point. The better the signals match each other after warping, the more accurately changes are reflected in the warping path. If a certain optimal power consumption profile is used as a reference signal (e. g., obtained by mathematical model), the discrepancy can be used to find the points where potential problems exist. These problems, then, need to be investigated by other methods. Thus, the algorithm can be used to find the sections of power consumption profiles to be optimized.

Figures 6-8 illustrate the warping results with modified weighting matrix (10).

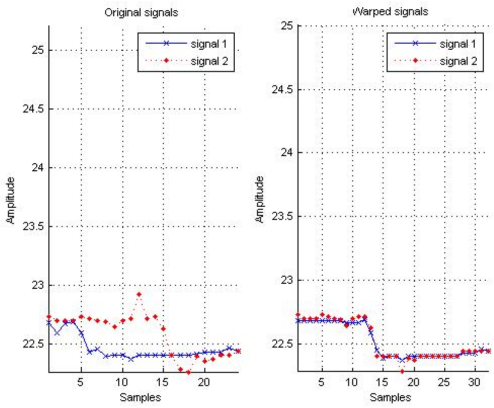


Figure 6: Warping diagrams for modified DTW. Left graph – signals before warping, right graph – signals after warping.

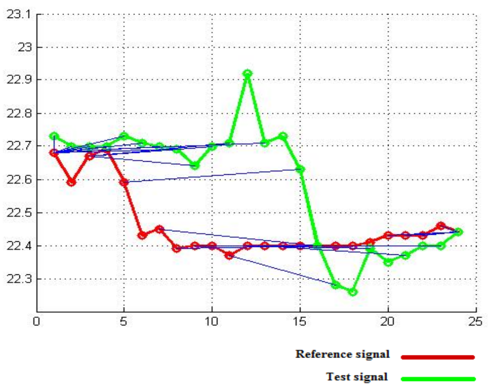


Figure 7: Matching diagram for modified DTW. Reference signal is green, test signal is red. Blue lines show shifting distances of points.

The matching diagram of the modified algorithm has fewer multiple matching points, but it also has unconnected points that can lead to the loss of significant points.

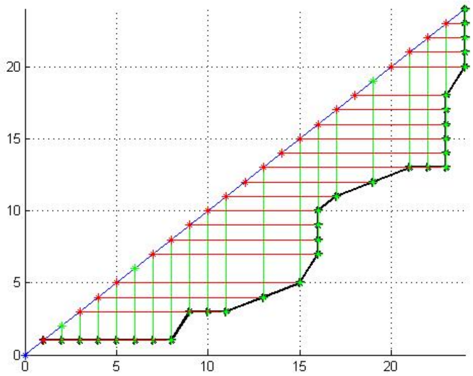


Figure 8: Warping patch along with diagonal for modified DTW. Lines show distances between path and diagonal.

The study showed that the algorithm with a modified weight matrix is more accurate than the classical one. In addition, this algorithm has nearly no cases of multiple matching, but there are points at the matching diagram that do not have connections. This gives potential ability for skipping these points. If the skipped point is significant (e.g. it demonstrates a significant decrease of power consumption), skipping the point leads to incorrect interpretation of the warping path curve.

Further studies showed that if the discrepancy of signals increases, both the number of multiple matching cases in the classical algorithm and the number of missing points in the modified algorithm increase (Figures 9-13).

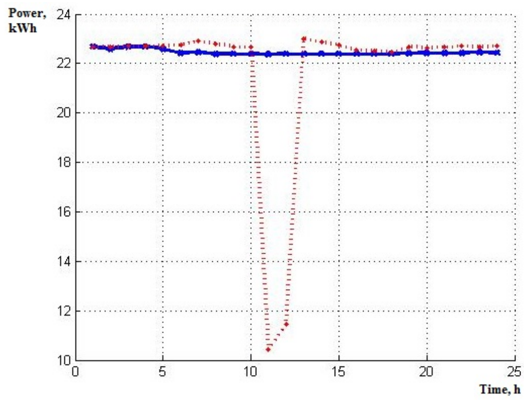


Figure 9: Example of a profile with a big deviation. Dotted line - reference signal, solid line - test one.

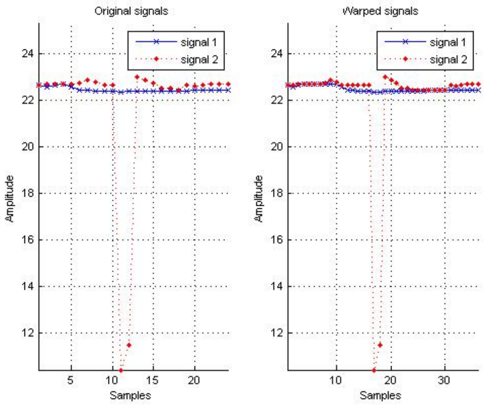


Figure 10: Warping diagrams for classical DTW. Left graph – signals before warping, right graph – signals after warping.

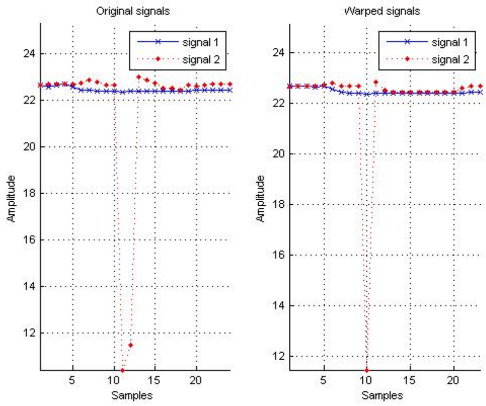


Figure 12: Warping diagrams for modified DTW. Left graph – signals before warping, right graph – signals after warping.

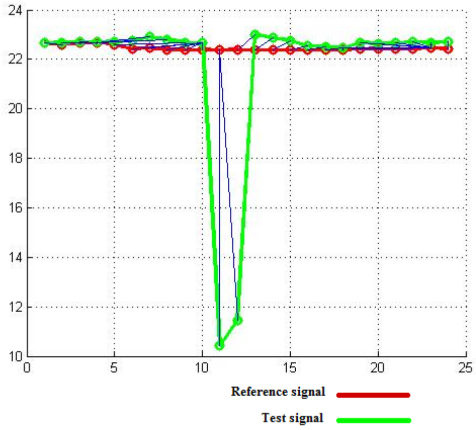


Figure 11: Matching diagram for classical DTW. Reference signal is green, test signal is red. Blue lines show shifting distances of points.

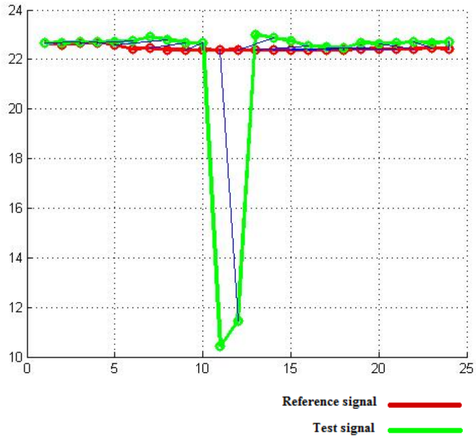


Figure 13: Matching diagram for modified DTW. Blue lines show shifting distances of points.

The accuracy of both algorithms decreases.

As a limiting case, the transient process of an emergency motor shutdown was considered (Figure 14).

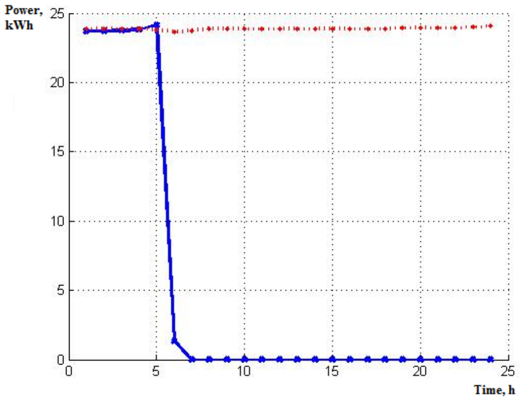


Figure 14: Example of profile with zero-valued segment. Dotted line - reference signal, solid line - test one.

The specified signal in this case becomes zero. The accuracy of both algorithms decreases significantly in this case (Figures 14-15). In addition, at the zero-valued segment of the test signal, the warping path curve in classical DTW algorithm matches with the diagonal (Figure 16), which makes it uninformative.

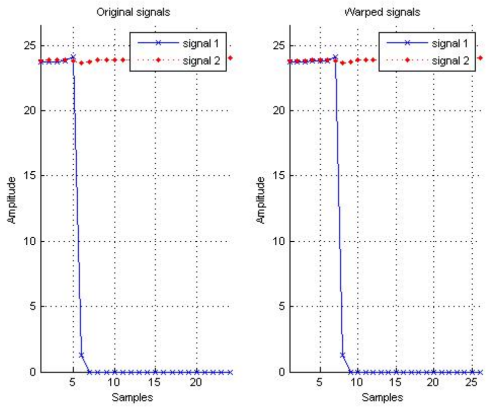


Figure 15: Warping diagrams for classical DTW. Left graph – signals before warping, right graph – signals after warping.

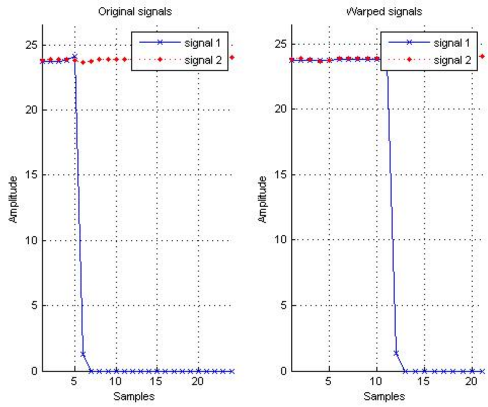


Figure 16: Warping diagrams for modified DTW. Left graph – signals before warping, right graph – signals after warping.

An additional study showed that classical DTW is not appropriate for comparing signals with straight lines. Modified algorithm in this case gives unreliable results and is also not able to be used.

The effects described appear because only the time axis is warped, so the algorithm recognizes properly the horizontal changes in the signal, but vertical changes are not recognized well. One possible solution of this problem is warping the whole plane [17] that allows to transform both time and value axes. The accuracy of the modified algorithm exceeds the accuracy of the classical one. When difference between signals is not significant, the modified DTW has more correct results but this algorithm allows unconnected points that can lead to significant change loss.

5 CONCLUSIONS

The study showed that parameters of the technological process and mechanical pump characteristics have weak influence to the instant changes of well equipment electric parameters. These changes caused only by changing EPS operational regime. It corresponds to theoretical statements described in [1][6].

Cross-sectional analysis of signals obtained from wells with similar production conditions was carried out to study instant changes patterns in electric power consumption profiles. Two modifications of DTW algorithm with different weighting matrices were used for the analysis. The methodic based on measuring distances between warping path and diagonal was used for DTW results interpretation.

Study allowed to define restrictions of described algorithm modifications on precision of changes detection when recognizing signal differences. When analyzing tiny magnitude changes in signals both modifications of the algorithm had precise results. This fact corresponded to the conclusions given in [9], [10]. However, none of these modifications was able to correctly recognize large magnitude changes between signals. In addition, when studying signals with large straight sections (both zero-valued and not), presence of which is a feature of considered profiles, results interpretation is impossible due to incorrect form of warping curve.

The study showed that classical DTW algorithm recognized changes with less precision then modified one. This problem also considers in [14] [16]. Nevertheless, when using modified algorithm, mismatching points appears. This may cause significant decreasing of recognition precision when mismatched point corresponds to significant regime change. Although in several works [8][10] – [13] these effects are not consider, they constrain use of this algorithm for described task.

Experiments showed that algorithm better recognizes changes in width of signals (shifting points along the time axis) than in magnitude. The possible solution for this problem is use of two-dimensional warping algorithm [17]. Moreover, this algorithm has variety of modifications [14] – [17] eliminating some negative effects when analyzing signals with different specific features.

The research highlighted features of DTW algorithm that restricted its use for analyzing changes in power consumption profiles. It also depicted basic features of the power consumption profiles themselves. Obtained results will be the basis for further investigations that will conform the algorithm to specific features of studied signals.

REFERENCES

[1] L. Hailong, "The numerical simulation for multistage fractured horizontal well in low-permeability reservoirs based on modified Darcy's equation", Journal of Petroleum Exploration and Production Technology, vol. 7(3), pp. 735-746, 2017, DOI:10.1007/s13202-016-0283-1

[2] V. V. Alekseev, A. P. Emel'yanov, and A. E. Kozyaruk, "Analysis of the dynamic performance of a variable-frequency induction motor drive using various control structures and algorithms", Russian Electrical Engineering vol. 87, no. 4, pp. 181-188, DOI: 10.3103/S1068371216040027

[3] D. C. Montgomery, "Design and Analysis of Experiments", Wiley, 2012, 752 p.

[4] Chatfield, C. "Time-series forecasting", New York: Chapman and Hall, 2001;

[5] Gurol Irzik "Causal Modeling and the Statistical Analysis of Causation", 1996;

[6] G. Takacs, Electrical submersible pump manual: design, operations, and maintenance, Gulf Professional Publishing, 2009, 420 p.

[7] J.F.Gülich, Centrifugal pumps, 2nd Edition, Springer, 2010, 998 p.

[8] Sakoe H., Chiba S., A Dynamic Programming Approach to Continuous Speech Recognition, In Proceedings of the 7th International Congress on Acoustics, vol. 3, 1971, pp. 65-69.

[9] I. Luzyanin, A. Petrochenkov, B. Krause, Problems of tiny changes analysis in complex time series using dynamic time warping algorithm, Proceedings of the XIX International Conference on Soft Computing and Measurements. SCM`2016, 2016, pp. 419-422, DOI: 10.1109/SCM.2016.7519799

[10] Y. Zhang, T. F. Edgar, A Robust Dynamic Time Warping Algorithm for Batch Trajectory Synchronization, American Control Conference, Seattle, June 2008, pp. 2864-2869.

[11] 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.

[12] 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.

[13] 10. M. Müller, Dynamic Time Warping, Information Retrieval for Music and Motion, Berlin Heidelberg: Springer 2007, pp. 69-84.

[14] Y. Jeong, M. K. Jeong, O. A. Omitaomu, Weighted Dynamic Time Warping for Time Series Classification, Pattern Recognition, No. 44, 2011, pp. 2231-2240.

[15] M. Kotas, J. M. Leski, and T. Moró, (2016). Dynamic time warping based on modified alignment costs for evoked potentials averaging DOI:10.1007/978-3-319-23437-3_26.

[16] 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.

[17] 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.

The Use of News Reports to Predict the Values of Macroeconomic Indicators and Indices Represented by Time Series

Artur Mikhailov and Natalia Gergel

[1]PermNational Research Polytechnic University,Komsomolsky ave. 29, Perm, Russia
mihailovarthur@rambler.ru, natalia_gergel@mail.ru

Keywords: Forecast, Prediction, Model, Text-mining, Machine Learning, Classification, Time Series.

Abstract: The use of forecasts and predictive models highly affects the process of making decisions. The use of given forecasts allows to increase economic effectiveness of individual entities as well as the corporations. The aim of the article is the investigation of the influence of the weakly formalized factors on the forecasts' accuracy. The study is based on the problem of classification for determining the trends of changing the indicators and the levels of external factors' influences on a change of the referencing parameter. The dataset which contains 25 daily news headings gathered during 8 years was used to make the calculations. The chosen news headlines are related to the stock market and were published by the most authoritative sources such as: Russia Today, Reuters, Scientific American, The Guardian. It was demonstrated that the record of the influence of the information in the news reports on the change of the referencing parameter (using the example of the NASDAQ index) allows clarifying the forecasts taken with the use of functional methods. Therefore, it leads to minimizing mistakes and maximizing the forecasts' reliability.

1 INTRODUCTION

Increasing of forecasting accuracy due to the use of earlier unrecorded factors and the work with large amounts of information are becoming more valuable, because the decisions that are being made directly depend on its quality.

As for now, when working with the macroeconomic indicators, the fact of the influence of scientific discoveries, political changes and the public figures' opinions have to be considered. This type of information mainly comes from the mass media. That is why the use of sources of information where the data is presented as a text deserves a special attention.

The method of machine learning is used in order to work with weakly structured information. When working with the given methods the quality of solutions depends on the quality of the data, amount of information and algorithms that are being used.

The processing of textual information is divided into two stages: preparation and processing of the data.

When preparing the data, some certain questions should be considered: 1) selection of data [1], 2) clearing of data (minimization of noises), 3) choosing the type of textual information for its use in machine algorithms (for example, LDA [2] – Latent Dirichlet Allocation, an approach based on using n-grams [3]), 4) reduce the number of attributes used (frequency algorithms that work with the concepts [4] and terms [5] are used for it), 5) setting correspondence between the dataset and numerical measures (for example, binary feature representation [6], Inverse Document Frequency Method [7]).

In order to process the textual information, the authors are trying to use methods such as: Support Vector Machine (SVM) [8], Regression Algorithms[9], Naïve Bayes [10], Decision Rules or Trees [11], k-NN [12] and [13]. The common factor of these methods is the detection of the relationships between features (which are usually words or phrases), such as input data and target.

2 DATA PREPARATION

For the purposes of this paper the opportunity of using these news reports in order to solve the task of forecasting macroeconomic index using NASDAQ index as an example has to be considered (the data can be taken from the website https://www.finam.ru).

2.1 Description of a data feature format

For the experiment, a dataset (taken from https://www.kaggle.com/aaron7sun/stocknews), which contains 25 news annotations for each day of the previous 8 years from 08.08.2008 to 01.07.2016 was used (the example of data is in the table 1). The data was selected from the authoritative mass media sources on the topic related to economics. Their news headlines became the data, since in comparison to the full news reports, the information used in these headings is straight to the point.

Table 1: The stricter of the news headings data set.

Date	**Heading 1**	**Heading 2**	**…**	**Heading 25**
08.08.2008	b"Georgia 'downs two Russian warplanes' as countries move to brink of war"	b'BREAKING: Musharraf to be impeached.'	…	b"No Help for Mexico's Kidnapping Surge"
11.08.2008	b'Why wont America and Nato help us? If they wont help us now, why did we help them in Iraq?'	b'Bush puts foot down on Georgian conflict'	…	b'All signs point to the US encouraging Georgia to invade South Ossetia. Goddamnit Bush.'
12.08.2008	b'Remember that adorable 9-year-old who sang at the opening ceremonies? That was fake, too.'	b"Russia 'ends Georgia operation'"	…	b"BBC NEWS | Asia-Pacific | Extinction 'by man not climate'"
⋮	⋮	⋮	⋱	⋮
01.07.2016	A 117-year-old woman in Mexico City finally received her birth certificate, and died a few hours later. Trinidad Alvarez Lira had waited years for proof that she had been born in 1898.	IMF chief backs Athens as permanent Olympic host	…	Ozone layer hole seems to be healing - US & UK team shows it's shrunk & may slowly recover.

2.2 Preparation of a training sample

In order to solve the task of classification the following classes of interconnection with the values of NASDAQ index need to be emphasized: strong growth, average growth, weak growth, weak decline, average decline, and strong decline. In order to train the model, it is necessary to link every day of the data set to the corresponding classes. In order to do so, the following steps were taken: 1) the difference of NASDAQ values taken from the current and the previous day was calculated, (delta value), 2) a step of growth and a step of decline were calculated in order to determine conditions for each of the classes in a training sample. Step of growth: $S_g = 2/3 \cdot \bar{G}$ (decline $S_d = 2/3 \cdot \bar{D}$), where $\bar{G}$ - is an average value of growth, $\bar{D}$ - is an average value of decline.

The classes were identified based on the following conditions:

- strong growth $\Delta \geq 2 \cdot S_g$ (value +3);
- average growth $2 \cdot S_g > \Delta \geq S_g$ (value +2);
- weak growth $S_g > \Delta \geq 0$ (value +1);
- weak decline $0 > \Delta \geq S_d$ (value -1);
- average decline $S_d > \Delta \geq 2 \cdot S_d$ (value -2);
- strong decline $2 \cdot S_d > \Delta$ (value -3).

As a result, there is a table 2 in which each of the classes corresponds to the following values: strong decline – 241, average decline – 200, weak decline – 431, weak growth – 479, average growth – 316, strong growth – 280.

Table 2: The example of the NASDAQ index classes' table.

Date	**Nasdaq (close)**	**Δ**	**Class**
08.08.2008	2414,1	-	-
11.08.2008	2439,95	25,85	1
12.08.2008	2430,61	-9,34	-1
⋮	⋮	⋮	⋮
01.07.2016	4862,693	19,953	1

2.3 Preparation of a training sample

In order to use the headings of the news reports (that were presented in English) aspects such as: articles, punctuation marks, numbers and other meaningless words, were excluded. The whole data set was written in a lowercase. These changes were needed in order to conduct private analysis of the whole textual data and the following private analysis of the headlines of each day.

3 DEVELOPMENT OF A MODEL FOR SOLVING THE TASKS OF CLASSIFICATION

3.1 Choosing a machine learning model

Using the inductive approach to the analysis of the results, it can be stated that the models in most of the cases were able to solve the task of classification. The best results were made by the Naïve Bayes model. One of the particular qualities of this method is not being able to work with new features that were not a part of the testing data, which was used to train a model. Based on this, the model that was used in order to solve the task of classification was the Random Forest Method. Its results were better than all of the other methods, except Naïve Bayes.

The models such as Logistic Regression, Naïve Bayes, Random Forest, k-NN were compared in order to choose the method of solving the task of classification. The work of these models was evaluated using different amounts of the data set: the whole data set and partial sample (25% of the whole data set). This investigation showed the behaviour of each of the models of machine learning when using different amounts of data set. It allowed to choose a model of machine learning in order to solve the task of classification. (Table 3)

Table 3: The results of learning and checking the work of methods for the classifying weak growth and average decline of the retrospective data.

Model	Partial sample (25% of the whole dataset)				Full sample (100% of the whole dataset)			
	Probability of the prediction (%)		numerical value (result out of the whole value)		Probability of the prediction (%)		numerical value (result out of the whole value)	
	Chosen classifier	Other classifiers	Chosen classifier	Other classifiers	Chosen classifier	Other classifiers	Chosen classifier	Other classifiers
Weak growth (+1)								
Naïve Bayes	100	92	121/121	337/365	100	97	478/478	1426/1470
Random Forest	44	99	53/121	361/365	69	99	328/478	1452/1470
Logistic Regression	95	99	115/121	364/365	97,5	99	466/478	1469/1470
k-NN	39,5	83,5	48/121	305/365	42,5	84,5	204/478	1244/1470
Average decline (-2)								
Naïve Bayes	100	76	47/47	335/439	100	95	202/202	1655/1746
Random Forest	93,5	100	44/47	439/439	100	100	202/202	1746/1746
Logistic Regression	13	99	6/47	435/439	18	99	37/202	1745/1746
k-NN	15	98	7/47	430/439	22	97	44/202	1692/1746

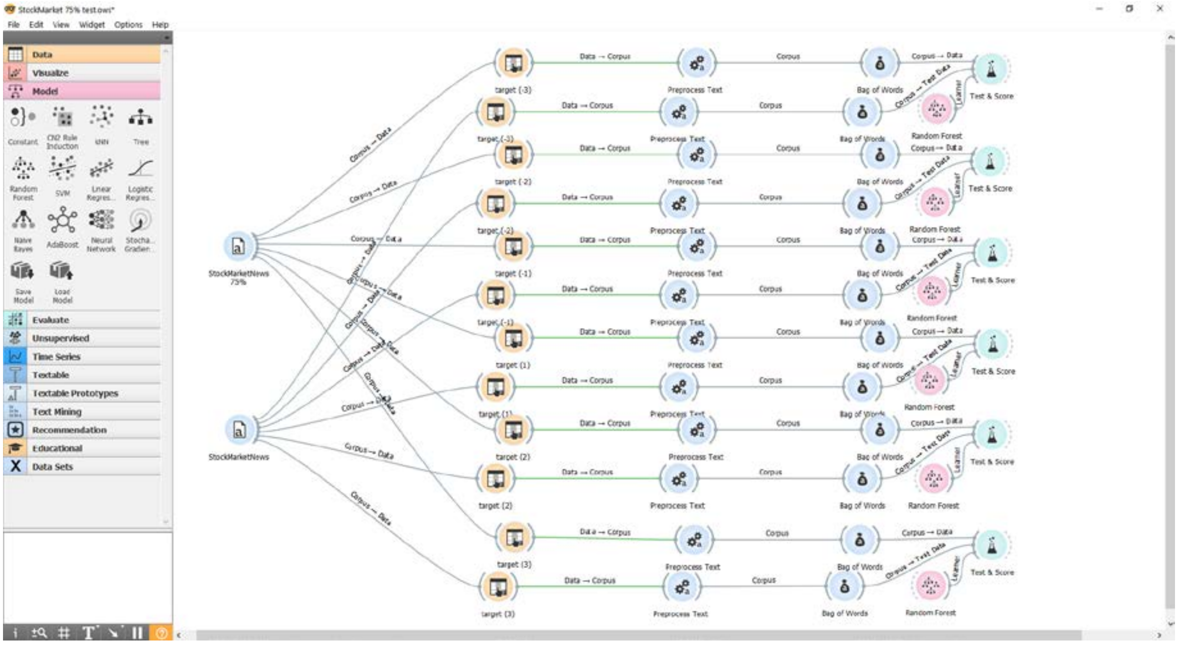


Figure 1: The model of solving the tasks of classification, which was developed in Orange.

3.2 Solving the tasks of classification

As a result, the model (figure 1) that was trained on 75% of data was developed to solve the task of classification. When checking on a retrospective data (for the 6 classes stated above) the method solved the task of classification correctly in 1826 out of 1949 cases (it corresponds to the probability of the correct classification which is more than 90%).

4 USING THE RESULTS OF THE SOLVED TASKS OF CLASSIFICATION TO FORECAST THE TIME SERIES VALUES

The main factor of the time series that describes the stock indices (such as NASDAQ) is the lack of seasonality, periodicity and known sequences. In this case the functional methods of forecasting [14] of the reference parameter do not work (the methods give bad results and do not pass the test of checking the adequacy of significant amount of steps – figure 2).

Our algorithm has to be built based on the following recurrent formula:

NASDAQ (t+1) = NASDAQ (t) + C, where C is the corrective coefficient that depends of the values of the expressions $3 \cdot S_g$, $2 \cdot S_g$, S_g, S_d, $2 \cdot S_d$, $3 \cdot S_d$, S_d and S_g (which are described in the 2nd part) and can be picked up based on the first values of the testing data set (or their parts).

The results were obtained using the recurrent formula (figure 3).

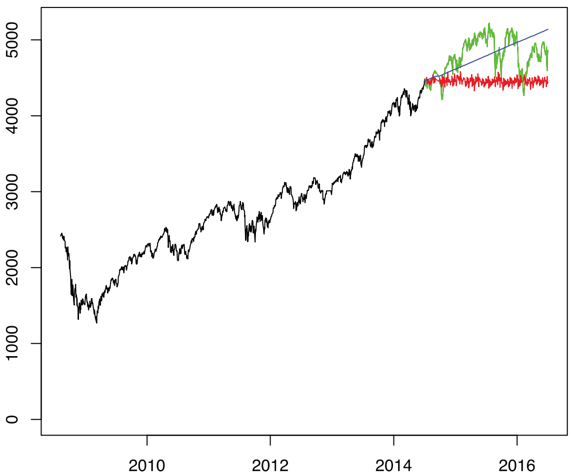


Figure 2: These changes of the NASDAQ index (black line – training sample, green line - testing sample) and the forecasts values that were obtained using the autoregressive method (blue line), fractal method (red line).

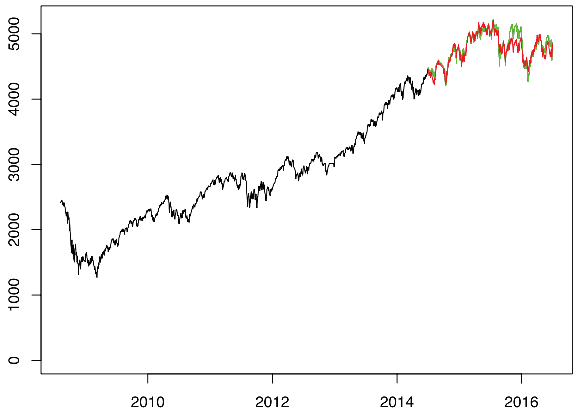


Figure 3: The results of forecasting the NASDAQ index (black line – learning sample, green line – testing sample) with the use of the offered recurrent formula (red line).

The Pearson's chi-squared test provides the best values of adequacy of the obtained result when checking the testing sample, narrowing of the confidence interval and greater forecasting horizon when using the offered recurrent dependence (Table 5).

Table 5: The results of the Pearson's chi-squared test.

Predictive model	The amount of the calculation steps in which the method stays adequate
Autoregressive method (ARIMA)	54
Fractal method	44
The offered recurrent method	338

In this case the offered method allows obtaining better results for a long period of time.

5 CONCLUSIONS

The results of the investigation showed the weak fitness of the forecasting functional methods used for the data description without any expressed sequences. At the same time, the results showed that the classification data could be used to solve the tasks of forecasting for which the algorithms on the 4th figure should be used.

The perspective of the use of classification methods for solving the tasks of forecasting and the opportunity of developing the forecast based on the use of testing data can be shown in this case.

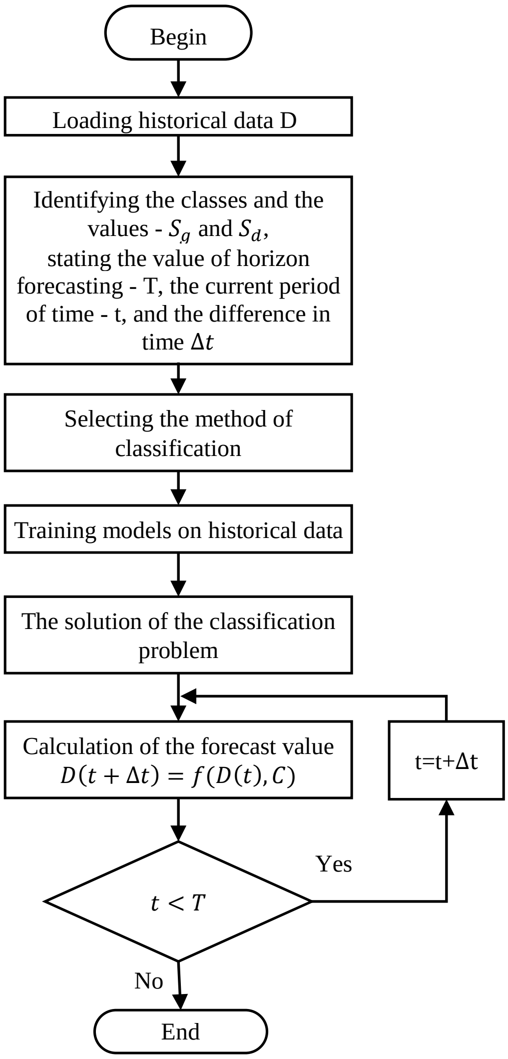


Figure 4: General scheme of using the classifiers in the time series forecasting algorithms.

ACKNOWLEDGMENTS

The authors thank the government of Perm Krai for the support of the project for "Development of software and economic and mathematical models for supporting innovation project management processes in production systems", which is being implemented in accordance with decree №166-п of 06.04.2011.

The reported study was partially supported by the Government of Perm Krai, research project No. C-26/058.

REFERENCES

[1] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, “Text mining for market prediction: A systematic review,” *Expert Syst. Appl.*, vol. 41, no. 16, pp. 7653–7670, Nov. 2014.

[2] F. Jin, N. Self, P. Saraf, P. Butler, W. Wang, and N. Ramakrishnan, “Forex-foreteller: currency trend modeling using news articles,” 2013, p. 1470.

[3] M. Butler and V. Kešelj, “Financial Forecasting Using Character N-Gram Analysis and Readability Scores of Annual Reports,” in *Advances in Artificial Intelligence*, vol. 5549, Y. Gao and N. Japkowicz, Springer Berlin Heidelberg, 2009, pp. 39–51.

[4] Y. Zhai, A. Hsu, and S. K. Halgamuge, “Combining News and Technical Indicators in Daily Stock Price Trends Prediction,” in *Advances in Neural Networks – ISNN 2007*, vol. 4493, D. Liu, S. Fei, Z. Hou, H. Zhang, and C. Sun, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 1087–1096.

[5] M.-A. Mittermayer, “Forecasting Intraday stock price trends with text mining techniques,” 2004, p. 10 pp.

[6] R. P. Schumaker and H. Chen, “Textual analysis of stock market prediction using breaking financial news: The AZFin text system,” *ACM Trans. Inf. Syst.*, vol. 27, no. 2, pp. 1–19, Feb. 2009.

[7] X. Zhou and Australasian Database Conference, Eds., *Database technologies 2002: proceedings of the Thirteenth Australasian Database Conference ; Monash University, Melbourne, January/February 2002*. Sydney: Australian Computer Society, 2002.

[8] T. Joachims, “Text categorization with Support Vector Machines: Learning with many relevant features,” in *Machine Learning: ECML-98*, vol. 1398, C. Nédellec and C. Rouveirol, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 137–142.

[9] S. Henrard, N. Speybroeck, and C. Hermans, “Classification and regression tree analysis vs. multivariable linear and logistic regression methods as statistical tools for studying haemophilia,” *Haemophilia*, vol. 21, no. 6, pp. 715–722, Nov. 2015.

[10] G. M. Di Nunzio, “Using scatterplots to understand and improve probabilistic models for text categorization and retrieval,” *Int. J. Approx. Reason.*, vol. 50, no. 7, pp. 945–956, Jul. 2009.

[11] T. M. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.

[12] Association for Computing Machinery, W. B. Croft, International Conference on Research and Development in Information Retrieval, and Trinity College Dublin, Eds., *SIGIR ’94: proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, 3 - 6 July 1994, Dublin, Ireland*. London: Springer, 1994.

[13] L. Mylnikov, B. Krause, M. Kuetz, K. Bade, and I. Shmidt, *Intelligent data analysis in the management of production systems (approaches and methods)*. Moscow: BIBLIO-GLOBUS, 2017.

[14] L. A. Mylnikov, A. V. Seledkova, and B. Krause, “Forecasting characteristics of time series to support managerial decision making process in production-And-economic systems,” *Proc. 2017 20th IEEE Int. Conf. Soft Comput. Meas. SCM 2017 6 July 2017*, pp. 853–855.

The Difference Between Precision-recall and ROC Curves for Evaluating the Performance of Credit Card Fraud Detection Models

Rustam Fayzrakhmanov, Alexandr Kulikov and Polina Repp
Information Technologies and Computer-Based System Department,
Perm National Research Polytechnic University,
Komsomolsky Prospekt 29, 614990, Perm, Perm Krai, Russia
fayzrakhmanov@gmail.com, thewato@gmail.com, polina.repp@gmail.com

Keywords: Credit Card Fraud Detection, Weighted Logistic Regression, Random Undersampling, Precision-Recall curve, ROC Curve

Abstract: The study is devoted to the actual problem of fraudulent transactions detecting with use of machine learning. Presently the receiver Operator Characteristic (ROC) curves are commonly used to present results for binary decision problems in machine learning. However, for a skewed dataset ROC curves don't reflect the difference between classifiers and depend on the largest value of precision or recall metrics. So the financial companies are interested in high values of both precision and recall. For solving this problem the precision-recall curves are described as an approach. Weighted logistic regression is used as an algorithm-level technique and random undersampling is proposed as data-level technique to build credit card fraud classifier. To perform computations a logistic regression as a model for prediction of fraud and Python with sklearn, pandas and numpy libraries has been used. As a result of this research it is determined that precision-recall curves have more advantages than ROC curves in dealing with credit card fraud detection. The proposed method can be effectively used in the banking sector.

1 INTRODUCTION

Fraud detection is generally considered as a data mining classification problem, where the objective is to classify the credit card transactions as legitimate or fraudulent correctly. Detection of fraudulent transactions combined with machine learning has become an exciting subject of research over the last years [1].

The credit card fraud exhibits unique characteristics which render the task extremely challenging for any machine learning technique. The most common characteristic is that the credit card datasets are highly unbalanced, which means they admit and uneven distribution of class transactions. The fraud class is represented by only a small number of examples (minority class) while the legal class makes up the rest (majority class). The ratio from legal class size to fraud class size can vary up to hundred fold [2]. Using these datasets as training sets in the learning process can bias the learning algorithm resulting in poor accuracy on the minority class but high accuracy on the majority class [3].

Approaches of solving the problem of unbalanced classes are divided into data-level methods and algorithm-level methods (or combinations of these techniques). Data-level methods are focused on modifying the training set to make it suitable for a standard learning algorithm. There are distinguish approaches which generate new objects for minority groups (oversampling) and which remove examples from majority groups (undersampling). Algorithm-level methods are focused on modifying existing learners to alleviate their bias towards majority groups. This requires a good insight into the modified learning algorithm and a precise identification of reasons for its failure in mining skewed distributions. The most popular branch is cost-sensitive approaches, such as weighted logistic regression [4].

To evaluate the performance these approaches [5][6] use Receiver Operator Characteristic (ROC) curves, which show how the number of correctly classified positive examples varies with the number of incorrectly classified negative examples. However, ROC curves can present an overly optimistic view of an algorithm's performance if there is a large skew.

Precision-Recall (PR) curve is an alternative to ROC curves for tasks with a large skew in the class distribution, such as a credit card fraud. Precision-recall curves are highly informative about the performance of binary classifiers, and the area under these curves is a popular scalar performance measure for comparing different classifiers [7].

In this article, a model for detecting a credit card fraud using weighted logistic regression and random undersampling techniques was built and ROC and PR curves for them were analysed.

2 EVALUATION OF A CLASSIFICATION MODEL

The aim of detection a credit card fraud is to design a binary classifier with a highest possible accuracy of fraud transactions. To design it many different machine learning technics are used; the most wide-spread of them are logistic regression, decision trees, support vector machine, its varieties and assembles. In this case the sets of data (containing dozens and hundreds of features) have become online payment transactions belonged to financial companies. Features are different information about an online purchase, such as the transaction's amount, IP-address, payment card data, etc. Since fraud transactions usually present less than 1% of the total number of transactions, the process of a classifier design is called imbalance learning, and the data is called imbalance dataset.

Since the credit card fraud task is binary a confusion matrix to evaluate a performance of approaches is used. The confusion matrix summarizes information about actual and predicted classifications performed by a classifier. Confusion matrix for binary classifiers is shown in Table 1. The table shows that four different forecast results are possible. Really positive and really negative outcomes are the correct classification, while the false positive and false negative outcomes are two possible types of errors [8].

Table 1: Confusion matrix.

Actual	Predicted	
	Positive class	Negative class
Positive class	True Positive (TP)	False Negative (FN)
Negative class	False Positive (FP)	True Negative (TN)

A false positive example is a negative example class that is wrongly classified as a positive one (legitimate transactions as fraudulent in context of the paper) and a false negative example is a positive example of the class that is wrongly classified as a negative (fraudulent as legitimate) one.

Standard performance metrics such as predictive accuracy and error rate can be derived from the confusion matrix:

$$Predictive\ Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Error\ Rate = \frac{FP + FN}{TP + FP + TN + FN}$$

The usage of a predictive accuracy and error rate leads to a poor performance for the minority class [9]. For that reason, a variety of common evaluation metrics based on confusion matrix are developed to assess the performance of classifiers for imbalanced data sets:

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN}$$

These metrics are developed from the fields of information retrieval. They are used in situations when performance for the positive class (the minority class) is preferred, since both precision and recall are defined with respect to the positive class.

Alternatively, the Receiver Operating Characteristic (ROC) can be employed to evaluate the overall classification performance. The ROC is a graphical representation that plots the relationship between the benefits (TPR) and costs (FPR) as the decision threshold varies. The ROC curve provides evidence that the true positive rate is directly proportional to the false positive rate [10].

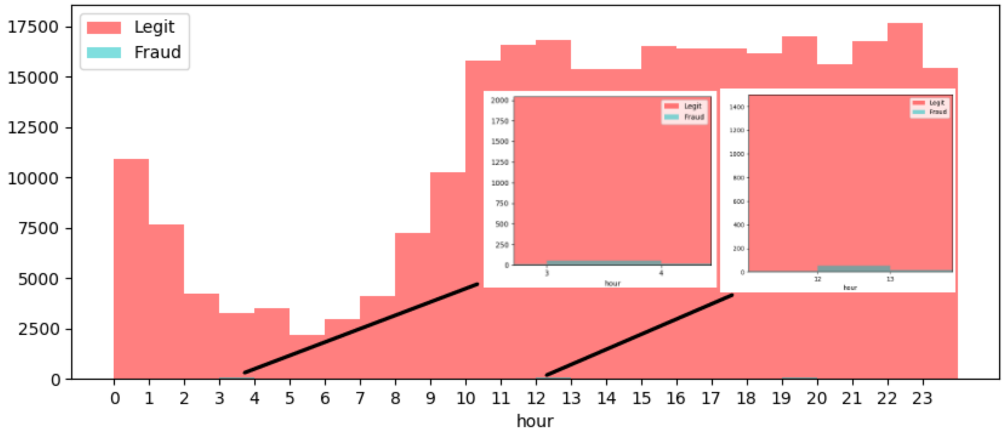


Figure 1: Distribution of the dataset.

Precision-recall (PR) curves, like the ROC curves, are an evaluation tool for binary classification that allows performance visualization. PR curves are increasingly used in the machine learning community, particularly for imbalanced datasets. On these imbalanced or skewed data sets, PR curves are a useful alternative to ROC curves that can highlight performance differences that are lost in ROC curves.

The area under curve (AUC) measure summarizes the performance of the classifier into a single quantitative measure, usually to determining what classifier is more superior. Generally, a better performing classifier has a larger AUC than that of an inferior one.

ROC and PR curves facilitate clear visualization comparisons between two or more classifiers over a large span of operating points.

Financial companies don't want to miss catching fraud (FN), therefore recall is important. However, it is necessary to consider that an accuracy lost (FP) is also money lost for companies, because they have to call the customer and verify that the purchase was authentic indeed which takes resources. Therefore, it is important to obtain high precision and recall values for the classifier.

3 EXPERIMENTS

Consider the dataset that contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where there are 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) amounts 0.172% of all transactions. It contains only numerical input variables which are the result of a principal component analysis (PCA) transformation. Due to confidentiality issues, there is no possibility to obtain the original features and more background information about the data. Features V1-V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise. Dataset is illustrated in Figure 1.

	Time	V1	V2	...	V28	Amount	Class
0	0.0	-1.359807	-0.072781	...	-0.021053	149.62	0
1	0.0	1.191857	0.266151	...	0.014724	2.69	0
2	1.0	-1.358354	-1.340163	...	-0.059752	378.66	0
3	1.0	-0.966272	-0.185226	...	0.061458	123.50	0
4	2.0	-1.158233	0.877737	...	0.215153	69.99	0

Figure 2: Dataset example.

The distribution of the dataset is illustrated in Figure 2. The data is totally unbalanced. This is a clear example where a typical accuracy score to evaluate our classification algorithm is used. For example, in case having just used a majority class to assign values to all records, a high accuracy still will be had, but all fraudulent transactions would be classified incorrectly.

To perform computations a logistic regression as a model for prediction of fraud and Python with sklearn, pandas and numpy libraries has been chosen. Consider the confusion matrix, precision and recall metrics on the raw dataset. The matrix is illustrated in Figure 3.

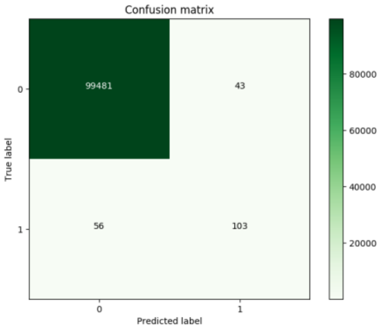


Figure 3: Confusion matrix of a model.

The recall of this model is 0.64 and precision is 0.71. These are fairly low scores. This is due to the fact that logistic regression as a standard classifier algorithm has bias to classes that have a number of instances. They tend only to predict data of most classes. The characteristics of the minority class are considered as noise and are often ignored. Thus, there is a high probability of mistaken classification of the minority class in comparison with the majority class [11]. This problem can be solved by algorithms of a family of decision trees, such as a random forest, but such algorithms are not stable to high overfitting [12].

To solve the unbalanced problem, a weighted logistic regression as an algorithm-level method and random undersampling as a data-level method was used.

Weighting is a procedure that weights the data to compensate the differences in a sample and population. In rare events, such as a credit card fraud, we tend to sample all the 1's (rare events) and a fraction of 0's (non-events). In such cases the observations have to be weighed accordingly.

Some arbitrary weights for a model to illustrate the tradeoff between precision and recall are specified. The weights to n = {1, 5, 10, 25, 50, 100, 500, 1000, 10000} are set. The results are shown in Table 2.

Table 2: Results of the model with different weight parameters.

Weight	Precision	Recall
1	0.65	0.71
5	0.68	0.71
10	0.77	0.65
25	0.81	0.41
50	0.84	0.46
100	0.85	0.27
500	0.90	0.08
1,000	0.94	0.04
10,000	0.97	0.005

Clustering, as an effective data-level technique [13], can be used. However, since the dataset has anonymous data, random undersampling is a better choice. Undersampling is one of the techniques used for handling class imbalance. In this technique, we under sample the majority class to match the minority class. So in our case, a random sample of non-fraud class to match number of fraud samples is taken. This makes sure that the training data has equal amount of fraud and non-fraud samples [14]. And then the model to the whole dataset is applied.

For undersampling random 25%, 10% and 1% legitimate samples of dataset are taken as well as random 492, 984 and 1476 legitimate samples (1x, 2x and 3x of fraud samples). The results are shown in Table 3.

Table 3: Results of the model with different random legitimate samples.

Samples	Precision	Recall
56862	0.81	0.81
28431	0.74	0.82
2843	0.28	0.88
1476	0.12	0.90
984	0.11	0.90
492	0.04	0.93

Due to the manually selecting a range of weights to boost the minority class and undersampling minority class our model has been improved to have a better recall, and in some cases, a better precision also. Recall and precision are usually tradeoffs of each other, so when both are improved at the same time, our model's overall performance is undeniably improved.

4 ANALYSIS OF PR AND ROC CURVES

For financial companies, as it has earlier been mentioned, both the high accuracy and the high completeness are important. To calculate the specific values of these metrics, different companies develop their own evaluation algorithms based on their financial strategy or use universal ones like Economic Efficiency [15]. Thus, for our

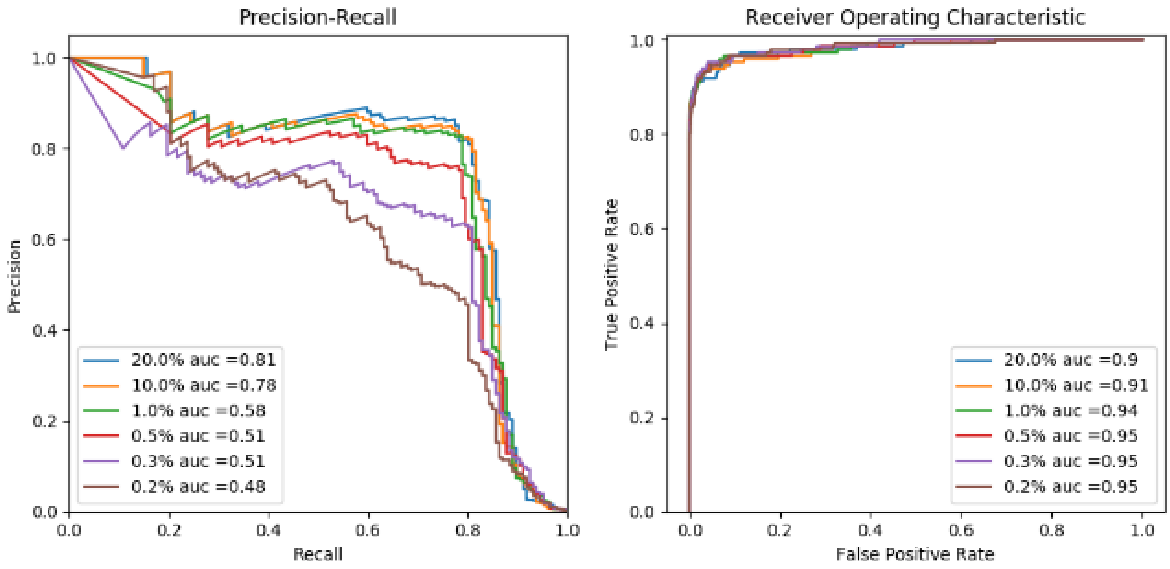


Figure 4: PR and ROC curves for random undersampling technique.

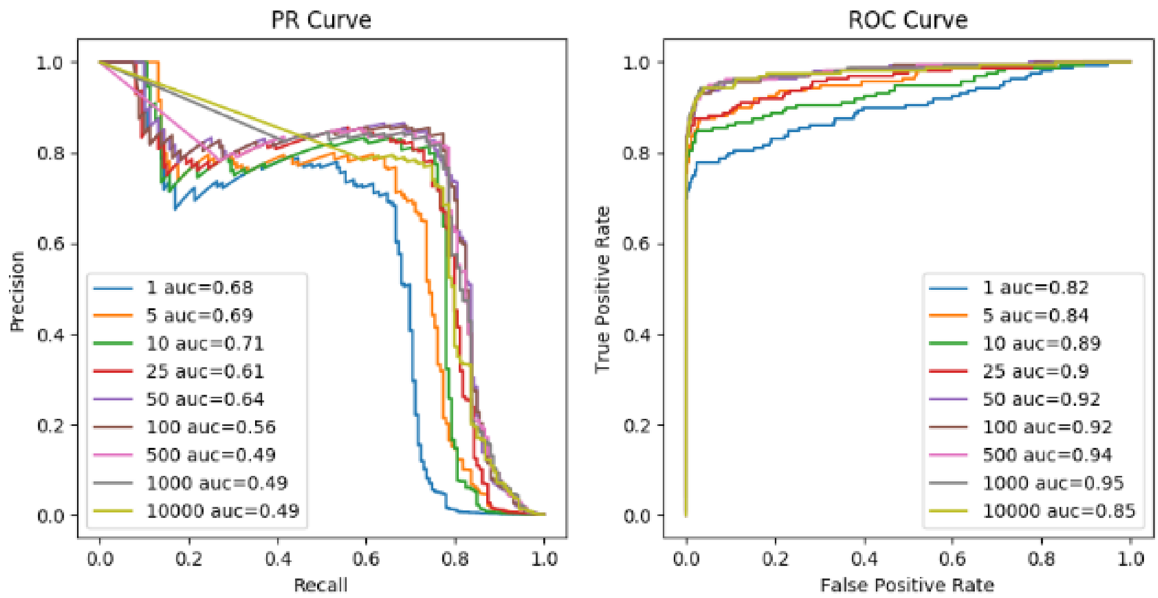


Figure 5: PR and ROC curves for weighted logistic regression.

calculations, a combination of the most possible values of precision and recall is used. To do this PR and ROC curves for both techniques (weighted logistic regression and random undersampling) are built and the area under curves (AUC) as a metric to evaluate both precision and recall is calculated. Plots of curves are illustrated on Figure 4 and Figure 5

For a PR curve, a good classifier aims at the upper right corner of the chart but the upper left corner aims at the ROC curve.

While PR and ROC curves use the same data, i.e. the real class labels and predicted probability for the class labels, different behaviour is observed, with some weights and samples seem to perform better in ROC than in the PR curve. This difference exists because the number of legitimate transactions greatly exceeds the number of fraud transactions in this domain. Consequently, a large change in the number of false positives can lead to a small change in the false positive rate used in ROC analysis. Precision, on the other hand, by comparing false positives to true positives rather than true negatives, captures the effect of the large number of negative examples on the techniques performance [16].

Such a difference can lead to wrong conclusions. In case to evaluate using only ROC curves for undersampling technique, it is seen that reducing the examples of the majority class to the size of the minority class leads to a better performance of the model. But according to Section 3, if the kind of the transformation is made the maximum recall will be really obtained, at the same time a low precision (20 false positives occur for each fraud) will be received. PR curves reflect the real picture: 20% of the majority class leads to the maximum possible values of recall and precision.

A similar situation is observed for the weighing technique. Using ROC curves, we can see that the maximum efficiency is achieved with weights of 500-1000, whereas PR curves show the maximum efficiency for 5-10. Compared these values with the obtained values of precision and recall, it is valid at weights 5-10 that precision and recall have the most effective values.

Precision is directly influenced by class imbalance since FP is affected, whereas TPR only depends on positives. That is why ROC curves do not capture such effects. Therefore, for cases where both precision and recall are important, in skewed data, such as credit card fraud detection, PR curves have a greater advantage over ROC curves.

5 CONCLUSIONS

As a result of this research it is determined that precision-recall curves have more advantages than ROC curves in dealing with credit card fraud detection. Area under precision-recall curve correlates with actual values precision and recall for both algorithm-level and data-level techniques. Since the credit card fraud is an unbalanced task, where the ratio of classes is less than 1%, ROC curve doesn't capture the effect of improving both precision and recall metrics.

Future research will be concentrated on improving machine learning methods to detecting credit card fraud using precision-recall curve as a main metric to evaluating performance of classifiers.

REFERENCES

[1] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, “Feature engineering strategies for credit card fraud detection,” Expert Systems with Applications, vol. 51, pp. 134–142, 2016.

[2] Lu Q., Ju C., “Research on credit card fraud detection model based on class weighted support vector machine,” Journal of Convergence Information Technology, vol. 6, no. 1, Jan. 2011.

[3] Monard, M.C., Batista, G.E., “Learning with skewed class distributions,” Advances in Logic, Artificial Intelligence and Robotics (LAPTEC'02), pp. 173-180, 2002.

[4] Zhou, Zhi-Hua, and Xu-Ying Liu, " On multi-class cost-sensitive learning," Computational Intelligence, vol. 26, no. 3, pp. 232-257, 2010.

[5] Dal Pozzolo, Andrea, et al, "Learned lessons in credit card fraud detection from a practitioner perspective," Expert systems with applications, vol .41, no. 10, pp. 4915-4928, 2014.

[6] Anis, Maira, Mohsin Ali, and Amit Yadav, "A comparative study of decision tree algorithms for class imbalanced learning in credit card fraud detection," International Journal of Economics, Commerce and Management, vol. 3, no. 12, 2015.

[7] Keilwagen, Jens, Ivo Grosse, and Jan Grau, "Area under precision-recall curves for weighted and unweighted data," PLoS One, vol. 9, no. 3, 2014.

[8] Novaković, Jasmina Dj, et al, "Evaluation of Classification Models in Machine Learning," Theory and Applications of Mathematics & Computer Science, vol. 1, no. 1, pp. 39-46, 2017.

[9] Fawcett, Tom, "An introduction to ROC analysis," Pattern recognition letters, vol. 27, no. 8, pp. 861-874, 2006.

[10] Yong, Terence Koon Beh, Chuan Tan Swee, and Theng Yeo Hwee, "Building classification models from imbalanced fraud detection data," Malaysian Journal of Computing, vol. 2, no. 2, pp. 1-21, 2014.

[11] King, Gary, and Langche Zeng, "Logistic regression in rare events data," Political analysis, vol. 9, no. 2, pp. 137-163, 2001.

[12] Fayzrakhmanov, R. A., Kulikov, A.S., et al, "Prediction of the need for narcotic and psychotropic drugs in the region using random forest model," artificial intelligence in solving actual social and economic problems of the XXI century, pp. 108-111, 2016.

[13] Fayzrakhmanov, R. A., et al, "Aplplication of cluster analysis in developing approaches to the selection and designation of treatment regimens for HIV-infected patients," Bulletin of Siberian Medicine, vol. 16, no. 3, pp. 52-60, 2017.

[14] Bhattacharyya, Siddhartha, et al, "Data mining for credit card fraud: A comparative study," Decision Support Systems, vol. 50, no.3. pp. 602-613, 2011.

[15] Lima, Rafael Franca, and Adriano CM Pereira, "Feature Selection Approaches to Fraud Detection in e-Payment Systems," International Conference on Electronic Commerce and Web Technologies, Springer, Cham, 2016.

[16] Davis, Jesse, and Mark Goadrich, "The relationship between Precision-Recall and ROC curves," Proceedings of the 23rd international conference on Machine learning. ACM, 2006.

Influence of Fuzzy Clustering on the Accuracy of Electrical Equipment Diagnostic Models

Denis Eltyshev and Ksenia Gnutova
Electrotechnical Department, Perm National Research Polytechnic University, Komsomolsky Ave. 29, Perm, Russia
eltyshev@msa.pstu.ru, k.gnutova@gmail.com

Keywords: Diagnostic, Electrical Equipment, Fuzzy Clustering, Fuzzy logic, Membership function

Abstract: The development of electric power industry is oriented on high reliability, flexibility and efficiency of managing power grids of arbitrary configuration. For such grids an information infrastructure is required. It should consist of various software and hardware, including systems of electrical equipment monitoring and diagnostics for accumulating information about its parameters with controlling and managing its condition. As a rule, data about the electrical equipment is heterogeneous. Thus there is a necessity of certain mechanisms for data processing to provide a possibility of constructing diagnostic models in an automated mode and adapting them to power grid operating conditions. The aim of this work is to develop a mechanism for automated calculation of the electrical equipment diagnostic models parameters. It supposes using historical data analyzing algorithms that ensure high reliability of the diagnosis. Base on this the application of fuzzy clustering for constructing membership functions to assess the features of equipment condition change in fuzzy diagnostic models was considered. Different fuzzy clustering algorithms were analyzed, and a technique for processing data on the equipment operation with constructing membership functions based on fuzzy partition matrix and clusters centres was proposed. The technique allows to approximate the membership functions by typical curves and to choose the most effective variant of clustering in terms of electrical equipment diagnostic reliability. The testing of fuzzy models for assessing the condition of power transformer equipment using clustering results was performed. A high level of compliance of simulation data with the conclusions of specialized organizations performing monitoring of equipment in power supply systems for oil production facilities was obtained. The practical relevance of the results is in using the technique in the synthesis of intelligent expert-diagnostic systems for increasing the electrical equipment diagnostic reliability and reducing the duration of its unplanned downtime.

1 INTRODUCTION

The development of efficient new generation electric power systems based on the Smart Grid technology with high reliability, fault tolerance, flexibility and adaptability is associated with the introduction of modern information and telecommunication technologies [1] - [3]. One of the tasks of such systems is to provide control over the condition of electrical equipment (EE), maintain its operability and operatively manage its operation modes. Modern information and diagnostic systems for assessing the EE condition should be integrated with monitoring systems to collect and accumulate data necessary for subsequent processing and analysis. Since this possibility can not always be ensured by technical and economic factors, to obtain information about the object both stationary and mobile software and hardware can be used [1][2], [4][5]. Considering the heterogeneity and ambiguousness of the EE operating data it is advisable to use intelligent technologies for constructing diagnostic systems, based on statistical data and the experience of qualified experts [4] – [8].

2 SETTING GOALS AND OBJECTIVES OF THE STUDY

The intelligent diagnostic procedure in general can be represented as the defining of the ratio [6][7]:

$$Y \in (y_1, y_2, \ldots, y_n) \rightarrow \mathbf{X} = (x_1, x_2, \ldots, x_m) \quad (1)$$

where Y is the set of classes of EE condition or types of defects in equipment elements; $\mathbf{X}$ is the vector of controlled technical parameters (diagnostic features); → is a set of rules linking the values of the parameters with the level of EE condition.

To solve the problem that is difficult to formalize, models based on fuzzy logic can be used. In this case it is necessary: to define for each variable x_i, i=1:n the number of terms for their verbal estimates (for example, the vibration level is “low”, “medium”, “high”); describe each term as a membership function (MF); construct a rule base that connect the values of the variables **X** with the class y_j, j=1:m. As a rule, these operations are carried out manually, by an expert, which does not provide the necessary flexibility of diagnostic models.

We can automate this process using the available data on the EE operation, or the data generated by the monitoring system.

In order to determine the parameters of the MF in automated mode, as well as to build the rule base, it is proposed to apply methods of data mining, in particular, fuzzy clustering [8] -[10]. This will greatly simplify the process of building expert-diagnostic systems, oriented on the assessing and management of the EE condition.

The aim of the work is to research mechanism of using fuzzy clustering algorithms for automated defining the MF in the EE diagnostic models that ensure high reliability of the diagnosis. The tasks include: developing the technique of fuzzy clustering based on the EE operating data for defining the MF parameters with approximation of MF by typical curves; estimation the influence of clustering algorithms on accuracy of EE diagnostics models.

3 DESCRIPTION OF THE RESEARCH METHODOLOGY

The procedure of the EE operating data analysis for synthesising expert-diagnostic models with automated defining of a MF can be represented as follows (Figure 1).

At the first stage (block 1) an array of initial data on the measurements of each monitored diagnostic parameter is formed.

Fuzzy clustering consists in determining the coordinates of clusters centers **C** = (c_1, c_2, ..., c_m) and the matrix **U** = [u_{ij}] showing the belonging of parameter measurements for each cluster (block 3).

Let’s consider the process of clustering using the fuzzy c-means algorithms (FCM), Gustafson-Kessel algorithms (GK) and Gath-Geva algorithms (GG) [11-13]. The difference between the methods is in the objective function and special metric, which makes it possible to allocate clusters of different shapes. For example, the FCM method minimizes the following objective function:

$$J(Z;U,V)=\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^{m}\left\|z_k-v_i\right\|_A^2 \qquad (2)$$

where z_k is the data array, m is the exponential weight, μ_{ik} is the partition matrix, c is the number of clusters, and **A** is the diagonal matrix.

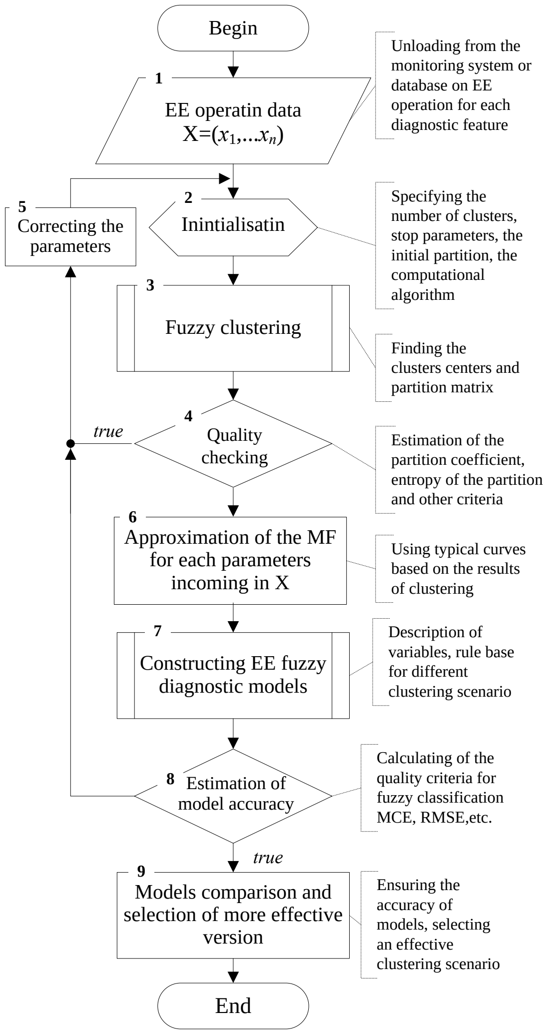


Figure 1: Block diagram of the technique for using fuzzy clustering in constructing EE diagnostics models.

The GK method uses the adaptive norm for each j-th cluster due to the individual matrix $\mathbf{A}_i$. It is a fuzzy covariance matrix of the cluster changing during the iteration process with the following objective function:

$$J(Z;U,V,\{A_i\}) = \sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^{m} D_{ikA_i}^{2} \quad (3)$$

where $\mathbf{A}_i$ is a fuzzy covariance matrix of the cluster.

The GG method uses the Gaussian distribution, and minimizes the sum of the squares between the data points and the prototype η_i of the being formed group.

In general the clustering algorithm can be divided into two stages: 1) initialization stage (block 2) when the membership matrix for all elements of the input set is randomly populated and the necessary parameters (number of clusters, exponential weight etc.) are selected. 2) The calculation stage (block 3), when the following steps are performed for each iteration:

1) Finding cluster centres:

$$v_i^{(l)} = \frac{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m z_k}{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m}, \; i = \overline{1,c} \quad (4)$$

2) Finding the covariance matrix of the cluster for the GK and GG methods:

$$F_i = \frac{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m \left(z_k - v_i^{(l)}\right)\left(z_k - v_i^{(l)}\right)^T}{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^m}, \; i = \overline{1,c} \quad (5)$$

3) Calculation of the distances between the new cluster centres and the points of data set. For the FCM method, the Euclidean distance is used:

$$D_{ik\mathbf{A}}^2 = \left(z_k - v_i^{(l)}\right)^T \mathbf{A}\left(z_k - v_i^{(l)}\right), \; i = \overline{1,c}, \; k = \overline{1,N} \quad (6)$$

For the GK method:

$$D_{ik\mathbf{A}_i}^2 = \left(z_k - v_i^{(l)}\right)^T \left[\rho_i \det(F_i)^{\frac{1}{n}} F_i^{-1}\right]\left(z_k - v_i^{(l)}\right), \quad (7)$$

For the GG method:

$$D_{ikA_i}^2 = \frac{(2\pi)^{\left(\frac{n}{2}\right)}\sqrt{\det(F_i)}}{a_i} \exp\left(\frac{1}{2}\left(z_k - v_i^{(l)}\right)^T F_i^{-1}\left(z_k - v_i^{(l)}\right)\right) \quad (8)$$

$$a_i = \frac{1}{N}\sum_{k=1}^{N}\mu_{ik} \quad (9)$$

4) Recalculation of the partition matrix.

The FCM method uses the condition:

If $D_{ik\mathbf{A}} > 0$, then

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c}\left(D_{ik\mathbf{A}} / D_{jk\mathbf{A}}\right)^{\frac{2}{(m-1)}}} \quad (10)$$

otherwise

$$\mu_{ik}^{(l)} = 0, \text{ if } D_{ikA} > 0 \text{ and}$$

$$\mu_{ik}^{(l)} \in [0,1], \; \sum_{i=1}^{c}\mu_{ik}^{(l)} = 1. \quad (11)$$

For GK and GG methods the partition matrix is recalculated in the same way, but $D_{ik\mathbf{A}}$ is used instead of D_{ikA_i}.

5) Stopping the algorithm. The fuzzy clustering algorithm stops when the next condition is fulfilled:

$$\max_{k=\overline{1,M},\, i=\overline{1,c}}\left(\left|\mu_{ki} - \mu_{ki}^{*}\right|\right) < \varepsilon, \text{ or } \max_{i=\overline{1,c}}\left(\left|V_i - V_i^{*}\right|\right) < \varepsilon \quad (12)$$

To assess the effectiveness of clustering a complex analysis of the results quality is carried out (block 4). To construct the MF we use the obtained coordinates of the cluster centers and the fuzzy partition matrix (block 6). For these purposes the technique of MF approximation by typical curves (for example, Gaussian, bell-shaped, pi-like) is used.

In accordance with the results of the MF construction fuzzy models, connecting the variables **X** and *Y* (block 7), are formed, the rules base is synthesized and the accuracy of the models by the criteria of the mean classification error (MCE), root mean square error (RMSE) between model and experimental data, or complex criteria [14] is determined.

4 INVESTIGATION OF THE METHODOLOGY

Let's consider (Figure 2) a simplified example of using fuzzy clustering in the problem of assessing the thermal condition of the power oil-filled transformer equipment elements (POTE). The transformer TDNT-16000/110-U1 with a voltage of 110/35/6 kV, which is typical equipment for power supply systems of the oil-producing fields in the Perm Krai was chosen as the object of the study.

We considered a fuzzy diagnostic model with input parameters $\mathbf{X} = (x_1, x_2)$, characterizing the condition of the transformer tank, where x_1 is the excess temperature of the bolted tank bell connections; x_2 is the maximum temperature difference over the surface of the tank and the elements of the cooling system [7].

We used the results of thermal imaging control of transformers from Perm Krai oil production facilities for the period 2010-2013 as the initial data. Also we supplemented initial data by simulation modelling, taking into account the boundary values of the transformer parameters given in the technical documentation for the electrical equipment operation (sample size - 250 measurements).

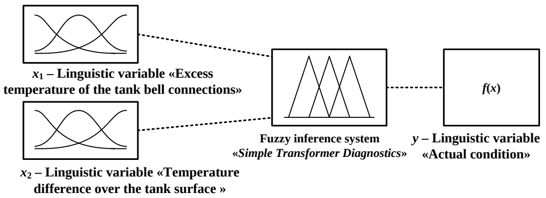


Figure 2: The structure of the fuzzy model for diagnostic POTE condition in the Fuzzy Logic Toolbox of the MATLAB.

We formed data set on the basis of monitoring protocols compiled by specialized service organizations using the thermograms of the transformer elements. In accordance with the technique of the POTE diagnostics using by the service organization and the recommendations of the normative and technical documentation we selected the classes “No defects” (1), “Developed defect” (2), “Critical defect” (3) for assessing the condition of the transformer tank *Y*. A fragment of the initial data is given in Table 1.

Table 1: Representation of the initial data for the POTE thermal diagnostic model.

№	x_1, °C	x_2, °C	Class of condition	Conclusion on the condition
1	4.9	25	1	No thermal defects
16	15	24.1	2	Developed defect (increased contact heating)
67	18	13	2	Developed defect (increased contact heating)
155	2.2	5	2	Developed defect (reduced circulation of oil in the tank)
184	32	20	3	Critical defect (contact overheating)
199	1.7	27.1	1	No thermal defects
246	7	39	3	Critical defect (increased tank heating)

We used 70% of data sample and three clusters for data processing by all fuzzy algorithms (FCM, GK, GG). The results of clustering (ordered and normalized data) are shown on Figure 3. The efficiency of clustering was estimated by partition coefficient (PC) and entropy (CE).

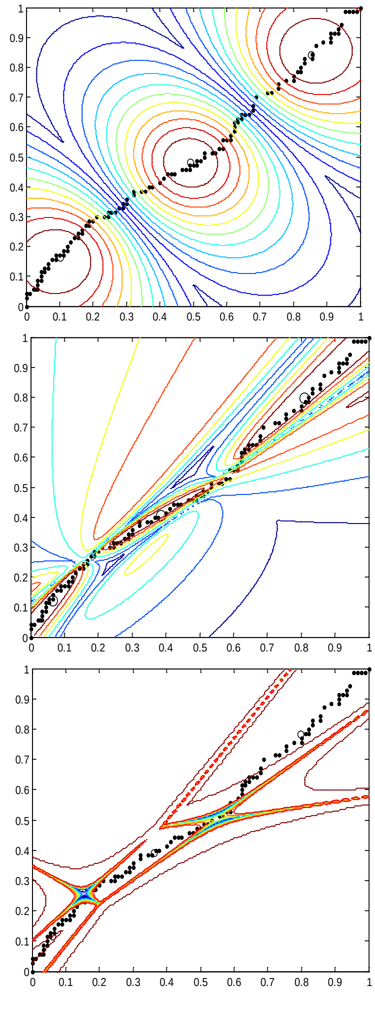


Figure 3: Fuzzy clustering of data on the POTE elements temperature, FCM (PC=0.82, CE=0.34) (a), GK (PC=0.91, CE=0.18) (b), GG (PC=0.96, CE=0.07) (c).

In order to construct fuzzy models we used the fuzzy partition matrixes for each cluster and constructed MF with three terms for assessing POTE elements temperature: “Low” (L), “Medium” (M), and “High” (H). The example of constructing the MF using fuzzy partition matrix and after approximation of the clustering results by Gaussian curves is shown in Figure 4.

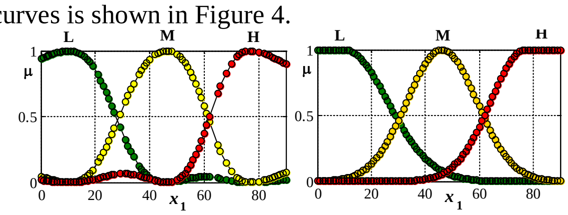


Figure 4: Constructing MF using fuzzy partition matrix (a) and approximation (b).

The rule bases of POTE fuzzy diagnostic models consist of 9th rules of following type: IF x1 = “Low” AND x2 = “Low”, then y = “No defects”.

5 RESULTS OF MODEL ACCURACY ESTIMATION

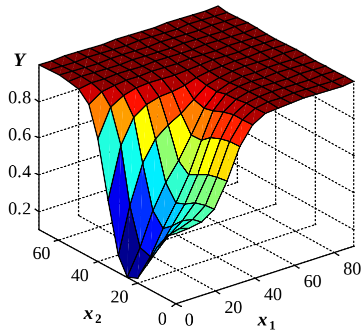
To assess the accuracy of fuzzy diagnostic models obtained by using clustering algorithms the simulation results and the conclusions of the service organizations from POTE monitoring protocols were compared.

We used 70% of data sample (190 positions) to learn fuzzy models and set the weights of the rules from the rule base by the gradient methods and complex criterion [14].

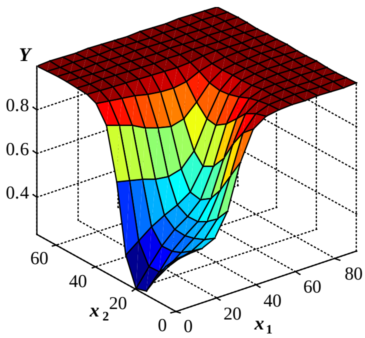
During the models adequacy verification using testing data set (30% of data sample) by χ^2 criterion we got the following results: χ^2 = 7.33 for model using FCM, χ^2 = 4.33 for model using GK, χ^2 = 6.01 for model using GG.

When a critical value of χ^2 = 34.77 the hypotheses of the models adequacy are accepted with 100% probability.

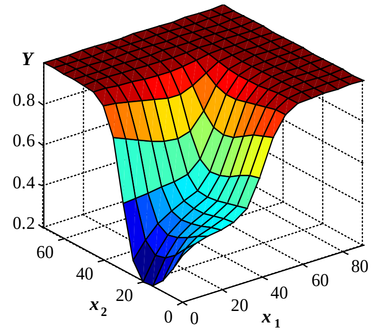
It can be seen from Figure 5 that the response surfaces of the models constructed by clustering algorithms differ little from each other.



a)



b)



c)

Figure 5: Surfaces of response of models, FCM (a), GK (b), GG (c).

To assess the effectiveness of clustering a mean fuzzy classification error (MCE) was calculated (Figure 6) on various data sets. We should note that for different cluster algorithms the number of correctly recognized classes (defects in equipment) varies from 89.2 to 92.4 %, which is a pretty good result. Thus using cluster algorithms for MF constructing provides small error of fault diagnosis of PTOE condition. For GG algorithm the error is less by 3.2% than for traditional FCM algorithm.

Reduction of errors in recognition of defects will allow in practice to reduce the equipment downtime both for the cause of the accident and according to plan. In the future, to confirm the effectiveness of the technique, it can be compared with other classification methods (for example, neural networks) in assessing the EE condition.

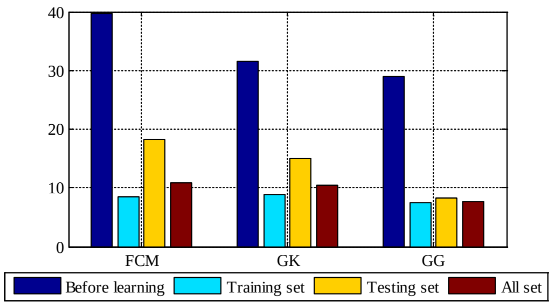


Figure 6: Results of the evaluation of the accuracy of fuzzy models of diagnostics of the POTE. The average number of misidentified MCE states, %.

It can be seen from the results that the application of clustering for the construction of the MF provides the accuracy of the diagnosis to 92.4% and more efficiently than the use of the expert method. At the same time there is an increase in the reliability of the assessment of the state by 8-12%.

6 CONCLUSIONS

Let us consider the most important results of the work:

1. The use of fuzzy clustering allows building the MF in automated mode, minimizing the expert's participation and taking into account the specifics of the electrical equipment operation contained in the available statistical data.

2. We can see deterioration or improvement in the quality of the equipment diagnostic models depending on the fuzzy clustering parameters.

3. The application of both fuzzy clustering and fuzzy modeling gives high results in the reliability of the EE condition assessment and provides adaptability of diagnostic models with possibility of the data volume increasing.

We should note that it is necessary to have an implemented technology for monitoring the EE condition and a database aggregating the results of the monitoring with processing and converting the data to a convenient form for fuzzy clustering and classification in order to provide the effectiveness of the technique. The technique is sensitive to initial data, clustering parameters, and algorithms for fuzzy models training.

We can define following directions of the further researches: formation of the rule base for diagnostic models in the automated mode; an analysis of ways to improve the accuracy of diagnostic models, including valid choice of the curve type for MF approximation and algorithms for model parameters setting; approbation of the technique on other EE diagnostics problems.

The results can be further used in the automated synthesis of expert-diagnostic models in intelligent systems for assessing the EE condition.

REFERENCES

[1] A.V. Kychkin, G.P. Mikriukov, "Applied data analysis in energy monitoring system," Journal of Problems regional energy, vol. 2(31), pp. 84-92, 2016.

[2] A.V. Kychkin, "Synthesizing A System For Remote Energy Monitoring In Manufacturing," Metallurgist, vol. 59, № 9-10, pp. 752-760, 2016.

[3] A.B. Petrochenkov, T. Frank, A.V. Romodin, A.V. Kychkin, "Hardware-in-the-loop Simulation of an Active-adaptive Power Grid," Russian Electrical Engineering, vol. 84, № 11, pp. 652-658, 2013.

[4] N.I. Khoroshev, V.P. Kazantsev, "Management Support of Electroengineering Equipment Servicing Based on the Actual Technical Condition," Automation and Remote Control, vol. 76, № 6, pp. 1058-1069, 2015.

[5] A.B. Petrochenkov, S.V. Bochkarev, A.V. Romodin, D.K. Eltyshev, "The planning operation process of electrotechnical equipment using the Markov process theory," Russian Electrical Engineering, vol. 82, № 11, pp. 592‐596, 2011.

[6] D.K. Eltyshev, V.V. Boyarshinova, "Intelligent Decision Support in the Electrical Equipment Diagnostics," Proceedings of the 19th International Conference on Soft Computing and Measurements, SCM 2016, pp. 157-160, 2016.

[7] D.K. Eltyshev, N.I. Khoroshev, "Diagnostics of the Power Oil-filled Transformer Equipment of Thermal Power Plants," Thermal Engineering, vol. 63, № 8, pp. 558-566, 2016.

[8] N.I. Khoroshev, R.N. Pogorazdov, "Adaptive Clustering Method in Intelligent Automated Decision Support Systems," Proceedings of the 19th International Conference on Soft Computing and Measurements, SCM 2016, pp. 296-298, 2016.

[9] K.A. Gnutova, D.K. Eltyshev, "Using Cluster Analysis in the Synthesis of Electrical Equipment Diagnostic Models," Proceedings of the 5th International Conference on Applied Innovations in IT 2017, pp. 119-124, 2017.

[10] D.K. Eltyshev, K.A. Gnutova, "Identification of the parameters of electrical equipment condition assessment models using fuzzy clustering," Proceedings of the 20th IEEE International Conference on Soft Computing and Measurements, SCM 2017, pp. 142-144, 2017.

[11] J.C. Bezdek, R. Ehrlich., W Full, "FCM: The fuzzy c-means clustering algorithm," Computers & Geosciences, vol. 10, № 2–3, pp. 191-203, 1984,

[12] R. Babuska, P.J. van der Veen, U. Kaymak, "Improved covariance estimation for Gustafson-Kessel clustering," Proceedings of the 2002 IEEE international conference on fuzzy systems, Honolulu: Technische Universitet Eindhoven, 2002, pp. 1081-1085.

[13] I. Gath, A.B. Geva, "Unsupervised optimal fuzzy clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, vol. 11, № 7, pp.773-781.

[14] S.D. Shtovba, O.D. Pankevich, A.V. Nagorna, "Analyzing the Criteria for Fuzzy Classifier Learning," Automatic Control and Computer Sciences. 2015, vol. 49. № 3, pp. 123–132, doi:10.3103/S0146411615030098.

Modelling the Generalized Multi-objective Vehicle Routing Problem Based on Costs

Viktor Kubil, Vasily Mokhov and Dmitry Grinchenkov
Platov South-Russian State Polytechnic University (NPI),
Prosveschenia Str. 132, Novocherkassk, Rostov Region, Russia
vksend@gmail.com, mokhov_v@mail.ru, grindv@yandex.ru

Keywords: Multi-Objective Optimization, Mathematical Model, Vehicle Routing Problem, Combinatorial Optimization, Graph Theory.

Abstract: The following article addresses a complex combinatorial optimization and integer-programming problem, referred to as the vehicle routing problem, which is typically related to the field of transportation logistics. The aim of the research is to combine a set of objective functions, number of common generalizations and extensions of the problem, arising in distributed services or goods supply. For this purpose, literature on the subject has been analysed, leading to the mathematical modelling method being applied. At the current moment such complicated variants of the problem present high importance for research because of both practical applications and high complexity. The paper proposes a new generalized multi-objective vehicle routing problem with multiple depots and heterogeneous vehicles fleet with regard to various factors affecting costs. The problem statement is presented as a mixed integer linear program. Objectives scalarization approach is proposed in order to reduce decision-maker participation. Shortcomings of the single-criterion formulation and negative effects of replacing the criteria with constraints are shown. The results provide initial data for solving a large number of transportation problems that are reduced to the vehicle routing problem. In particular, the application of the ant colony optimization as a method for solving the problem is discussed.

1 INTRODUCTION

Transportation affects many stages of production and distribution systems and represents an important component of the final product cost [1]. Route planning largely determines the effectiveness of transportation and is often reduced to one of the vehicle routing problems (VRP).

VRP is a combinatorial optimization and integer programming problem, which calls for the determination of the optimal set of routes to be performed by a fleet of vehicles to serve a given set of customers. In general, the objective is to minimize the overall transportation cost. It is one of the most attractive topics in operation research, logistics, and supply chain management [2, 3]. This interest in VRP is motivated by both its practical relevance and its considerable difficulty.

Indeed, a large number of real-world applications have widely shown that the use of software optimization and automated procedures for solving the VRP yields substantial savings in the global transportation costs [4]. However, the successful application of optimization techniques requires a mathematical model of the problem under consideration.

For today, there are many variants of VRP and its formulations, differing mainly by various additional restrictions. These variants have extended the applicability in real-life cases, but they are often based on models that do not take into account many factors.

Typically, a VRP model contains a single criterion to be minimized, which is the cost proportionate to the total trip distance or time. In fact, most of route planning problems are multi-criteria, and there is no unified solution, which simultaneously satisfies all the objectives.

The purpose of the research is to design a mathematical model of multi-objective VRP and a scalarization approach for reducing decision-maker participation. This paper considers more complex and generalized variant of the VRP, that can be categorized as “rich” VRP [5, 6], that is closer to the practical distribution problems. It makes the model universally applicable and allows for exclusion of extra specifics and characteristics, if needed.

2 PROBLEM OVERVIEW

First of all, let us consider the classical VRP to demonstrate the need for multiple criteria or constraints.

2.1 The classical VRP

The solution of the classical VRP problem is a set of routes, which all begin and end in one depot, and which satisfy the constraint that all the customers are served only once (Figure 1). The transportation cost can be diminished by reducing the total travelled distance and by reducing the number of the required vehicles.

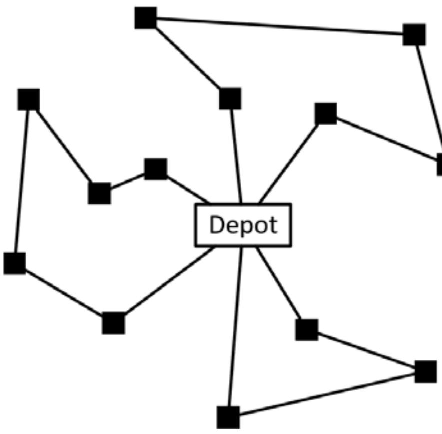


Figure 1: Example of a classical VRP solution: the black squares and the connecting lines represent clients and routes respectively.

2.2 Downside of a single criterion

In the case of the described classical VRP with one depot, the minimized sum of weights of all edges that make up the routes leads to the only optimal solution of having a single route for a single vehicle, i.e. VRP is reduced to less difficult TSP.

This conclusion follows from the structural features of the road network and its graph: the weight of any path between two vertices must be greater than or equal to the weight of the edge connecting them, therefore the triangle inequality is met for all edges: $c_{ik} + c_{kj} \geq c_{ij}, \forall i, j, k \in V$.

Since the increase in number of vehicles leads to an increase in number of the edges entering routes, it is possible to reduce the solution cost by connecting the last vertex of the previous route with the first vertex of the next (Figure 2).

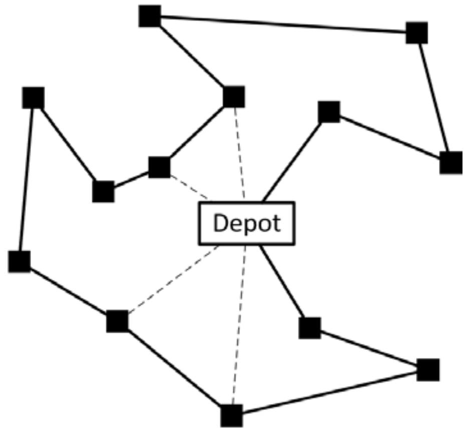


Figure 2: The total cost of the solution is reduced by connecting routes into one.

Obviously, in this case, the total value of all routes is reduced, but clients are served sequentially and in general, the implementation takes more time than using several vehicles. Thus, overall performance is reduced and another criteria or constraints are is required to resolve the dilemma.

Notice that the conclusion is valid for both symmetric and asymmetric matrix of costs if the triangle equality is satisfied.

2.3 Constraints instead of multiple objectives

Some objectives can be replaced by constraints in order to consider multiple criteria. It is suitable for some specific cases and usually greatly facilitates the search for solutions, but it poorly corresponds to reality in general.

For example, one of widespread approaches to obtain approximately equivalent routes is route balancing. There are different ways to balance routes by restrictions, like balancing the number of customers served by each active vehicle, balancing the distance of routes travelled by vehicles or balancing the waiting time required for the route.

In the first instance, the number of vertices in each route must not differ by more than specified (one, in extreme cases). This restriction allows to find solutions effectively even for a large number of vertices [7].

Thus, routes are balanced in the number of vertices, but not always distance-balanced (it can be seen in Figure 3 that routes are incommensurable in length).

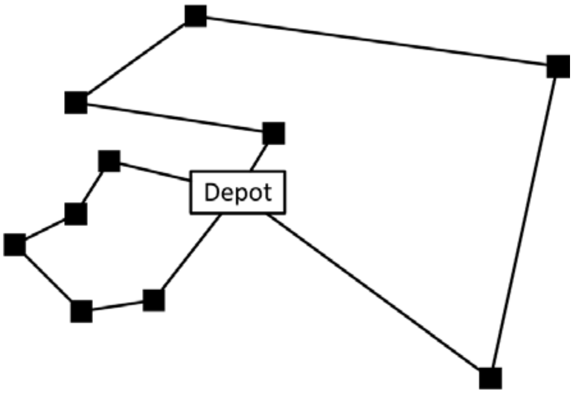


Figure 3: Balancing the number of vertices.

Contrariwise, the equalization of route distances may lead to irrational result, when a vertex is forced to belong to the route (Figure 4 shows that two vertices at the bottom are in different routes).

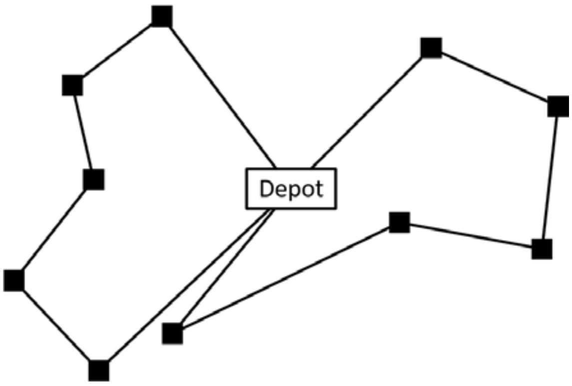


Figure 4: Balancing the distances.

It should be noted that balancing is sometimes proposed as an objective function [2]. However, it is rarely justified and associated with costs. Furthermore, it is clear, that a compromise among different balancing criteria is needed.

3 GENERALIZATIONS

In this paper, an asymmetric Heterogeneous Multi-Depot VRP (HMDVRP) mathematical model is used to formalize the described general multi-objective problem. It is a variant of the VRP characterized by multiple depots, multiple vehicle types and multiple asymmetric matrixes of initial data for each vehicle. Figure 1 provides an example for the solution with the use of four vehicles (a, b, c, d) and different depots.

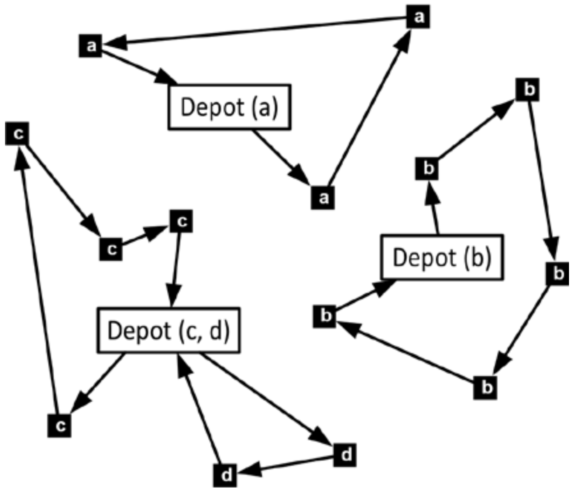


Figure 3: The solution for combined generalizations.

3.1 Symmetric and asymmetric

In the symmetric VRP, the distance between two customers is the same in each opposite direction, forming an undirected graph [8]. This assumption does not correspond to real conditions [9] and often leads to a certain gap between a theoretical project and practical application. Actually, the shortest path between two points of the road network usually depends on directions. Such differences are most noticeable at small scales, for example, in an urban environment. Therefore, from a practical point of view, it is advisable to consider the asymmetric VRP, which assumes different distances in each opposite direction, forming a directed graph. In addition, many software tools for routing provide data according to chosen direction.

On the other hand, the use of a directed graph significantly increases the solution space and, consequently, complicates the search for the optimum. To avoid this, depending on the particular application, costs (weights of edges) for opposite directions can be reduced to a certain average value.

Thus, symmetric problem can be considered as a particular case of asymmetric one, and since the purpose of this article is to pose a generalized version of multi-objective VRP, asymmetric variant will be used further.

3.2 Single and multiple depots

As noted above, route planning from depots to customers is a common and challenging task. Nevertheless, there is a rigid assumption that there may be only one depot. Although the single-depot VRPs have attracted so much attention, they are not suitable for some cases where a company has more than one depot, in which vehicles start and end their routes [10].

To resolve this limitation, this paper focuses on the VRP with multiple depots, or Multi-depot VRP (MDVRP). Multi-depot VRP is a generalization of the classical VRP, so it does not rule out the case of single depot.

Because there are additional depots, the decision makers usually have to determine which depots serve which customers, which is a grouping problem to be solved prior to the routing and scheduling problems. Obviously, this type of problem is more challenging and sophisticated than the single-depot VRPs.

3.3 Homogenous and heterogeneous fleet

Commonly, the fleet in VRP models is homogeneous, which does not always correspond to reality. Decisions relating routing heterogeneous fleets of vehicles are frequently taken into consideration in logistics operations [11].

The Heterogeneous Fleet VRP (HVRP) is a generalization of the classical VRP in which customers are served by several different types of vehicles with various characteristics. It is assumed in proposed model that the number of vehicles of each type is fixed and equal to a constant (for it to be unlimited the number of vehicles just should be big enough).

It is harder to solve heterogeneous fleet problem than the homogeneous one. Therefore, if the difference among vehicles is not significant, characteristics can be considered the same.

4 MATHEMATICAL MODEL

Taking into account everything above, we have constructed a mathematical model of the problem based on a linear programming formulation in terms of graph theory.

4.1 Problem formulation

Let $G = (V, A)$ be a directed graph, where $V = V_C \cup V_D$ is the set of vertices $\{1, \ldots, n\}$, V_C represents clients, V_D represents depots and $A = \{(i, j) : i, j \in V, i \neq j\}$ is a set of arcs defined between each pair of vertices. Heterogeneous fleet K of vehicles available and there is a bijection between the set K of vehicles and the set V_D of depot-vertices. Note that technically the location of depot-vertices can coincide if some vehicles belong to a single depot. The demand D_j is set for each customer j and the carrying capacity Q_k for each vehicle k. The end goal is to determine a minimum-cost set of routes in the feasible region considering all n criteria.

The formulation uses a set X of binary variables x_{ijk} equal to 1 if vehicle k travels directly from i to j, and to 0 otherwise.

According to the established assumptions, the Generalized Multi-Objective Vehicle Routing Problem can be stated as follows:

$$\min\left\{f_1(X),\ldots,f_n(X)\right\}, \tag{1}$$

subject to:

$$\sum_{k\in K}\sum_{j\in V} x_{ijk} = 1, \quad \forall j \in V_C; \tag{2}$$

$$\sum_{j\in V} x_{ijk} - \sum_{j\in V} x_{jik} = 0, \quad \forall k \in K \wedge \forall i \in V; \tag{3}$$

$$\sum_{i\in V_D}\sum_{j\in V} x_{ijk} \leq 1, \quad \forall k \in K; \tag{4}$$

$$\sum_{j\in V} x_{ijk} = 0, \quad \forall i \in V_D \wedge \forall k \in K, \quad i \neq k; \tag{5}$$

$$\sum_{j\in V_C} D_j \sum_{i\in V} x_{ijk} \leq Q_k, \quad \forall k \in K. \tag{6}$$

Equation (1) contains a vector of objective functions to minimize. Constraints (2) guarantee that each customer will be visited exactly once. Flow conservation constraints are expressed in (3). Constraints (4) mean that each vehicle departs from the depot once or doesn't depart at all. It is given that the fleet is heterogeneous, therefore it is important to consider that each vehicle belongs to its own depot (5). Finally, the limitation of the carrying capacity, which cannot be less than the total demand of the visited customers for each vehicle, presented in (6).

Such flexible formulation makes it possible to exclude insignificant limitations from consideration leaving only the needed constraints.

4.2 Problem parameters

Roads and clients are characterized by several parameters for each type of the vehicle, and defined as the initial data of the problem:

- c_{ijk} – fixed fare between vertices i and j for vehicle k;
- d_{ijk} – distance from vertex i to vertex j for vehicle k;
- t_{ijk} – travel time between vertices i and j for vehicle k, including the service time.

Such choice of parameters is based on real basic information about the path between two points, which can be obtained using modern navigation tools. It is supposed that data on roads are defined for various types of transport; therefore, an extra parameter index k corresponding to a particular type of vehicle is used.

4.3 Objectives

Let us form a vector of objective functions (1) allowing to take into account the basic criteria for estimating the cost of a solution. In construction of the mathematical model, the following principle was adopted: all decisions ultimately affect the enterprise profits and costs, which can be predicted at least roughly. Otherwise, additional methods of decision-making are required to find the weights of objective functions. Integrated expert estimates for decision-making support can simplify this problem [12].

1. Number of involved vehicles:

$$f_1(X) = \sum_{k \in K} \lambda_{1k} \sum_{i \in V_D} \sum_{j \in V} x_{ijk}.$$

Minimizing the number of involved vehicles is one of the key objectives of VRP [13]. If the route contains a single vertex, the vehicle does not leave the depot and is considered uninvolved. The penalty value λ_{1k} is applied to each involved vehicle k. Expenses on preparation of the vehicle and the courier determine the size of a penalty.

2. Total travelled distance:

$$f_2(X) = \sum_{k \in K} \lambda_{2k} \sum_{i \in V} \sum_{j \in V} d_{ijk} x_{ijk}.$$

The total distance of all routes determines mainly expenses on fuel and transport servicing. As a rule, the fuel consumption per distance unit and the type of used gasoline are known for all vehicles. Based on these data, fuel cost per distance unit can be obtained for each vehicle k. Maintenance costs are estimated per distance unit according to particular vehicle characteristics. In addition, different risk costs are calculated with consideration of distance. For instance, the risk associated with traffic accidents can be estimated as a multiplication of the probability of the accident and the average cost of its consequences. In a similar manner, other expenses associated with the distance travelled can be estimated. Thus, the coefficient of the objective function λ_{2k} is the sum of all components.

3. Total travelled time:

$$f_3(X) = \sum_{k \in K} \lambda_{3k} \sum_{i \in V} \sum_{j \in V} t_{ijk} x_{ijk}.$$

The cumulative time to be spent by all vehicles is necessary to take into account for the time rate wage payment calculation. So, λ_{3k} is estimated as wages per unit time for each courier.

4. The completion time:

$$f_4(X) = \max\left(\lambda_4 \sum_{i \in V} \sum_{j \in V} t_{ijk} x_{ijk} \right).$$

The time since departure of the first vehicle to return of the last vehicle determines costs of maintaining the transportation system as λ_4 and allows for calculation of the prior basic compensation of all couriers regardless of their employment. All involved vehicles move simultaneously therefore the completion time is determined by the most prolonged route and is evaluated using Chebyshev scalarization function. Coefficient λ_4 corresponds to the enterprise costs per time unit and does not depend on specific routes.

5. Fixed costs:

$$f_5(X) = \sum_{k \in K} \sum_{i \in V} \sum_{j \in V} c_{ijk} x_{ijk}.$$

The movement between two vertices may be associated with a priori costs, for instance, toll road taxes. In some cases, the cost depends on the type of vehicle, thence coefficient value is set for each.

Thus, for all considered objective functions, the value is now expressed in uniform units; therefore, a linear scalarization (weighted sum) can be used:

$$F = \sum_{i=1}^{n} f_i(X).$$

The proposed criteria are not exhaustive for all practical problems, but as analysis has shown, they have a significant impact on the cost of the solution. Any other criteria can be added in a similar way, but it is important to consider that the difficulty of finding a solution depends, among other things, on the complexity of an objective function. If it is impossible to relate the significance of the criterion to costs, other decision-making methods may be needed, especially for nonlinearity.

5 SOLUTION METHOD

In the field of combinatorial optimization, the VRP is regarded as one of the most challenging problems. It is indeed NP-hard, so that the task of finding the best set of vehicle tours by solving optimization models is computationally prohibitive for real-world applications [14]. As a result, different types of heuristic methodologies are usually applied. Furthermore, in view of conflicting objective functions it is difficult to accomplish the task of clustering. Clustering is likely to eliminate the optimal before the start of the search and keep suboptimal. Existing algorithms are not designed to solve the proposed generalized multicriteria vehicle routing problem.

Therefore, to solve this problem a modified multi-objective ant colony optimization algorithm (ACO) is being developed. ACO is a probabilistic technique for finding good paths through graphs and it is suitable for multi-objective problems [15, 16]. Swarm metaheuristics like ACO are useful for a large search space, especially using methods of increase in effectiveness [17]. The results obtained using the algorithm allow us to conclude that the solution of the problem depends substantially on the chosen weighting coefficients of the objective function.

Figures 5-6 show an example of how different solutions of the problem with two of the above objectives, the total travelled distance and the completion time. These criteria are highly conflicting and therewith illustrative. In the first case (Figure 5), the sum of distances is minimized, but the routes are not balanced and the time required to completion is excessive.

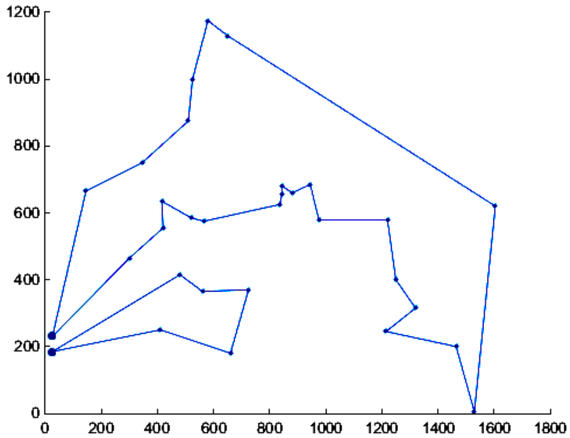


Figure 5: Minimization of the total travelled distance.

Alternatively, in the second case, the time of the longest route is minimized primarily, so the routes are approximately of equal length though they do not look optimal separately.

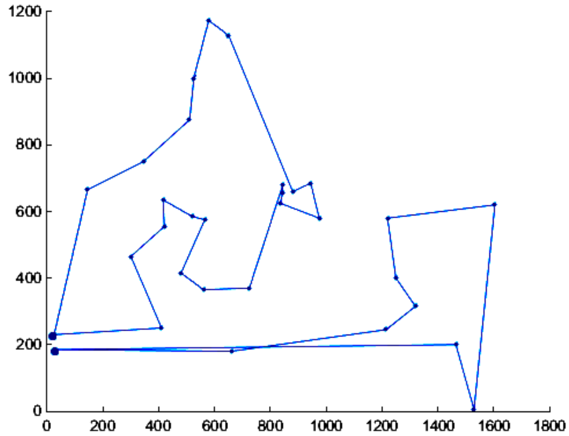


Figure 6: Minimization of the completion time.

6 CONCLUSIONS

A general mixed-integer linear mathematical programming formulation of Multi-Objective Vehicle Routing Problem with multiple depots and heterogeneous vehicles fleet is designed. The proposed set of objective functions considers factors affecting costs, but the considered list is not exhaustive. The presented approach allows supplementing the model with other objectives in the same way reducing decision-maker participation. In addition, it is possible to exclude some extra parameters, which are caused by generalization of the model, if required.

The model stated in this paper can be used effectively not only to solve problems concerning the delivery or collection of goods but for the solution of different real-world applications rising in transportation systems as well. Several examples of specific real-life situations [18, 19, 20] involving multi-objective routing problems are presented below:

- cargo transportation;
- fast food delivery;
- school bus routing;
- solid waste and trash collection;
- merchandise transport routing;
- tour planning for mobile healthcare facilities;
- postal services;
- maintenance engineering.

Thus, the new mathematical model, unlike existing ones, expands the practical applicability in cases of distributed depots and customers, and, at the same time, aims to global cost savings. In the forthcoming work, an algorithm for solving the problem will be reviewed in detail.

REFERENCES

[1] J.P. Rodrigue, C. Comtois, and B. Slack, “The geography of transport systems”, Routledge, July 2013.

[2] W. Zhou, T. Song, F. He and X. Liu, “Multiobjective Vehicle Routing Problem with Route Balance Based on Genetic Algorithm”, Discrete Dynamics in Nature and Society, December 2013.

[3] B.L. Golden, S. Raghavan and E.A. Wasil, “The vehicle routing problem: latest advances and new challenges”, Springer Science & Business Media, 2008.

[4] P. Toth and D. Vigo, “Vehicle routing: problems, methods, and applications”, SIAM, December 2014.

[5] A. Goel and V. Gruhn. “A general vehicle routing problem”, European Journal of Operational Research, pp.650–660, 2008.

[6] V.N. Kubil and V.A. Mokhov, “On the application of swarm intelligence in the transport logistics problems”, Problems of modernization of engineering education in Russia: collection of scientific articles on problems of higher school, Platov South-Russian State Polytechnic University (NPI), Novocherkassk, pp.140-144, 2014. (rus)

[7] Y.L. Kostyuk, M.S. Pozhidaev, “New heuristics is proposed for approximate solution of Vehicle Routing Problem (VRP) with capacity restriction and metric distances between vertices”. Tomsk State University Journal, 2010. (rus)

[8] D. Vigo, “A heuristic algorithm for the asymmetric capacitated vehicle routing problem”, European Journal of Operational Research, Vol. 89, No. 1, pp.108–126, 1996.

[9] R. Herrero, A. Rodríguez, J. Cáceres-Cruz and A.A. Juan, “Solving vehicle routing problems with asymmetric costs and heterogeneous fleets”, Int. J. Advanced Operations Management, Vol. 6, No. 1, pp.58–80, 2014.

[10] W. Ho, G.T. Ho, P. Ji and H.C. Lau, “A hybrid genetic algorithm for the multi-depot vehicle routing problem”, Engineering Applications of Artificial Intelligence, pp.548-557, June 2008.

[11] C.D. Tarantilis, C.T. Kiranoudis, and V.S. Vassiliadis, “A threshold accepting metaheuristic for the heterogeneous fixed fleet vehicle routing problem”, European Journal of Operational Research, 152(1), pp.148-158, 2004.

[12] D.V. Grinchenkov and 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.

[13] H. Li, A. Lim, “A metaheuristic for the pickup and delivery problem with time windows”, International Journal on Artificial Intelligence Tools, pp.173-186, 2003.

[14] R. Dondo, and J. Cerda, “A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows”, European Journal of Operational Research, pp.1478-1507, 2007.

[15] V.N. Kubil and V.A. Mokhov, “Application of the ant colony algorithms in multi-objective optimization problems with changing conditions”, Information and telecommunication systems and technologies, Kemerovo, pp.80-81, 2015.

[16] J.E. Bell and P.R. McMullen, “Ant colony optimization techniques for the vehicle routing problem”, Advanced Engineering Informatics, 2004.

[17] V.A. Mokhov, V.N. Kubil, F.A. Turovski and V.S. Filatov, “On the classification of methods to improve efficiency of swarms metaheuristics”, Scientific and technical conference and exhibition of innovative projects, Platov South-Russian State Polytechnic University (NPI), Novocherkassk, pp.74-78, 2014.

[18] N. Jozefowiez, F. Semet and E.G. Talbi, “Multi-objective vehicle routing problems”, European journal of operational research, 2008, pp.293-309.

[19] V.N. Kubil and V.A. Mokhov, “Automation of fast food restaurant networks using ant colony optimization”, The theory, design methods, software and technical platform of corporate information systems, Platov South-Russian State Polytechnic University (NPI), Novocherkassk, pp.44-46, 2014.

[20] V.N. Kubil, “Optimization of the courier services performance based on ant colony algorithm”, Scientific and technical conference and exhibition of innovative projects, Platov South-Russian State Polytechnic University (NPI), Novocherkassk, pp.119-120, 2014.

Cutting Time Optimization Using Technology for CNC Machines

Rustam Fayzrakhmanov, Rustam Murzakaev and Anatoly Polyakov
*Departmet of Information Technologies and Computer-Based System,
Perm National Research Polytechnic University, Perm, Russia
fayzrakhmanov@gmail.com, rustmur@gmail.com, squarepants_07@mail.ru*

Keywords: Cutting Time Optimization, Common Cut, Cutting Tool, CNC, Equidistant, Restrictions.

Abstract: The paper considers a problem of cutting tool route formation in a generalized formulation. The paper also proposes a mathematical model of total cutting time minimization based on standard, chained and common cutting technologies. Simple and combined equidistant types of transitions between equidistant (cutting and idle), as well as the entry points (insertions) and exits (ejections) of the cutting instrument, are used as technological limitations. Total cutting time equals to the sum of idle moves, total stroke time and the amount of time spent on material insertion and cutting initialization. The problem is solved in two main steps. First step (preliminary) includes determining areas of common cutting and equidistant contours generation, considering information about common cutting. Second step (optimization) includes creation of entry and exit points array, followed by optimization working stroke and idle moves ratio, as well as optimization of entry points count to minimize total cutting time. Algorithms for determining common cutting areas and generation common equidistant are shown. The proposed model has been tested via "ITAS NESTING" software complex. The Great Deluge Algorithm has been used for computing experiment conduction. The results of experiment obtained using waterjet instrument shows that common cutting technology usage leads to shortening of total cutting time due to reduction of idle moves and number of needed entry points.

1 INTRODUCTION

A necessary stage in the automation of the preparation of control programs for CNC cutting machines is the solution of the task of forming a cutting tool path (TP) in accordance with a nesting layout containing details and information on their location on a sheet [1][2].

Usually, in works devoted to the optimization of the TP route, the idling length [3], [4] or the cost of cutting, [5][6] are used as the criterion. Standard cutting (all contours are processed by a continuous line, transitions are carried out at idle speed) or chain cutting (there are transitions on the run) are provided as a cutting technology, but when cutting with a waterjet machine, it is advantageous to use the common cutting technology, especially for long product pieces. This makes it possible to reduce the total cutting time due to the decrease of the length of the working stroke and the number of necessary tool insertion points. To use this technology in the formation of the TP, it is necessary to take into account new limitations on the model and algorithms.

The aim of the work is to develop and test an experimental model for minimizing the total cutting time, taking into account cutting technologies.

2 TASK DESCRIPTION

The input information for the formation of the TP is the nesting layout and technological parameters of the cutting (cutting method, idling speed, working stroke speed, insertion time, cutting width, etc.).

Let's consider that the input nesting map contains m parts located at some distance from each other. The distance between the parts is greater than or equal to the predetermined gap size.

Some pairs of parts can be at the distance of cutting width d, with $d < gap$, which means that a common cut will be performed on this section. Parts could be nested inside each others to increase the cutting ratio.

Each part consists of one external and an arbitrary number of internal contours. We introduce a number of notation:

$outer_k$ – outer contour of the k-th part, $k = \overline{1,m}$;

$inner_k$– set of internal contours of the k-th part, $k = \overline{1,m}$;

In the situation where the part has no internal contours, $inner_k = \emptyset$.

A contour is to be considered as a closed geometric object (GO), made up of a set of segments and arcs, since most cutting machines work only with geometric primitives mentioned above [6].

In general, the formation of a TP consists of performing a number of steps that can be performed in different order:

– creating equidistant lines along which the contours of the parts will be cut (half the cutting width);

– generating the order of contour processing;

– setting the direction of processing for each contour (CW - clockwise or CCW - counterclockwise);

– determination of coordinates of entry (insertion) and exit (ejection) points for TP;

– selection and placement of specialized cutting technologies, such as bridges, jumpers, loops at contour corners, etc.

While forming a TP the technological features of cutting, including technological limitations (TL) and the cutting technology (CT) used, must be taken into account.

2.1 The model for minimizing the total cutting time

It is assumed that on the input nesting layout the parts can be paired, which complicates the model of minimizing the total cutting time, since the combined contours will have a common entry and exit point, i. e. not every contour will have its own entry and exit points.

To solve this problem, it is proposed to match the entry / exit points not for contours, but for their equidistant. Moreover, for contours that have a common cut, a common equidistant will be generated.

Now let's consider an ordered set of contours $C = (c_1, c_2, \dots, c_n)$ and equidistants $E = (e_1, e_2, \dots, e_n)$.

The function $eq : C \to E$ associates the contour with its equidistant. In the case where there are no pairs of aligned parts on the nesting layout, $n = m$. Otherwise, $m < n$, since the contours of the combined parts have one common equidistant.

Each i-th equidistant corresponds to a set of potential entry points P_i and exit points Q_i. In the process of forming the RI route, for each equidistant, a single insertion point $P_i^k \in P^i$ and the corresponding exit point $Q_i^k \in Q^i$ are chosen. The distance between the k-th output point of the i-th circuit and the s-th entry point of the j-th contour in the plane is denoted by $L(q_i^k, p_j^s)$.

The point on the i-th equidistant to which the transition from the insert point occurs, is denoted by $\overline{p_i}^k$, and the point from which the transition from the contour equidistant to the exit point occurs by $\overline{q_i}^k$. The distance $L(p_i^k, \overline{p_i}^k)$ is equal to the length of the approach on the i-th equidistant, the distance $L(\overline{q_i}^k, q_i^k)$ is equal to the retraction length.

The complete structure v, which uniquely identifies the TP, includes the following information:

P_{start} – start point of the cutting tool route;

$cord_j$ – order of processing the j-th contour in TP, $j = \overline{1,n}$ (contour number);

$ord(1), ord(2), \dots, ord(m)$ – permutation defining the sequence of equidistant contours traversal, where *ord(i)* is the equidistant number visited i-th in TP;

$e_i = eq(c_j)$ – j-th contour equidistant, $i = \overline{1,m}, j = \overline{1,n}$;

dir_i – i-th equidistant cutting direction (CW or CCW), $i = \overline{1,m}$;

p_i^k – i-th equidistant entry point, $i = \overline{1,m}$;

$\overline{p_i}^k$, – i-th equidistant cutting start point, $i = \overline{1,m}$;

q_i^k– i-th equidistant exit point, $i = \overline{1,m}$;

$\overline{q_i}^k$– i-th equidistant cutting end point, $i = \overline{1,m}$;

P_{end} – end point of the cutting tool route;

$trans_{ord(i),ord(i+1)}$ – a parameter denoting the type of transition between the exit point of the i-th equidistant and the entry point of the (i + 1) -th equidistant in the order of contour processing (has two values: 0 - working, 1 - idle) $i = \overline{1,m}$.

Objective function of the task of minimizing the total cutting time:

$$T_{cutting}(v) = \frac{L_{st}(v)}{V_{st}} + \frac{L_{im}(v)}{V_{im}} + T_{ins} \times N_{inp}(v) \to min \quad (1)$$

where $T_{cutting}(v)$ – total cutting time; $L_{st}(v)$– total length of tool stroke; V_{st}– stroke speed, i.e. tool cutting speed; $L_{im}(v)$– total length of idle movements; V_{im} – tool idle movement speed; $N_{inp}(v)$– number of insertion points; T_{ins}– constant, time of one insertion.

Function $L_{im}(v)$ depends on TP as follows:
$L_{im}(v) = L(p_{start}, p^s_{ord(1)}) + \sum_{i=1}^{m-1}[L(p^k_{ord(i)}, p^s_{ord(i+1)}) \times trans_{ord(i),ord(i+1)}] + L(q^k_{ord(m)}, p_{end})$ (2)

i. e. total length of idle movements between the contours of parts is reduced if there are working transitions in the route.

Total length of working strokes $L_{st}(v)$

$L_{st}(v) = \sum_{i=1}^{m-1}[L(e_i) + L(p^k_i, \overline{p_i}^k) + L(\overline{q_i}^k, q^k_i)] + \sum_{i=1}^{m-1}[L(q^k_{ord(i)}, q^s_{ord(i+1)}) \times (1 - trans_{ord(i),ord(i+1)})]$ (3)

where $L(e_i)$ the sum of the lengths of geometric primitives equidistant e_i (in the case of a regular equidistant, the sum is equal to the perimeter; in the case of a combined equidistant, the sum of the perimeters of the equidistant contours minus the length of the common face, since it must be taken into account only once).

Thus, if working transitions (chain cutting) are used instead of idle transitions between the contours, then the value $L_{st}(v)$ is increased on the nesting layout.

Consider $N_{inp}(v)$:

$N_{inp}(v) = m - \sum_{i=1}^{m-1}(1 - trans_{ord(i),ord(i+1)})$ (4)

i.e. the number of insertion points decreases with increasing the number of working transitions.

2.2 TP restrictions

In order to form an acceptable TP , it is necessary to take into account a number of technological limitations.

1. Equidistant. The cutting process assumes that to preserve the geometry of the work piece, the cutting must be carried out at some distance from the contour, called the *equidistant.* For external contours of parts, the equidistant is displaced outward, and for internal contours it is displaced inward [6].

2. Insertion points. The punching of the material is accompanied by various physical processes at the insertion point (deformations, heating, melting, etc.), therefore it is performed at some distance from the cutting contour. The offset of the insertion point is defined relative to the equidistant of the original contour and coincides with the direction of the equidistant displacement (outward or inward of the contour).

For example, when processing an external contour with a waterjet machine, it is recommended to place insertion points only on convex corners of the contour formed by geometric primitives. In the case of processing the inner contour - on the contrary, the best place is in the center of the arc or segment. For a laser machine, the position of the tapping point can be arbitrary [1].

3. Direction of entry and exit. Entry and exit on the *i*-th equidistant is performed in the direction coinciding with the cutting direction, specified by parameter dir_i.

4. Start and end of TP. TP begins at the point P_{start} (usually, this is the origin of the coordinate system) and ends in point P_{end} (may coincide with P_{start}).

5. The order of contour processing. The order of contour processing must satisfy two constraints, called preconditions:

− processing of the external contour of a part can be carried out only after processing all of its internal contours;

− processing of parts embedded in the inner contour of another part must be made before the contour in which they are located.

To take into account the preceding conditions, the following rule is proposed. If there is an *i*-th contour inside the *j*-th contour $i \neq j$, then the *i*-th contour must be processed before the *j*-th contour, i.e.
$belongs(i,j) \Rightarrow cord_i < cord_j$,
where $belongs(i,j)$ – a predicate equal to *true* if the *i*-th contour belongs to the inner region of the *j*-th contour and false otherwise; $cord_i, cord_j$– number of the *i*-th contour and the *j*-th contour in the contour processing sequence.

Thus, a model for minimizing the total cutting time is constructed, taking into account constraints in the form of equidistant points, entry and exit points, preconditions and technologies of standard, chain and common cutting.

2.3 Basic stages of problem solving

To solve the task, it is necessary to perform a number of actions related to the preliminary stage:

1) construct a matrix of precedence conditions for contours [4];

2) determine the areas of the common cut on the nesting layout;

3) generate equidistant contours taking into account information about the common cut.

The obtained information will be used at the stage of TP optimization, which is also proposed to be divided into a number of actions:

1) generate a set of potential entry and exit points;

2) perform optimization of idle movement and stroking time, as well as the number of insertion

points. It should be noted that it is possible to minimize only the idle time.

2.4 Implementation features

Let's consider key implementation features of the general cut and equidistant accounting.

1. Common cut section determination. To use a common cut, groups of tightly laid parts with a common face are created. Because the contours are separated by a common cutting line, there is no need to cut the segment twice. A common straight line allows to place pieces at a distance *d*, which saves the material and reduces the total length of the cut. It should be noted that implementation of a general cut between arcs is not considered in this paper.

The admissibility of a common cut for a pair of geometric objects belonging to two different contours is determined by the fulfillment of a number of conditions:

1) both geometric objects are straight;
2) geometric objects are parallel to each other;
3) the lengths of geometric objects are equal;
4) none of the straight lines goes beyond the border of the other (aligned with each other);
5) the distance between geometric objects is equal to the cutting width *d*.

Determination of the common cut sections is performed between all pairs of contours. The geometric objects of these contours are compared in pairs with each other, and if all the TR are fulfilled, it is assumed that a common cut is allowed between the GO data. The complexity of the algorithm for determining sections of a common cut is $O(n^2)$, where *n* – is the total number of geometric objects on the nesting layout.

2. Equidistant alignment. A special situation occurs when it is necessary to construct an equidistant for a pair of combined contours. In this case, the following actions are performed:

1) construct equidistant [7] for all internal and external contours (Figure 1, a);
2) check whether the contours are aligned (that is, the conditions for a general cut are met);
3) if the contours are combined, combine the equidistant contours into one common equidistant (Figure 1, b).

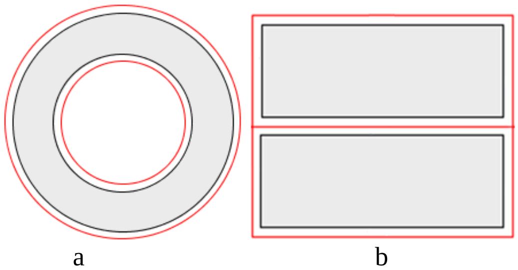


Figure 1: An example of simple equidistant for the inner and outer contours (a)
and a combined equidistant (b).

3 EXPERIMENTAL VERIFICATION OF THE MODEL

To verify the adequacy of the proposed model, we perform a numerical experiment using a nesting layout having several correct areas of the common cut. The experiment will also demonstrate the effect of a common cut on the cutting time.

The *ITAS Nesting* software complex is used as the environment for the experiment. As an algorithm for the formation of TP, the Great Deluge Algorithm is used, which gradually accomplishes the tasks of creating the sequence of parts machining and selecting insertion points on them. It should be noted that any other algorithm that allows to build a route taking into account technological limitations is suitable for using the model [8].

To determine the total cutting time, it is needed to know the cutting speed, idle speed and the time required for inserting the tool into the material. But since these parameters depend on the model of the machine, the type of material and its thickness, we use a special case of cutting a steel sheet of Russian St3 grade 50 mm thick.

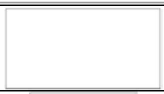
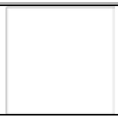
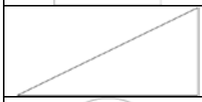
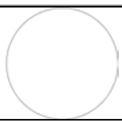
With this cutting, the following parameters are typical for the waterjet machine:

- the tool cutting speed is 25 mm / min.
- speed of tool idling - 1000 mm / min.
- material insertion time - 20 s.
- tool delay time before cutting starts - 3 s.

To test the effectiveness of the model, we will construct TP using general snake algorithm, original TP with and without common cutting (CC) for four different test cases (Figures 2-5).

First test case consists of 5 different parts packed in some irregular way. The content of this test case is presented in Table 1.

Table 1: Test case, 1 part.

Part	Width, mm	Height, mm	Count
	210,0	1050,0	2
	400,0	200,0	3
	200,0	200,0	2
	420,0	200,0	2
	100,0	100,0	2
	400,0	200,0	2

The aim of this test is to see how model will work on usual irregular nesting layout with basic parts. The layout is presented in Figure 2.

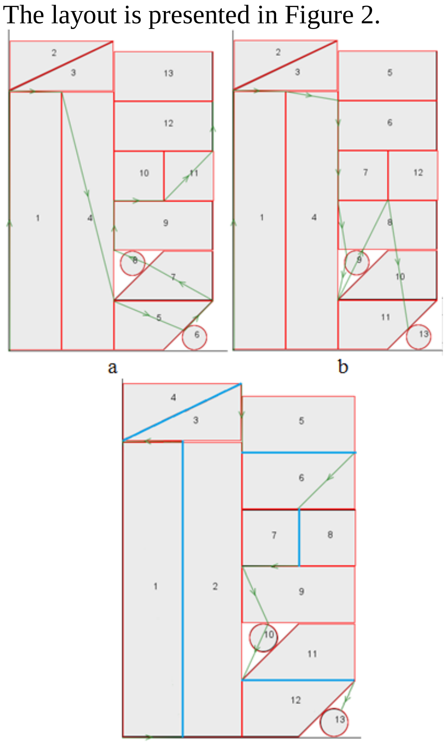
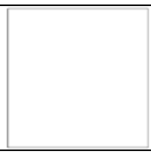
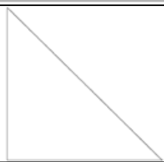


Figure 2: Test case 1 with default snake strategy (a), TP without use of common cut (b) and with it (c).

In Figure 2c, pairs of contours are processed by common cut: 1-2, 3-4, 5-6, 7-8, 11-12 (the common face is shown in blue). The contour numbers correspond to the order number of the processing.

Second test case consists of 3 different parts packed in pairs. The content of this test case is presented in Table 2.

Table 2: Test case, 2 parts.

Part	Width, mm	Height, mm	Count
	100,0	100,0	4
	200,0	100,0	2
	100,0	100,0	16

The aim of this test is to see total amount of profit provided by CC for nesting layout with high number of conjoined parts. The layout is presented in Figure 3.

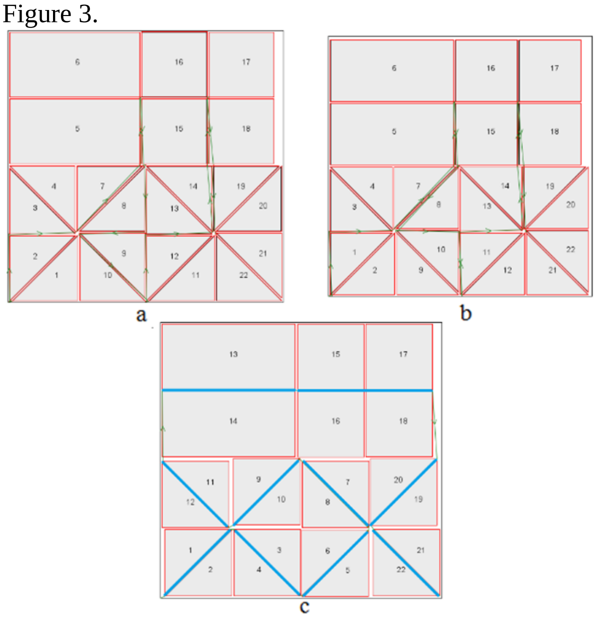
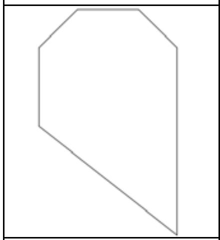
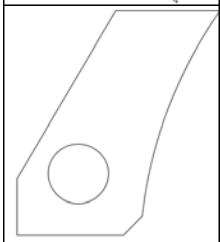


Figure 3: Test case 2 with default snake strategy (a), TP without use of common cut (b) and with it (c).

In Figure 3c, pairs of contours are processed by common cut: 1-2, 3-4, 5-6, 7-8, 11-12, 13-14, 15-16, 17-18, 19-20, 21-22. It may be noted that half of the parts were conjoined in this nesting layout.

Third test case consists of 2 different non-basic parts providing possible and impossible CC areas at the same time. The content of this test case is presented in Table 3.

Table 3: Test case, 3 parts.

Part	Width, mm	Height, mm	Count
	130,0	212,0	8
	53,5	60	19

The aim of this test is to see profit on complex parts with equal possible and impossible CC ratio for nesting layout square. The layout is presented in Figure 4.

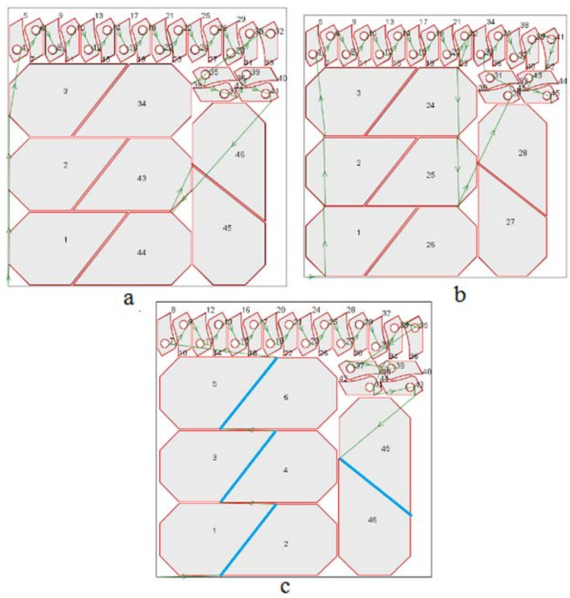
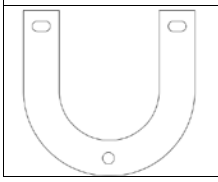


Figure 4: Test case 3 with default snake strategy (a), TP without use of common cut (b) and with it (c).

In Figure 4c, pairs of contours are processed by common cut: 1-2, 3-4, 5-6, 45-46. Most of the parts are not affected by CC.

Fourth test case consists of 1 part type packed in regular way without any CC possibility. The content of the test case is presented in Table 4.

Table 4: Test case, 4 parts.

Part	Width, mm	Height, mm	Count
	240,0	232,3	40

The aim of this test is to see if any profit can be achieved without CC on regularly nested parts since snake algorithm works best on such layouts. The layout is presented in Figure 4.

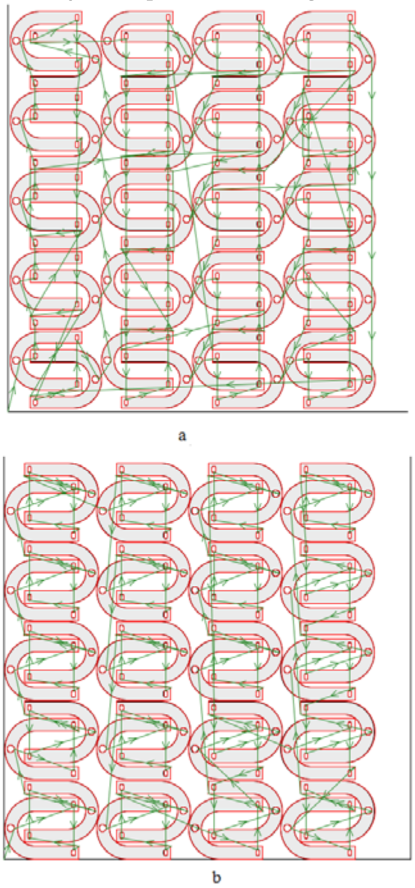


Figure 5: Test case 4 with default snake strategy (a), TP with and without use of common cut (b).

In Figure 4b, one can see that no parts were affected by CC.

The results of the experiment are summarized in Table 5. Default snake method is presented as absolute value while original TP is expressed as percentage difference to snake method.

Table 5: The results of a numerical experiment.

Parameter	Snake (value)	Without CC (percentage difference)	With CC (percentage difference)
Test case 1			
Idling (mm)	7133,0	22,2%	76,4%
Idling time (s)	428,0	22,2%	76,4%
Working stroke (mm)	15270,8	0,0%	16,5%
Working stroke time (s)	36649,9	0,0%	16,5%
Insertion points (pc.)	13	0,0%	38,5%
Insertion time (s)	299,0	0,0%	38,5%
Cutting time (s)	37376,9	0,3%	17,3%
Test case 2			
Idling (mm)	1546,0	0,1%	83,2%
Idling time (s)	92,7	0,1%	83,2%
Working stroke (mm)	8375,6	0,0%	19,2%
Working stroke time (s)	20101,5	0,0%	19,2%
Insertion points (pc.)	22	0,0%	50,0%
Insertion time (s)	506,0	0,0%	50,0%
Cutting time (s)	20700,2	0,0%	20,3%
Test case 3			
Idling (mm)	2050,5	5,1%	10,8%
Idling time (s)	123,0	5,1%	10,8%
Working stroke (mm)	9304,1	0,0%	7,4%
Working stroke time (s)	22329,9	0,0%	7,4%
Insertion points (pc.)	45	0,0%	8,9%
Insertion time (s)	1035,0	0,0%	8,9%
Cutting time (s)	23487,9	0,1%	7,4%
Test case 4			
Idling (mm)	32234,6	5,8%	5,8%
Idling time (s)	2114,1	5,8%	5,8%
Working stroke (mm)	64517,8	0,0%	0,0%
Working stroke time (s)	154842	0,0%	0,0%
Insertion points (pc.)	160	0,0%	0,0%
Insertion time (s)	3680,0	0,0%	0,0%
Cutting time (s)	23487,9	0,1%	0,1%

The numerical experiment showed a decrease in the total cutting time due to:

– the total idle movement time is slightly reduced for original TP without CC in range of 0,1-22%;

– the total idle movement time is greatly reduced (10,8-83,2%) for original TP with CC along with 7-19,2% reduction of working stroke;

– reduction of insertion time depends on count of conjoined parts and is up to 50% better for original TP with CC.

The decrease in the idling movements for the waterjet machine gives minor profit since its main time consumption is working stroke. It explains why original TP without CC gives only up to 0,3% total time profit. At the same time original TP with CC gives great time reduction up to 20% as in the test case 2 because of extremely low waterjet working speed.

It should be noted that best results can be achieved on irregular nesting layouts with high CC availability provided by conjoined parts with long straight lines. But even in the worst case, original TP is not worse than snake path giving some time profit on idle movement as 5,8% in the test case 4.

For further research, new tests with different idle/working speeds, including typical ones for laser machine, should be carried out. The difference between idle movement and idle stroke of laser machines is usually small since they work with thick materials [9]-[10]. Therefore, idle time reduction of around 80% as in the test cases 1 and 2 for CC, should give major total cutting time reduction.

4 CONCLUSIONS

As a result, the mathematical model of cutting time minimization has been developed. It differs from existing ones by:

– taking into account the technologies of standard, chain and common cutting;

– usage of equidistant contours, entry / exit points as technological limitations, conditions of contours and parts precedence.

The total cutting time is estimated taking into account the lengths of the equidistant, the lengths of the approaches / ejections and the types of transitions between equidistant contours (idle or working).

The numerical experiment carried out for the waterjet machine showed the adequacy of the constructed model. It is revealed that this model could give the results not worse than default snake strategy on regular nesting of complex parts, and up to 20% total cutting time decrease in high CC cases. Further research should be carried out, including laser machine modeling and introducing this technology in real industrial environment.

REFERENCES

[1] Fayzrakhmanov R.A., Murzakaev R.T., Mezentsev A.S., Shilov V.S. Formation of an energy-efficient cutting tool path in waterjet and laser cutting cnc machines. Russian Electrical Engineering Vol. 85 (2015), Ppp. 651-655.

[2] Petunin A.A., Chentsov A.G., Chentsov P.A., Chentsov A.A. Elements of dynamic programming in local improvement constructions for heuristic solutions of routing problems with constraints. Automation and Remote Control Vol. 78 (2017) pp. 666-681.

[3] Leung, S., Lin, Y., Zhang, D. Extended local search algortihm based on nonlinear pro-gramming for two-dimensional irregular strip packing problem. Computers & Operations Research Vol. 39 (2012), pp. 678–686.

[4] Xie S.Q., Gan J. Optimal process planning for compound laser cutting and punch using genetic algorimthms. Int. J. Mechatronics and Manufacturing Systems. – 2009 Vol. 2. Pp. 22-38.

[5] Roli A., Blum C. Metaheuristics in combinatorial optimization: overview and conceptual comparison. ACM Computing Surveys. Vol. 35 (2003) pp. 268-308.

[6] Bennell, J., Oliveira, J. The geometry of nesting problems: A tutorial. European Journal of Operational Research Vol. 184 (2008), pp. 397–415.

[7] Dewil R., Vansteenwegen P and Dirk Cattrysse. Sheet Metal Laser Cutting Tool Path Generation: Dealing with Overlooked Problem Aspects // Key Engineering Materials – 2015. Vol. 639. Ppp. 517-524.

[8] Dewil R., Vansteenwegen P and Dirk Cattrysse. Construction heuristics for generating tool paths for laser cutters. International Journal Of Production Research –Vol. 52(20) (2014) pp. 1-23.

[9] Dewil R., Vansteenwegen P and Dirk Cattrysse. A review of cutting path algorithms for laser cutters. International Journal of Advanced Manufacturing Technology – Vol. 87(5-8)(2016) pp.1865-1884.

[10] Umar S., Jawahar N., Balamurali M. Sequential optimization approach for nesting and cutting sequence in laser cutting – Journal of Manufacturing Systems Vol. 33(4) (2014).

Synchronized Control of Four or More Stepper Motors for Computer Numerical Controled Machines and 3D Printers

Nikola Jovanovski[1] and Josif Kjosev[2]

[1,2]Institute of Electronics, Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and MethodiusUniversity, Skopje, Republic of Macedonia
jnikola@feit.ukim.edu.mk, josif@feit.ukim.edu.mk

Keywords: CNC machine, CNC machining, CNC milling machine, 3D printer, 3D printing, Stepper motor control.

Abstract: A simple and effective algorithm for synchronized control of four or more stepper motors for computer numerical control machines is presented. Advantages and disadvantages of stepper and servo motors for CNC machines are discussed. Various control topologies for stepper motors are analyzed. Different types of stepper motor drivers and their characteristics are explained. The main emphasis is on the algorithm for synchronous control of multiple motors. The mathematical background for moving functions is explained in context of an overall 3D printing process. A practical implementation of the algorithm is evaluated. Possible directions for further research in this area are pointed.

1 INTRODUCTION

The Computer Numerical Control (CNC) is a technology that uses microcomputers to generate, parse and execute the sequential control that describes the end effector's behavior. The application of this technique is often used in turning, drilling, milling or extruding machines. Recently its application has expanded to other tasks, such as: electronic components insertion, tube welding, and cutting robots [1]. While these technologies are at the top of the third industrial revolution, today we are facing the fourth industrial revolution where one of the main driving forces will be 3D printing [12].

The idea behind this paper is to research the possibilities for designing an electro-mechanical system that can be used both as a 3D printer and a CNC machine. The focus in this paper is placed on the 3D printer issues, specifically, the synchronized control of 4 or more stepper motors. These results can be reused for CNC milling machines with some additional components [5].

The starting points are: the mechanical construction, mechanisms for movement and their integration. A 3D model of the machine is shown in figure 1. Speed and forces (torques) calculations define the motors and motor drivers selection. There are various types of printer heads but for this dual-purpose machine a thermal extruder type appears to be the most appropriate. The main controller should be also suitably selected for the dual purpose. These issues have been discussed in a previous paper [5].

CNC machine motion can be performed with AC or DC servo motors or stepper motors. Servo motors require additional positioning sensors and closed loop control systems. Stepper motors are very accurate in open loop systems [2][5] and are simple to control provided that the maximal required torque at the maximal required speed is available [4][6].

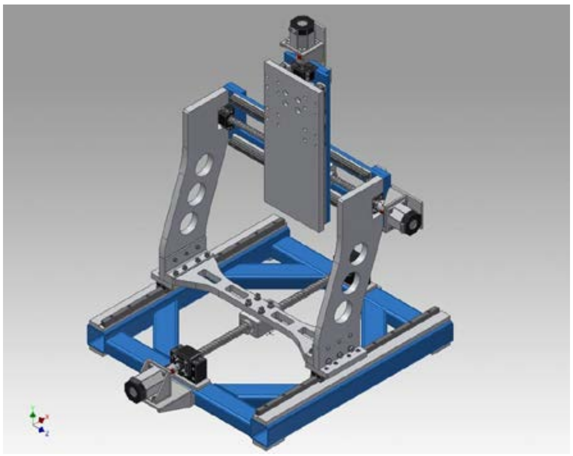


Figure 1: 3D printer model.

A specific issue regarding motion control is the multiple axes motion synchronization. This is a very complex task with servo motors but is also nontrivial with stepper motors. 3D printer manufacturers have proprietary control algorithms that are not accessible for performing experiments. Open source community projects are available for this purpose but, again, better performing algorithms are usually encapsulated in binary libraries. Some motor and controller manufacturers offer PLC controllers that provide 2-axes synchronization with speed control [13] but PLCs are an expensive overkill with the additional (unnecessary) options they have. They also do not provide solutions for 3 or more axis synchronization. In this paper we present a very efficient algorithm for synchronized stepper motors control that uses small number of operations compared with the algorithm for linear interpolation explained in the literature [3]. The results are based on real experiments with a 3D printer designed from scratch in our laboratory to experiment with different types of stepper motors and algorithms.

2 STEPPER MOTOR DRIVE TOPOLOGIES

The following is a brief review of the stepper motor operation and their drive topologies. Stepper motors usually have permanent magnet rotor and two (or four) groups of windings on the stator as presented in figure 2. The group of windings is presented as coils at 90 degrees which are called “electrical degrees” while the windings are positioned at certain spatial angle such that a 360° electrical angle corresponds to a much smaller spatial angle (eg. 7.2°).

The rotor is moved by changing the current through the coils in appropriate sequence such that magnetic forces attract or repel the rotor to support the rotation. If the currents through the coils change direction, the motor is named bipolar. Otherwise it is unipolar.

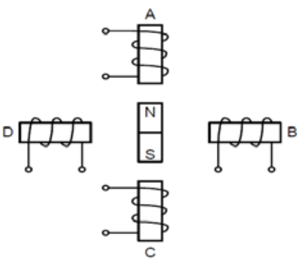


Figure 2: Stepper motor principle diagram.

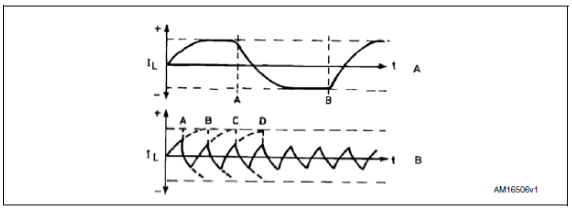


Figure 3: Peak winding current decreasing at higher speed, limited by motor inductance [9].

The current is obtained by connecting the coils to a voltage source V through appropriate electronic switches (a stepper motor driver). One step is a sequence of two coils (or four coils) sequentially energized/de-energized (pulsed) and it corresponds to a rotation by a predefined angle (typically 1.8° for a 200-step motor).

The coils resemble first order circuits with their inductance L and resistance R. This means the currents through the coils change exponentially:

$$i(t) = I(\infty) - \left[I(\infty) - I(0) \right] e^{-t/\tau}$$

where $I(0)$ is the starting current, $I(\infty)=V/R$ is the steady state current and $\tau=L/R$ is the time-constant of the circuit. A complete current change takes approximately 3τ. This limits the maximal applicable step rate i.e. rotational speed of the motor at full torque because the torque depends on the coil current amplitude – figure 3.

The time-constant can be reduced by increasing R, and correspondingly V, externally by a same factor n (typically 4-5) to retain the nominal current. Unfortunately this increases the power loses nRI^2 to an unacceptable level. The solution is the Pulse Width Modulation (PWM) technique which can provide (on average) the nominal coil current at much higher voltage with minimum power loses – figures 4 and 5.

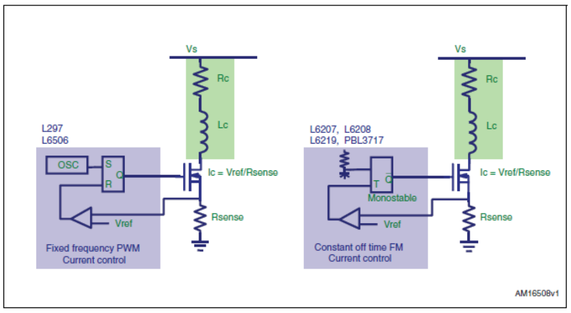


Figure 4: Switch mode drive implementation [9].

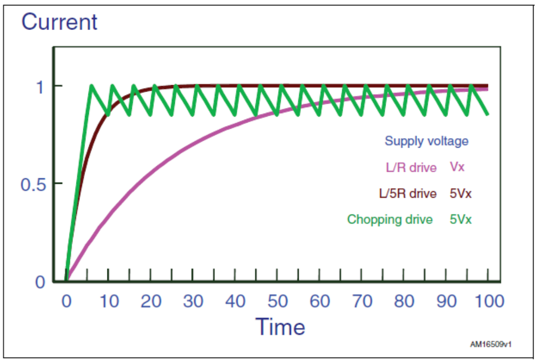


Figure 5: Current waveforms for L/R, L/nR and Switch mode drive [9].

3 STEPPER MOTOR DRIVERS

Simplest stepper motor drivers provide full-step and sometimes half-step driving. For a two-phase motor a full step consists of complete sequential de-energizing and energizing of the two coils and provides rotation of 90° electrical. Half-step is achieved by energizing both coils with the same polarity. This positions the rotor at 45° electrical and doubles the number of pulses for a revolution but increases the positioning resolution.

The previous idea can be extended further: if the currents in the energized coils are not equal, the rotor can be positioned at any angle between 0 an 90°, depending on the current ratio. This technique is called microstepping and provides substantial increase of the positioning resolution. Modern PWM stepper controllers are ideal for providing the microstepping control technique. They can typically increase the number of steps up to 16 times (eg. HY-DIV268N) or up to 256 times (eg. DQ860MA).

CNC machines convert the motor rotation into linear motion by ball screws. A standard value of 5mm displacement is obtained for each motor revolution. Thus a typical 200-pulse motor provides 0.025mm linear resolution. This is much above the required resolution. If microstepping with 16 microsteps is applied, the resolution becomes 1.56 micrometer which is satisfactory [11]. A 256 microsteps would provide resolution below 0.1 micrometer that is more than satisfactory [10]. Actually, uncompensated motor errors are much higher than this resolution [14] when absolute positioning is needed but position calculations can also accumulate errors that should be kept below the CAD resolution (which is typically at 1 micrometer). Additional benefits of high frequency microstepping are smooth motion, reduced motor noise, and reduced possibility for electro-mechanical resonance and oscillations [14].

Typical stepper motor driver has three control inputs:

- ENABLE/DISABLE (“zero” means the driver is disabled and the motor is turned off, while “one” puts the motor in the hold torque state).
- DIRECTION (“zero” drives the motor in clockwise direction, and “one” drives the motor in counter-clockwise direction)
- PULSE (each pulse at this input means rotation of 1 (micro-)step of the motor in the direction selected by DIRECTION input)

These signals are generated by the control unit’s firmware. Central part of this firmware is the optimized software algorithm for multiple motor driving, which is presented in the following sections.

4 3D PRINTER PRACTICAL IMPLEMENTATION

Besides the previously mentioned electromechanical components, the complete 3D printer includes:

-CAD software for designing and rasterizing the parts to be printed and generating commands in the standard CNC language named G-code,

-control unit (microcontroller),

-methods for transferring the G-code to the control unit,

-control unit firmware for coordinates calculations and motor driving signals generation,

-user interface.

The implemented CAD software is Pronterface (free open-source software) with it’s rendering engine Slic3r that slices the part into printable layers. The generated G-code specifies sequence of coordinates for linear X-Y axes movement and the required speed for each movement including the extruder for material deposition as coordinate E. The movement in Z direction is generated only once for each layer to be printed.

The control unit is based on STM32F4Discovery board with a CortexM4 microcontroller running on 168MHz.

G-commands are transferred through serial communication at the rate of 115 kbps. New command is transferred after the current is completed.

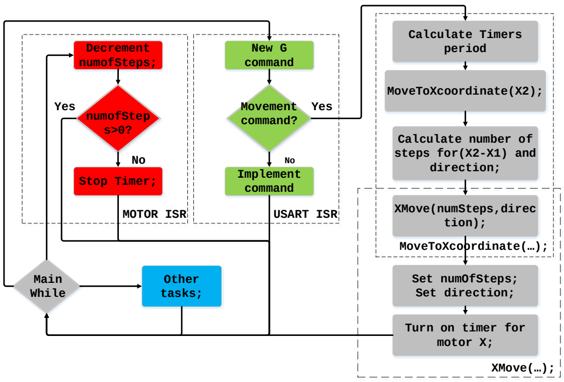


Figure 6: Firmware principle block diagram.

The firmware of the controller consists of many functions (communication, coordinate interpolation, motor control, extruder heating PID control, sensor scan, user interface through a graphical touch-screen, and some other housekeeping functions). Motor control functions are timer interrupt-driven and have the highest priority. Communication is also interrupt driven at a lower level and incorporates the linear interpolation calculations. The lowest interrupt driven function is the PID heater control (not shown here). All other functions are reaized in an infinite loop. The principal block-diagram of the firmware is presented in figure 6.

The current position is tracked through four global variables (*X*,*Y*,*Z*,*E*) which are updated on each motor step. The target position, sent by the CAD system, is kept in the variables *X*2, *Y*2, *E*2 and *speed*, (As previously mentioned, Z movement is accomplished separately between layers and will not be discussed here.)

The origin (0,0,0,0) is established at power-up of the machine. The printer moves into negative direction until proximity switches are activated to set the spatial origin (0,0,0) while the extruder coordinate at start-up is 0 by definition.

5 SOFTWARE IMPLEMENTATION OF LINEAR INTERPOLATION BETWEEN TWO POINTS

The need for linear interpolation can be depicted in figure 7. If movement is to be accomplished from point (X1,Y1) to point (X2,Y2) and the required numbers of steps for the X and Y axes movements are generated with the same frequency, a movement along the dotted line will be performed instead of a movement along the dashed line.

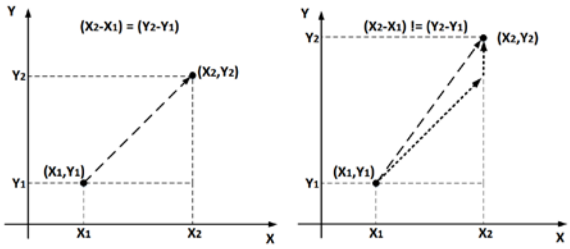


Figure 7: Movement path between two points without linear interpolation.

Obviously, the speeds along X and Y axes must be different to accomplish the required movement along the dashed line. The distance and the speed components along the X and Y axis are shown in Figure 8.

The problem here is to determine the values of the X and Y components of the velocity. Since the CAD system actually defines the speed of the material extrusion V (mm/min) it can be applied only to the direct distance between the points (1) and (2), Therefore first we needed to calculate the distance S from the starting point to the destination point. We calculate the coordinate changes ΔX and ΔY from the following equations:

$$\Delta X = | X_2 - X_1 |, \Delta Y = | Y_2 - Y_1 | .$$

and then the distance :

$$S = \sqrt{\Delta X^2 + \Delta Y^2}$$

Now the travel time is determined from the following equation:

$$t = \frac{S[mm]}{V[mm/\min]} .$$

Since the extruder travels the distances ΔX and ΔY during time t, the components of the velocity can be calculated by the following equations:

$$V_X = \frac{\Delta X[mm]}{t[\min]}, V_Y = \frac{\Delta Y[mm]}{t[\min]} .$$

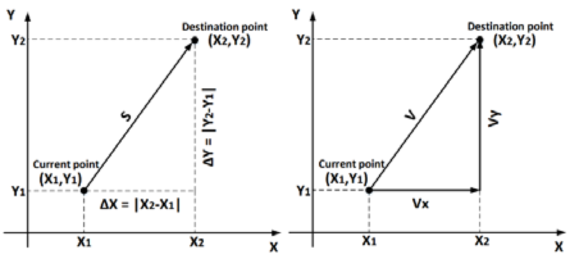


Figure 8: Interpolation of distance and velocity between two points.

The number of pulses for the required movement in the X and Y directions, ΔX and ΔY, can be calculated from the known koeficient of the spindle, 5mm/revolution, number of steps per revolution (200) and number of microsteps per step of the motor controler. Similar calculation holds for the extruder but here the distance is S.

The velocity in each direction is determined by the frequency of steps the motor performs i.e. the frequency of pulses applied to the PULSE imput of the controller. The easiest way to generate pulses with certain frequency (or period) is to use a timer/counter in PWM mode and set the period of the PWM waveform. The duty cycle is not important because the motor controller takes care of the PWM duty cycle applied to the motor. Therefore the pulse duration can be set to the maximum applicable value. If the current in the coil rampes above the maximal value, it would be limited by the power supply current limit.

By seting the timer prescalers, depending on the timer frequency and drivers microstepping factors, the following calculation can be established:

$$T_X = \frac{CONSTx}{V_X[mm/\min]}[us], T_Y = \frac{CONSTy}{V_Y[mm/\min]}[us].$$

Similar calculation holds for the extruder:

$$T_E = \frac{CONSTe}{\left|E_2 - E_1\right|} \cdot t[us]$$

where CONSTe is the appropriate scaling constant.

These values (T_X, T_Y, T_E) are directly written into the registers that define the timer periods associated with X, Y and E axes which is acomplished with the function:

```
void MoveXY(double x2, double y2,
double e2, double speed);
```

The same function calculates the number of steps for each coordinate and activates the appropriate timer. This is presented on the right-hand side of figure 6.

6 SOFTWARE IMPLEMENTATION TIMING ANALYSIS

As mentioned before, the current position of the printer is tracked through four global variables (X,Y,Z,E). These variables are updated on every motor step i.e. at every timer pulse in an interrupt routine. (Timers generate interrupts on the falling edge of every pulse). The interrupt routine (only for X-axis) is presented in red colour in Figure 6.

After completing the current movement, new point is obtained, calculations are performed and the process is repeated until a stop command is received.

The firmware is heavily interrupt driven and timing calculations have to be performed to check the processor utilization and confirm that every task can be completed.

The realized prototype has maximum linear speed of 600mm/min, microstepping factor of 128, 200 steps per motor revolution and 5mm linear motion per revolution. This gives the lowest interrupt period of T_m=19.5us. The non-optimized code produced interrupt routines of maximum T_{pm}=0.325us.

The communication between the controller and Pronterface transfers 34 bytes for each point. Since the maximal point distance is 0.1mm the 600mm/min requires 100 transfers per second. The four time_and_number_of_pulses calculation routines take approx. T_{pcalc}=0.35us each and are called at period of T_{calc}=10000us. Each byte is also collected by an interrupt that takes less than T_{pch}=0.15us at a period of approx. T_{ch}=100us.

The total utilization of the processor by the interrupt routines is:

$$Utilization = 4\left(\frac{T_{Pch}}{T_{ch}} + \frac{T_{Pcalc}}{T_{calc}} + \frac{T_{Pm}}{T_m}\right) = 0.073$$

This means that the processor has very low utilization of 7.3% and can easily control much more than 4 motors. This is accomplished by efficient implementation of the motor control algorithm using timer peripherals.

This also leaves space for linear velocity increase up to the maximally obtainable form Pronterface of 2000mm/min and it would not reduce substantially the number of motors because at higher speeds the microstepping factor is also reduced.

The higher speed may interfere with the current baud rate of 115kbps. At 10 bits per byte (1start+8data+1stop) 34 bytes take 3.4ms while the maximal point rate of 20000points/min means 3ms for a point. The remedy would be to increase the baud rate to the maximum specified by Pronterface of 250kbps [15].

7 CONCLUSIONS

This paper presents an efficient and simple algorithm for synchronized control of four or more stepper motors. The algorithm and the 3D printer prototype provide a solid base for further research in this area.

The prototype achieves high resolution at the expense of slightly reduced printing speed. This is due to the massive mechanical construction that provides higher accuracy but also requires higher power drive unit compared to typical commercial printers. On the other hand, the massive construction is needed for the alternative CNC subtractive processing.

Other possible direction for experimenting is variable printing speed. This option is not available in today's commercial 3D printers. Additionally, it should be emphasized that this kind of stepper motor implementation provides a reliable and accurate way of 3D printing without using a feedback link from the motors. This topics will be presented in a following paper.

REFERENCES

[1] Paulo Augusto Sherring da Rocha Junior,Rogerio Diogne de Silva e Souza, Maria Emilia de Lima Tostes, “Prototype CNC Machine Design”, 10.1109/INDUSCON.2010.5740068.

[2] J.M. Rosario, L.K. Rincon, D.Kubiak, D. Dumur, L.F.Melo, “Actuator Selection of CNC Machine Tool Based in Dynamical Modeling and Control”, International Symposium on Power Electronics, Electrical Drives, Automation and Motion, 2014.

[3] Dev P. Desai, Dr. D.M. Patel, “Design of Control Unit for CNC Machine Tool using Arduino based Embedded System”, International Conference on Smart Techologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), India, 6-8 May 2015, pp. 443-448.

[4] Industrial Circuits Application Note Stepper Motor Basics, http://www.solarbotics.net/library/pdflib/pdf/motorbas.pdf.

[5] Nikola Jovanovski, Josif Kjosev, “Hardware and Software aspects of 3D printers”, International Conference ETAI, 2016.

[6] Suk-Hwan Suh, Seong-Kyoon Kang, Dae-Hyuk Chung, Ian Stroud, “Theory and design of CNC Systems, 2008 Springer-Verlag London Limited.

[7] Jacob Bayless, Mo Chen, Bing Dai, “Wire Embedding 3D Printer”, Engineering Physics, University of British Columbia, April 12, 2010.

[8] Jonathan Heathcote, “Improving the Makerbot 3D Printer”, University of Manchester, School of Computer Science, Project Report 2012.

[9] Thomas Hopkins, Aplication note, Stepper motor driving, http://www.st.com/content/ccc/resource/technical/document/application_note/57/c8/7c/c1/0d/91/46/89/CD00003774.pdf/files/CD00003774.pdf/jcr:content/translations/en.CD00003774.pdf.

[10] Wantai Motor, DQ860MA, PWM currrent control, https://factorydaily.com/fdattachs/fdattachs7/112716014424312.pdf.

[11] HY-DIV268N-5A two phase hybrid stepper motor drive manual, http://wiki.kreitek.org/media/proyectos:cnc:div268n-5a-datasheet.pdf.

[12] Executive Perspectives “Accelerating the Next Industrial Revolution”, https://spare-parts-3d.com/2017/04/18/3d-printing-key-driver-4th-industrial-revolution-hp-ceo/

[13] Eaton Logic Controller Programming Manual, http://www.eaton.com/ecm/groups/public/@pub/@electrical/documents/content/mn05003003e.pdf

[14] Ericsson, Industrial Circuits Application Note, Microstepping, http://users.ece.utexas.edu/~valvano/Datasheets/StepperMicrostep.pdf

[15] Plastic Scribbler, Printing the Future in 3D (Pronterface tutorial), http://www.plasticscribbler.com/tutorial/getting-started/item/21-getting-started-with-pronterface#.WlGh4oWcEwE

Prediction-Based Planning in Production System Management through Subsystem Interaction

Mikhail Sadiakhmatov and Leonid Mylnikov
Perm National Research Polytechnic University
29 Komsomolsky Avenue, 614990 Perm, Russia
fortis3000@gmail.ru, leonid.mylnikov@pstu.ru

Keywords: Product Management, Production System, Industrial Engineering Problem, Optimization Problem, Criterion Function, Forecast, Prediction, Predictive Modelling, Reflexive Control.

Abstract: The research concerns the investigation of predictive models based on optimal control task. It allows increasing the management efficiency due to joint consideration and synchronization of internal and external processes towards the system. In this paper, the predictive model for solving multicriterion product management task was developed. To develop a model, automotive industry data was processed. The paper follows the reflexive approach and provides an application of simulation modelling to solve jointly the optimization problem taking into account the mutual influence of the production subsystems. The feasible solutions were received as functions of time. The solutions obtained were compared with the practical ones that based on historical data. The practical significance of the research lies in using market data to estimate company capabilities preliminarily whether they meet the market needs. At the same time, the objectivity of strategic decisions is increasing due to the formalization of process description, objective data preparation, and the company synchronization with the external environment.

1 INTRODUCTION

The market situation changes dynamically and the competition increases. The potential effect of the better management quality in production systems is concerned with company's susceptibility to the market changes and customer preferences. Companies have to implement new developments following fashion and customer preferences, create new markets, take into account the reduction of the product lifetime, the increase of modification number, product structural complexity, energy and resource intensity of production processes, a number of production systems involved in the production cycle. Furthermore, transferring from flow production to small-batch and even job (one-off) production for customer needs is an upward trend. The project developers start regarding production as a service. Thereby, there is an opportunity to order the service from different manufacturers in different countries, change the lot sizes, and make modifications. The production systems aspire to improve their universality and production processes flexibility, follow the path of progressive transformation of computer-aided manufacturing into automatic one and the virtual fabric. Therefore, the process of manufacturing tasks solving requires the better quality and the higher efficiency of the management decisions, especially for the small companies.

The factors that are not considered together previously, begin to influence the efficiency of the production system operation significantly. Thus, for example, it becomes impossible to consider management tasks only as industrial engineering and selling. The joint consideration of production and power-supply systems, warehouse operation, logistical organization, recycling and resource reusing tasks is required.

The subsystem interaction is considered with the time factor and traditionally based on using differential calculus. However, when developing the predictive model to manage production and technological processes it becomes complicated to formulate it because of the processes complexity. And when it is done the model obtained is usually insoluble. Therefore, the approaches based on the optimal control [1] and game theory [2] principles get widespread. When several subsystems are considered jointly, these approaches are confronted with the multicriteria problem. It causes an appearance of such methods as folding technique, criteria ranking, and reflexive control. The approaches increase task dimension, and solution finding faces NP-completeness and necessity of using metaheuristic methods.

The introduction of the time factor makes the task more complicated and necessitates simulation modelling of Δt or special condition principles. In this case, the solution will be found in the form of tabulated function. This formulation permits turning to the proactive management due to using the predictive models and solving the problem of lagging between decision and external situations regarding the system under consideration (internal processes synchronization and market condition). On the other hand, it requires better forecast quality.

2 THE PROBLEM STATEMENT

The predictive model development is based on forecasts [5].

The initial data of industry-oriented predictive model is generally represented in time series. The example is in Table 1.

Table 1: The initial data example described price changing.

Date	**Price of Ford Mustang**
01.11.2013	36654
01.12.2013	36652
…	…
01.01.2017	36284

Up-to-date models in the industrial engineering field are directed to the external and internal processes synchronization in order to reduce

- financial, time, and energy costs
- warehouse capacity
- path length travelled by components within a company
- negative impact on the environment, etc.

The models, therefore, are directed not only at process optimization but also at risk minimization [6], [7].

These tasks need to be considered according to the process proceeded dynamics. The description with using differentials prove to be too complex though, so simulation modelling is necessary.

In this case, each of the tasks is possible to describe with a separate criterion using reflexive approach. Moreover, we could find their solutions as a set of optimization problems, that have common parameters and use forecast-based data. Figure 1 shows the scheme of the models interaction. The received solution will be a tabulated time function with a fixed time step (the Δt principle).

Let us examine the tasks for the model described in Figure 1.

The economic lot-scheduling problem is mathematically described in the following way:

$$\sum_{ww^*} K_{ww^*}(C_w(t) * x_w(t) + C_{w^*}(t) * x_{w^*}(t)) \to \max,$$

$$\sum_{zw} R_{zwj} * S_{wk} * x_w(t) \le P_j, \forall j,$$

$$\sum_{w} S_{wz} * x_w(t) \le L_z, \forall z,$$

$$x_w(t) \le G_w(t), \forall w,$$

$$x_w(t) \ge 0, \forall w,$$

where K_{ih} — product w and w^*compliance coefficient;

w — product index;

x_w — production volume of product w;

C_w — net profit from product w manufacturing;

R_{zwj} — requirement in facility capacity for treatment material/ item/ component z of product w by facility j;

P_j — total capacity of facility j;

S_{wz} — requirement of material/ item/ component z per product unit w;

L_z — available material/ item/ component z;

z — index of material/ item/ component;

G_w — market/ demand/ order restriction for product w.

The purchase management task could be formulated as:

$$\sum_{zw} A_{zw}(t) * u_{zw}(t) + V_z(t) * L_z(t) + N_z(t) * y_z(t) \to \min,$$

$$L_z(t-1) + y_z(t) - B_z(t) = L_{zw}(t), \forall z,$$

$$\sum_{zw} R_{zwj} * S_{wk} * x_w(t) \le P_j, \forall j,$$

$$u_{zw}(t) \in [0,1], \forall z, w,$$

$$y_{zw}(t), L_{zw}(t) \ge 0, \forall z, w,$$

where u — Boolean flag representing if re-equipment/ revision/ reboot needed;

y — purchase volume;

A_{zw} — re-equipment/ revision/ reboot cost;

B_z — requirement/consumption of material/ item/ component z for manufacturing product w;

V_z — storage cost of material/ item/ component z;

N_z — cost of material/ item/ component z.

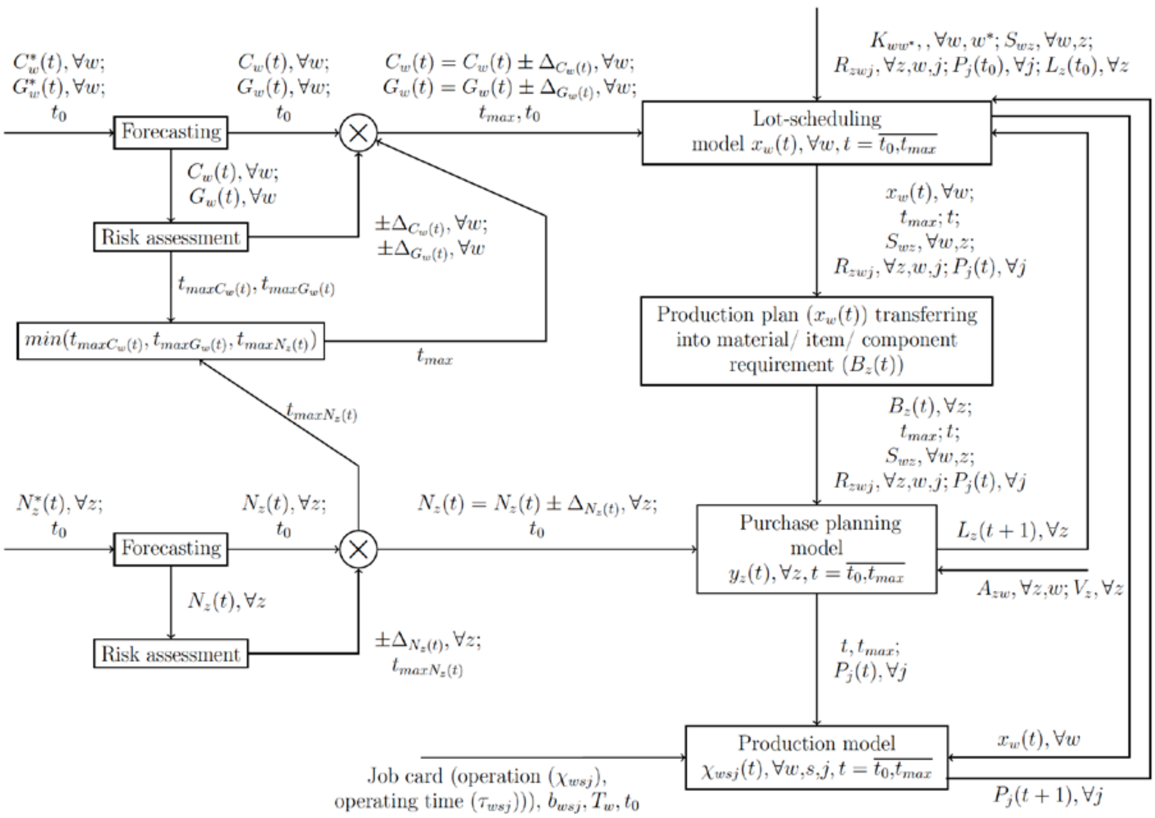


Figure 1: Structural scheme of predictive model interaction when their joint solution finding.

Below is a sequence of work planning task when assembling a product from a multitude of parts:

$$\sum_{wsj} k_{wsj} * \chi_{wsj}(t) * \tau_{wsj} \to \min,$$

$$\sum_{ws} \tau_{wsj} \le P_j(t), \forall j,$$

$$\sum_{js} \tau_{wsj} \le T_w * x_w(t), \forall j,$$

$$\sum_{j} k_{wsj} * \chi_{wsj}(t) * b_{wsj} = \sum_{j} \chi_{w(s+1)}(t) * \tau_{w(s+1)},$$

$$\forall w, s = \overline{1, (s^* - 1)},$$

where s – assembly step $s = 1, s^*$, s^* – last operation;

k_{wsj} – variable production costs

χ_{wsj} – number of operation at the point in time t

τ_{wsj} – time costs at step s on the facility j when manufacturing product w

b_{wsj} – coefficient of spoilage ($0 \ge b_{wsj} \ge 1$)

T_w – total time of manufacturing product w

Some parameters are computable. We could find their values using a job card. The general form of the job card is shown in Table 2.

For example, within the purchase management problem, B_z is the very parameter, which value depends on the production value of each product and total requirement of material/ item/ component z.

When considering the range of optimization problems as time problems, we are able to take into account a number of parameters as time functions. The set of item and storage costs time functions are represented in Table 3.

Table 2: Job card structure.

ID	Number/code of product (w)	Operation (χ)	Operation time(τ)	Previous operation (χ)	Facility (j)	Number of facilities (R)	Material/ component (z)	Consumption rate of material/ component (S)
1	1	$\chi_{1,s,j}$	τ_{1s1}	$\chi_{1,s-1,j}$	1,2	1,2	1	1
2	1	$\chi_{1,s,j}$	τ_{1s2}	$\chi_{1,s-1,j}$	2	1	2	1
3	1	$\chi_{1,s,j}$	τ_{1s3}	$\chi_{1,s-1,j}$	3	1	2,3	2,4
4	2	$\chi_{2,s,j}$	τ_{2s4}	$\chi_{2,s-1,j}$	3,4	1,1	2	1
5	2	$\chi_{2,s,j}$	τ_{2s5}	$\chi_{2,s-1,j}$	5	1	1,3	2,1
6	2	$\chi_{2,s,j}$	τ_{2s6}	$\chi_{2,s-1,j}$	6	1	1	1
…	…	…	…	…	…	…	…	…
	w	$\chi_{w,s,j}$	τ_{wsj}	$\chi_{w,s-1,j}$	j	R_{zwj}	z	S_{wz}
…	…	…	…	…	…	…	…	…

Table 3: Item and storage costs changing.

Date	Item	Item cost (RUB)	Storage cost (RUB)
Jan	1	100	10
	2	150	10
	3	220	12
Feb	1	100	10
	2	150	10
	3	220	12
Mar	1	110	10
	2	160	10
	3	230	12
Apr	1	110	10
	2	160	10
	3	240	12
…	…	…	…

A large number of parameters is defined with the time series and used them for forecast describing. Consequently, the result depends on forecast accuracy increases. In this case, the system behaviour investigation by modelling of predicted values deviation becomes actual. The use of forecasts leads to probabilistic models appearing based on risk assessment [8], Bayes theorem [9], and Monte-Carlo method [10].

3 PROBABILISTIC CHARACTER OF PARAMETER FORECASTING DEFINED WITH TIME SERIES

The use of forecasted values brings up the question of error estimate what computing risk assessment can be used for [11]. The risk assessment is calculated subject to the factors influence on the risk value: $r = \left|1 - \frac{a}{a^*}\right|$, where a – forecasted value of the estimated factor; a^*– exact value of the estimated factor.

In order to determine the planning horizon, we used the test sample. The continuous independent variables should be selected for considering management task, that uses forecasts of several factors. Consequently, the values of the variables are also independent events.

For independent parameters c_1 and c_2 the following dependence is correct:

$P(c_1c_2) = P(c_1)P(c_2)$, where $P(c)$ – probability of occurrence c.

Based on $P = 1 - r$, where P – probability, we could determine risk assessment values:

$$r(c_1c_2) = 1 - P(c_1c_2) = 1 - P(c_1)P(c_2) =$$

$= 1 - \left(1 - r(c_1)\right)\left(1 - r(c_2)\right) =$
$= 1 - 1 + r(c_1) + r(c_2) - r(c_1)r(c_2) =$
$= r(c_1) + r(c_2) - r(c_1)r(c_2).$

The forecasted values are time series. Therefore, the values could be considered jointly according to their simultaneous calculation.

Hence, risk assessment could be carried out with the cumulative sum. Parameter c_1 in the probability of occurrence formula takes values c_{1_1} and c_{1_2}:
$P\left(c_{1_1} + c_{1_2}\right) = P\left(c_{1_1}\right) + P\left(c_{1_2}\right) - P\left(c_{1_1}\right)P\left(c_{1_2}\right).$
$r\left(c_{1_1} + c_{1_2}\right) = 1 - P\left(c_{1_1} + c_{1_2}\right) = 1 - 1 +.$
$+r\left(c_{1_1}\right) - 1 + r\left(c_{1_2}\right) + 1 - r\left(c_{1_1}\right) - r\left(c_{1_2}\right) +$
$+r\left(c_{1_1}\right)r\left(c_{1_2}\right) = r\left(c_{1_1}\right)r\left(c_{1_2}\right).$

In order to calculate the following risk assessment values we used formula
$r(c_{1_{i-2}} + c_{1_{i-1}} + c_{1_i}) = r\left(c_{1_{i-2}} + c_{1_{i-1}}\right) +$
$+r\left(c_{1_{i-1}}\right)r\left(c_{1_i}\right)$

4 JOINT SOLUTION OF MANAGEMENT AND PURCHASE PROBLEMS

Each of optimization tasks received can be classified as a multiparameter task with non-linear restriction, some parameters of which are defined as time functions. The solutions of the tasks will be also time functions.

The gradient methods were the first to appear. They need the function to be twice differentiable and convex. The disadvantage of the methods is sensitivity towards the initial value, and also freezing in local extrema in the case of multiextremality, nonconvex restrictions, multiply connected feasible region etc.

Modern methods divided conditionally into three groups [12]: clustering, constraint propagation, and metaheuristic methods. When choosing the solving method, it should be taken into account that completeness is the most important feature of combinatorial optimization methods. The comprehensive method guarantees solution finding in the case of its existence. However, the large dimension of search space complicates the application of the method. In addition, solution search time might be unacceptable, e.g. because of decision time restriction. In case heuristic methods are used or combinatorial methods are supplemented with heuristic elements, the proof of the method completeness becomes more complicated. Heuristic search methods are for the most part incomplete.

In practice, hybrid methods are widespread. Moreover, any algorithm results would be improved due to joint solver constructing. In view of specialized solving method absence, it is reasonable to apply the evolutionary approach namely stochastic search method. The disadvantage of evolutionary methods is result and optimization time dependence on initial approximation.

For calculating tasks mentioned above, we used the genetic algorithm and its implementation in the programming language R – rgenoud package. The package combines evolutionary search algorithm with the methods based on derivatives (Newton or quasi-Newton) [13].

The example of product management problem solving, with the economic lot-scheduling subsystem considered, is described in Listing 1.

Listing 1. Function calculating production value

```
opt_GA_volume_time_plan <- function(C,
G, P, R, L, q1, q2, t){
    var <- length(C[1,])
     x <- matrix(NA, nrow = length(C[,1]
), ncol = var, byrow = TRUE)
    y<- NA
    for (i in 1:t) {
        x[i,] <- genoud(function(y) K[1
,2]*(C[i,1]*y[1] + C[i,2]*y[2]), nvars
= var, max = TRUE, starting.values = NU
LL, Domains = matrix(c(0, G[i,1], 0,G[i
,2]), ncol = 2, byrow = TRUE),data.type
.int = TRUE)$par
    }
    return(x)
}
```

The joint solution problems under discussion are represented algorithmically in Figure 2.

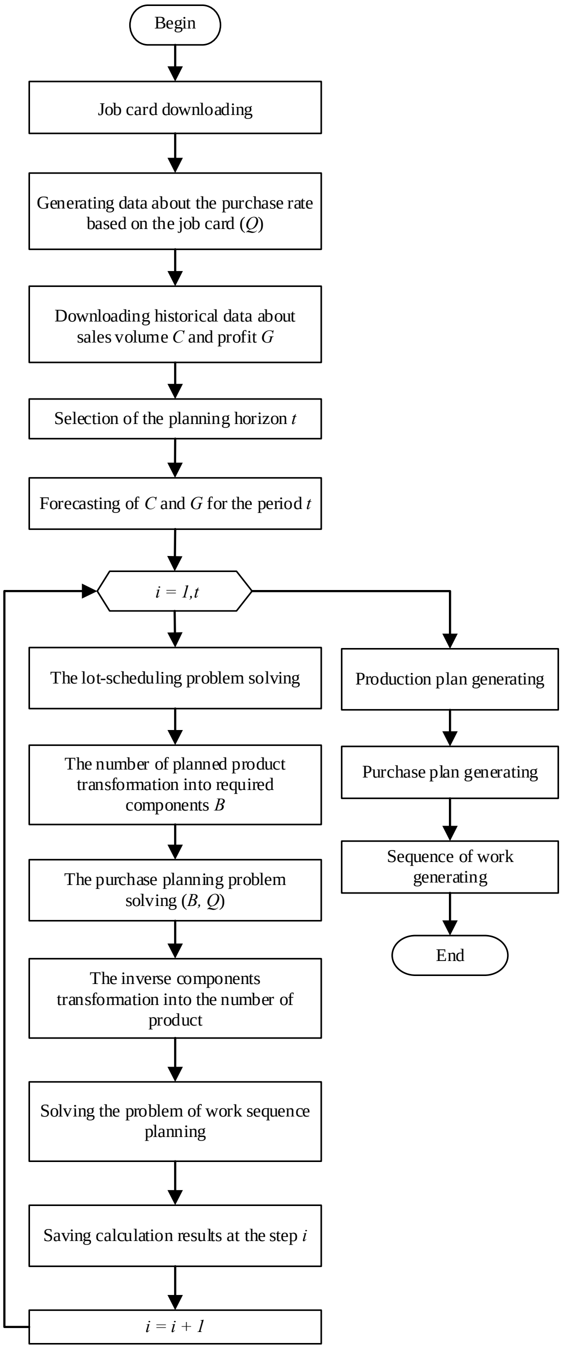


Figure 2: Flow chart of predictive model for joint optimization.

6 THE FINDINGS ANALYSIS

As a result, we received the set of findings:

- The planning horizon estimation for the methods used (Figure 3);
- The optimum production plan (Figure 4);
- The estimation of criterion function variation (Figure 5);
- The estimation of parameters sensitive to criterion function variation (Figure 6).

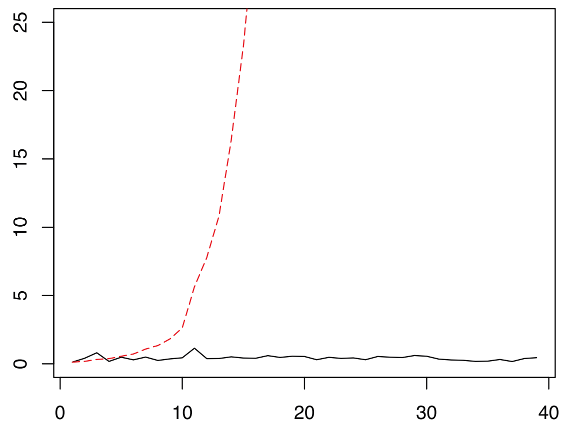


Figure 3: The magnitude of risk assessment obtained using the fractal forecast method.

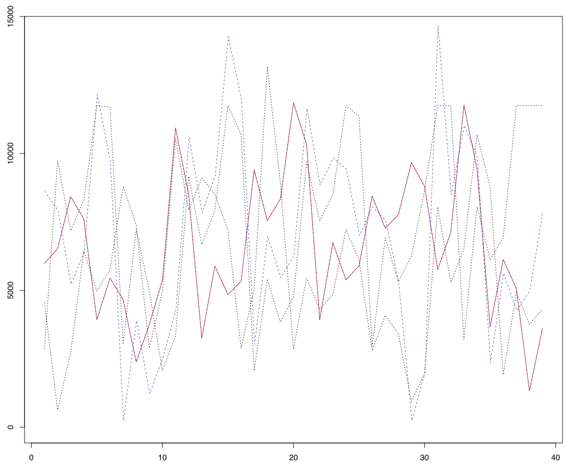


Figure 4: The lot-scheduling plan (plan and fact matching for two products).

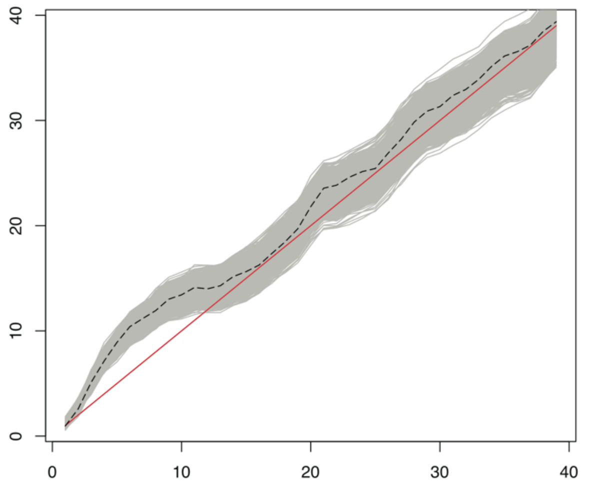


Figure 5: The reduced criterion variation obtained using the fractal forecast method.

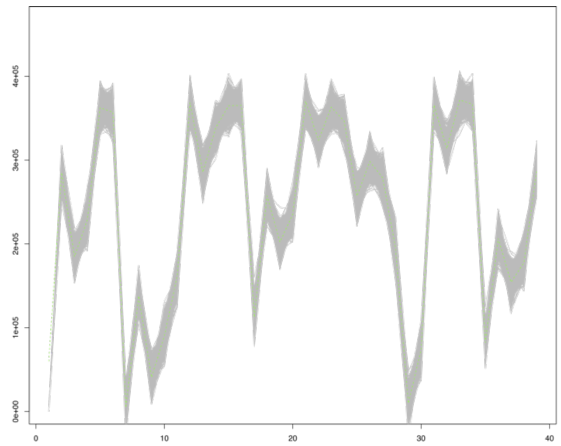


Figure 6: The requirement for one kind of item according to possible variation in the production plan.

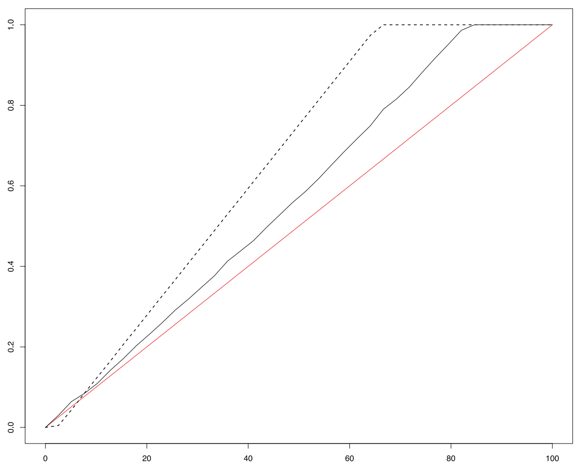


Figure 7: The comparison of loading (lot-scheduling) when implementing various production plans by ROC analysis.

The variation in production plan and management affects the production system. It is necessary to take into account during a process of management decision making. For example, in some cases, we can expect the system profitability increase with a decline in production value and facility/ warehouse/ staff loading. Thereby, the comparative analysis is needed. It would be made using the ROC curve (Figure 7).

6 CONCLUSION

The findings show, that the result depends on the forecast accuracy. It is worth noting that results do not consider delay and inertia factors, which take place in real production systems, load them more, and can cause an organizational change in production.

Despite it, the created models could be implemented. Expected that they improve the efficiency of the production system work during the transition to the virtual production and Industry 4.0 concepts. The described in this paper models take into account several factors such as energy and resource intensity of the production processes tending to increase.

The development of the model should be solving the tasks associated with:

- inertia factor
- costs of the production volume changing
- transport subsystem risks
- delivery of defective parts
- product returns.

ACKNOWLEDGMENTS

The authors thank the government of Perm Krai for the support of the project for “Development of software and economic and mathematical models for supporting innovation project management processes in production systems”, which is being implemented in accordance with decree №166-п of 06.04.2011.

The reported study was partially supported by the Government of Perm Krai, research project No. C-26/058.

REFERENCES

[1] A. Santander-Mercado and M. Jubiz-Diaz, ‘The economic lot scheduling problem: a survey’, *Int. J. Prod. Res.*, vol. 54, no. 16, pp. 4973–4992, Aug. 2016.

[2] M. Díaz-Madroñero, J. Mula, and D. Peidro, ‘A review of discrete-time optimization models for tactical production planning’, *Int. J. Prod. Res.*, vol. 52, no. 17, pp. 5171–5205, Sep. 2014.

[3] H. Jalali and I. V. Nieuwenhuyse, ‘Simulation optimization in inventory replenishment: a classification’, *IIE Trans.*, vol. 47, no. 11, pp. 1217–1235, Nov. 2015.

[4] M. E. Peralta Álvarez, M. Marcos Bárcena, and F. Aguayo González, ‘On the sustainability of machining processes. Proposal for a unified framework through the triple bottom-line from an understanding review’, *J. Clean. Prod.*, vol. 142, pp. 3890–3904, Jan. 2017.

[5] L. A. Mylnikov, A. V. Seledkova, and B. Krause, ‘Forecasting characteristics of time series to support managerial decision making process in production-And-economic systems’, *Proc. 2017 20th IEEE Int. Conf. Soft Comput. Meas. SCM 2017 6 July 2017*, pp. 853–855.

[6] L. A. Mylnikov, ‘Planning and Management Tasks of Innovation Projects in Production and Economic Systems’, in *Key Issues for Management of Innovative Projects*, B. L. Moya, M. D. S. de Gracia, and L. F. Mazadiego, Eds. InTech, 2017.

[7] L. Mylnikov and M. Kuetz, ‘The risk assessment method in prognostic models of production systems management with account of the time factor’, *Eur. Res. Stud. J.*, vol. 20, no. 3, pp. 291–310, 2017.

[8] L. Mylnikov, ‘The Management of Innovation Projects’ Implementation in Production-and-Economic Systems: The Use of Prognostic Models for the Risk Metrics of Obtained Decisions’, 2017.

[9] S. D. Tajbakhsh, E. del Castillo, and J. L. Rosenberger, ‘A Bayesian Approach to Sequential Optimization based on Computer Experiments’, *Qual. Reliab. Eng. Int.*, vol. 31, no. 6, pp. 1001–1012, Oct. 2015.

[10] K. S. Moghaddam, ‘Supplier selection and order allocation in closed-loop supply chain systems using hybrid Monte Carlo simulation and goal programming’, *Int. J. Prod. Res.*, vol. 53, no. 20, pp. 6320–6338, Oct. 2015.

[11] L. Mylnikov, ‘Risk Evaluation in Manufacturing Organization Tasks for Product Technological Projects and Establishment of Project Portfolio for Production Systems’, in *Proceedings of the 2016 International Conference on Applied mathematics, simulation and modelling*, Beijing, China, 2016, pp. 399–402.

[12] E. K. Burke, S. J. Landa, and E. Soubeiga, ‘Hyperheuristic Approaches for Multiobjective Optimization’, in *Proceeding of the MIC 2003*, Kyoto, Japan, 2003.

[13] Walter R. Mebane, Jr. and Jasjeet Singh Sekhon, ‘Package “rgenoud”’. 19-Jul-2015.

Computer-aided Control of Sensorimotor Skills Development in Operators of Manufacturing Installations

Rustam Fayzrakhmanov, Ivan Polevshchikov and Anatoly Polyakov
Perm National Research Polytechnic University,
Department for Information Technologies and Computer-Based Systems, Perm, Russia
fayzrakhmanov@gmail.com, i.s.polevshchikov@mail.ru, squarepants_07@mail.ru

Keywords: Training Simulation Complexes, Computer-Aided Training Systems, Sensorimotor Skills.

Abstract: In this paper we introduce a novel model and method to increase the effectiveness of training operators of manufacturing installations. The proposed mathematical model describes main aspects of the process of training and development of sensorimotor skills. In contrast to existing models, our mathematical model incorporates dynamic characteristics of the skill development process and exponential learning curves which are used in assessing the performance of motor actions and development of control actions. We implemented the model and method as a component of our AnyCrane training simulator for crane operators. AnyCrane training simulator is a framework for realistic simulation of environment in ports, cranes, personal, physics and different technological processes. Our case study demonstrated effectiveness of the approach presented in this paper. The time necessary for moving a cargo decreased by 19%, the precision of load-unload operations increased by 30%, and the smoothness of crane turning has been increased by 36%.

1 INTRODUCTION

An effective use of modern manufacturing installations is determined by the quality of training which operators had in the past. Low quality of technological processes and high level of accidents and injuries are often a consequence of insufficient development of necessary sensorimotor skills in operators.

An operator must have required skills to timely and accurately perform necessary technological operations according to the desired process trajectory, to have necessary knowledge, be able to evaluate the real course of the technological process and choose the most effective actions.

Therefore, the problem of increasing the effectiveness of training operators of technological processes in order to develop necessary level of professional sensorimotor skills within a relatively short period of time is of paramount concern. To achieve this goal different computer-aided training systems (CTSs) and computer simulators were developed.

However, despite the presence of a large number of research studies in the field of professional training automation, the problem of developing professional skills in operators of technological processes with the use of training simulators is not developed enough.

2 STATE OF THE ART

A lot of research is dedicated to the problem of automated management and control of the professional skills development and, in particular, professional sensorimotor skills of operators of technological systems. State-of-the-art research addresses the following aspects:
– problems of professional skills formation, in particular, mathematical modeling of skills developing processes [1] – [2];
– development and usage of training complexes for professional training of operators of various technological processes [3] – [4];
– models and methods for constructing CTSs and their individual components [5] – [6].

It is worth mentioning that there are still many open questions which are to be answered. For example, the most relevant to us is how to control the development of specific professional skills and adequately assess their level of maturity using computer simulators and serious games. Thus, it requires to have a full control on the training process selecting a specific set of assignments and setting a necessary level of their complexity.

An effective training with the use of simulators should take into account the following aspects:

– an algorithm for forming a set of relevant sensorimotor skills and based on the iterative nature of obtaining each skill;

– taking into account complex relations between various technological quality characteristics, used to access the success of training;

– defining the actual level of skill in case of incomplete initial data of established standards on performing technological operation;

– a large amount of motor functions combined with operating of several mechanisms at once.

3 MATHEMATICAL MODEL OF THE SKILLS DEVELOPMENT CONTROL SYSTEM

To solve this problem, a mathematical model of the automated control system for the development of operator's professional sensorimotor skills through the exercises using training simulation complexes (TSCs) is proposed. This model was integrated and tested on the AnyCrane framework [7].

We aim at the following problem: for the minimal training time (the execution of exercises designed for TSC), the performance of the technological operation must be brought to automatism at a given quality level. That is, $T_{\text{tr.}} = \sum_{g=1}^{N_{\text{ex.}}} T_g \to \min$, wherein for $\forall g = \overline{1, N_{\text{ex.}}}$:

$T_g = f_{\text{imp.}}(U_g^{\text{dec.}}, U_g^{\text{rek.}}) = \sum_{v=1}^{N_g^{\text{imp.}}} T_{gv}$ – the time T_g for multiple execution of the g-th exercise depends on the set of decisions $U_g^{\text{dec.}}$ at the end of each v-th execution and the set of advisory influences (recommendations) $U_g^{\text{rek.}}$ to the learner in the process of each v-th execution (T_{gv} – the time when the g-th exercise was performed);

$K_{gh} \geq K_{gh}^{\text{thr.}}$ – the quality of technological operation execution, determined by the parameters K_{gh} (where $h = \overline{1, N_{\text{par.}}}$), must comply with the standards, i.e. threshold values $K_{gh}^{\text{thr.}}$;

$N_g \geq N_g^{\text{scl.}}$ – the exercise is performed a specified $N_g^{\text{scl.}}$ number of times in a row at the required level of quality (the action is fixed to automatism).

The set of dependencies between mathematical model parameters is shown schematically in Figure 1.

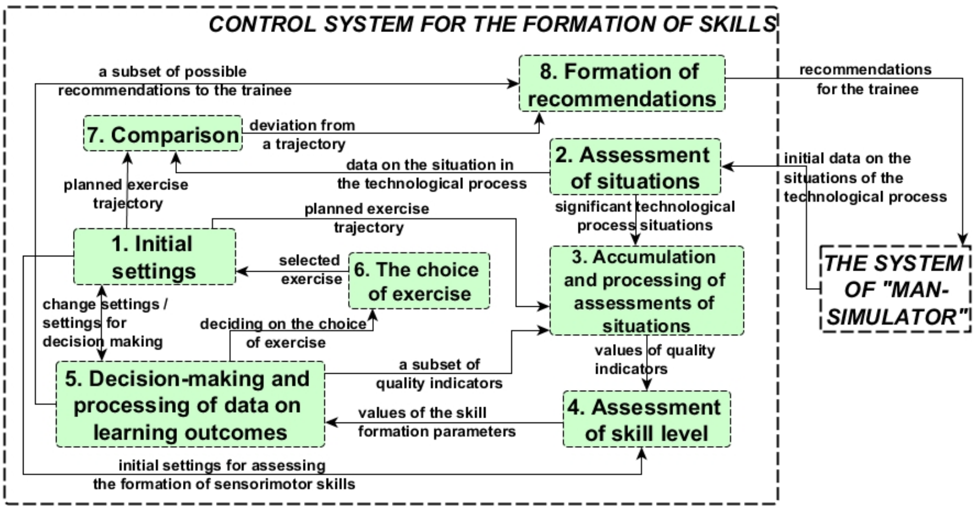


Figure 1: Scheme of control system for the formation of sensorimotor skills.

Block 3 (Figure 1) collects primary data on the process of sensorimotor skill formation (obtained in block 2), and on basis of processed data, the values of the quality indicators measured during the exercise are calculated, which constitute the set $X^{\text{used}}_{\text{qual.}} \subseteq X_{\text{qual.}}$, where $X_{\text{qual.}}$ is the set of all the quality indicators associated with the exercise ($|X_{\text{qual.}}| = N_{\text{qual.}}$). Quality indicators characterize the results and correctness of motor actions. Examples of indicators: the time of the operation; result accuracy; number of emergency situations; accuracy of the actions sequence compared with the optimal trajectory $T_{\text{pl.}}$.

In block 3, the following are formed: $P_{\text{trm.}} = \{P^{\text{trm.}}_r \mid r = \overline{1, N'_{\text{trm.}}}\}$ – set of terminal indicators values (calculated once after the completion of the exercise or part of it); $P_{\text{dyn.}} = \{P^{\text{dyn.}}_s(t_j) \mid s = \overline{1, N'_{\text{dyn.}}}, j = \overline{1, N^{\text{dyn.}}_s}\}$ – a set of dynamic indicators values (measured multiple times during the exercise) at time moment t_j ($N^{\text{dyn.}}_s$ – the number of measurements by the s-th indicator), wherein $|X^{\text{used}}_{\text{qual.}}| = N'_{\text{qual.}} = N'_{\text{trm.}} + N'_{\text{dyn.}}$.

Block 4 evaluates each step in the sensorimotor skill formation, the success of which is determined by whether the operator performed the exercise a specified number of times with the required quality or not. Thus, it calculates the degree of automatism in a trainee. Each new stage differs from the previous one by the addition of new estimated primary quality indicators.

The model's parameters (the coefficients of mastery) that determine the quality of the technological operation execution as the degree of compliance with the standards, represent the dimensionless values on the interval $[0;1]$ obtained by processing the data on the quality indicators calculated in block 3: $M_{\text{trm.}} = \{K^{\text{trm.}}_r \mid r = \overline{1, N'_{\text{trm.}}}\}$ and $M_{\text{dyn.}} = \{K^{\text{dyn.}}_s(t_j) \mid s = \overline{1, N'_{\text{dyn.}}}, j = \overline{1, N^{\text{dyn.}}_s}\}$ – sets of mastery coefficients by terminal and dynamic indicators, respectively; $M_{\text{agr.}} = \{K^{\text{agr.}}_s \mid s = \overline{1, N'_{\text{dyn.}}}\}$ – set of aggregated mastery coefficients by dynamic indicators, where

$$K^{\text{agr.}}_s = \sqrt[N^{\text{dyn.}}_s]{\prod_{j=1}^{N^{\text{dyn.}}_s} K^{\text{dyn.}}_s(t_j)}.$$

Values from $M_{\text{trm.}}$ and $M_{\text{dyn.}}$ are calculated, mainly, on the basis of Mamdani algorithm fuzzy inference procedure, where a system of production rules is constructed to compare each primary index of the mastery coefficient. Norms are defined by fuzzy sets determined as linguistic variables of the rules system. The usage of fuzzy inference required due to the incompleteness of the initial information on standards, identified with the involvement of experts [8].

The quality of technological operation performance as a whole is characterized by an integrated indicator – the complex mastery coefficient:

$$K_{\text{mast.}} = \frac{\sum_{r=1}^{N'_{\text{trm.}}} W^{\text{trm.}}_r K^{\text{trm.}}_r + \sum_{s=1}^{N'_{\text{dyn.}}} W^{\text{dyn.}}_s K^{\text{agr.}}_s}{\sum_{r=1}^{N'_{\text{trm.}}} W^{\text{trm.}}_r + \sum_{s=1}^{N'_{\text{dyn.}}} W^{\text{dyn.}}_s},$$ where

$W^{\text{trm.}}_r \in M_{\text{weig.}}$ and $W^{\text{dyn.}}_s \in M_{\text{weig.}}$, and $M_{\text{weig.}}$ – the set of weights of the quality indicators calculated in block 1.

Compliance with quality is reflected by the indicator $I_{\text{mast.}}$ of successful ($I_{\text{mast.}} = 1$) or unsuccessful ($I_{\text{mast.}} = 0$) exercise performance:

$$I_{\text{mast.}} = 1, \text{ if}$$
$$(\forall r = \overline{1, N'_{\text{trm.}}}, \Delta^{\text{trm.}}_r \geq 0) \wedge (\forall s = \overline{1, N'_{\text{dyn.}}}, \Delta^{\text{agr.}}_s \geq 0) \wedge (\Delta_{\text{mast.}} \geq 0);$$
$$I_{\text{mast.}} = 0, \text{ if}$$
$$(\exists r = \overline{1, N'_{\text{trm.}}}, \Delta^{\text{trm.}}_r < 0) \vee (\exists s = \overline{1, N'_{\text{dyn.}}}, \Delta^{\text{agr.}}_s < 0) \vee (\Delta_{\text{mast.}} < 0).$$

Here $\Delta_{\text{mast.}} = K_{\text{mast.}} - K^{\text{thr.}}_{\text{mast.}}$, $\Delta^{\text{trm.}}_r = K^{\text{trm.}}_r - K^{\text{trm.}}_{\text{thr.}\,r}$, $\Delta^{\text{agr.}}_s = K^{\text{agr.}}_s - K^{\text{agr.}}_{\text{thr.}\,s}$, where $K^{\text{thr.}}_{\text{mast.}}$, $K^{\text{trm.}}_{\text{thr.}\,r}$, $K^{\text{agr.}}_{\text{thr.}\,s}$ are the threshold values of the mastery coefficients determined in block 1.

Since the generated skill means learning execution of the technological operation at a given level of quality up to the automatism, a parameter

$\Delta_{\text{scl.}} = N_{\text{succ.}} - N_{\text{scl.}}$ is also calculated in block 4, where $N_{\text{succ.}}$ is the number of succeeded ($I_{\text{mast.}} = 1$) successive exercise executions with a given set of indicators, $N_{\text{scl.}}$ the minimum number of such executions in which the skill is considered formed, defined in block 1.

In block 5, the controls formation (decision making) $v_{\text{dec.}} = f_{\text{dec.}}(I_{\text{mast.}}, \Delta_{\text{scl.}})$ is carried out at the end of each and before the first exercise in case of necessity of: re-execution; modification of the measured quality indicators set; the determination of advisory influences subset in the repeated exercise execution.

To represent the logic-time features of $v_{\text{dec.}}$ controls generation with repeated execution of the exercises to form sensorimotor skills, an imitation model was constructed in the form of the Petri Net (Figure 2).

The use of the Petri network allows describing formalized decision-making algorithm (represented by the network positions) in order to organize the most effective information support for the learner, depending on the current skill level (determined by the conditions in the network transitions) [9].

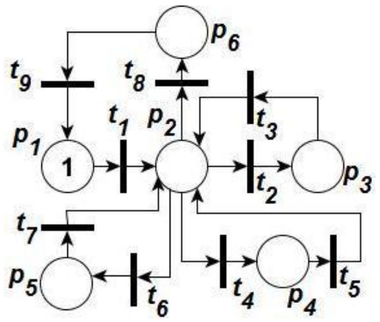


Figure 2: Petri Net.

Positions p_1 correspond to the initial subset formation of quality indicators that are evaluated when the exercise is performed, p_2 – the exercise is performed by the trainees. The position p_3 indicates the decision to repeat the exercise without modifying the subset of the indicators and introducing recommendations, p_4 – without modifying the subset of the indicators and without introducing recommendations, p_5 – with the modification of the subset of the indicators, p_6 –moving to the next exercise (selected in block 6).

Transition t_1 means that the trainee has started the first exercise, transitions t_3, t_5 t_7 mean that the trainee has started re-execution. The transition to the next exercise is designated as t_9. Transition t_2 means that the exercise is completed with the condition $I_{\text{mast.}} = 0$, t_4 – with the condition $(I_{\text{mast.}} = 1) \wedge (\Delta_{\text{scl.}} < 0)$, t_6 – with the condition $(I_{\text{mast.}} = 1) \wedge (\Delta_{\text{scl.}} = 0) \wedge (I_{\text{qual.}} = 1)$, t_8 – with the condition $(I_{\text{mast.}} = 1) \wedge (\Delta_{\text{scl.}} = 0) \wedge (I_{\text{qual.}} = 0)$, where $I_{\text{qual.}}$ is the indicator of availability ($I_{\text{qual.}} = 1$) or absence ($I_{\text{qual.}} = 0$) of quality parameters not yet used in successive exercises.

Modification of quality indicators subset is carried out in block 5 according to the dependencies

$$X''_{\text{qual.}} = X'_{\text{qual.}} \cup X^{\text{new}}_{\text{qual.}},$$

$$X^{\text{new}}_{\text{qual.}} = f(X_{\text{qual.}} \setminus X'_{\text{qual.}}, M_{\text{pr.}}),$$

where $X'_{\text{qual.}}$ – the current subset of used indicators (initially $X'_{\text{qual.}} = \varnothing$), $X''_{\text{qual.}}$ – the modified subset, $X^{\text{new}}_{\text{qual.}}$ – the subset of indicators with the highest priority among those not yet used, $M_{\text{pr.}} = \{W^{\text{pr.}}_i \mid i = \overline{1, N_{\text{qual.}}}\}$ – the set of indicators priorities calculated by the hierarchy analysis method in block 1.

A set of recommendations $M_{\text{rec.}}$ for re-executing the exercise is defined as

$$M_{\text{rec.}} = M^{\text{mast.}}_{\text{rec.}} \cup M^{\text{exp.}}_{\text{rec.}} \cup M^{\text{trm.}}_{\text{rec.}} \cup M^{\text{dyn.}}_{\text{rec.}},$$

where: $M^{\text{mast.}}_{\text{rec.}} = f^{\text{mast.}}_{\text{rec.}}(K_{\text{mast.}}, K^{\text{thr.}}_{\text{mast.}})$;

$M^{\text{exp.}}_{\text{rec.}} = f^{\text{exp.}}_{\text{rec.}}(K_{\text{mast.}}, K^{\text{exp.}}_{\text{mast.}})$, where

$$K^{\text{exp.}}_{\text{mast.}} = f_{\text{exp.}}(\gamma_{\text{scl.}}, N_{\text{imp.}});$$

$M^{\text{trm.}}_{\text{rec.}} = \{M^{\text{trm.}}_{\text{rec.}r} \mid r = \overline{1, N'_{\text{trm.}}}\}$, where

$$M^{\text{trm.}}_{\text{rec.}r} = f^{\text{trm.}}_{\text{rec.}}(K^{\text{trm.}}_r, K^{\text{trm.}}_{\text{thr.}r});$$

$M^{\text{dyn.}}_{\text{rec.}} = \{M^{\text{dyn.}}_{\text{rec.}s} \mid s = \overline{1, N'_{\text{dyn.}}}\}$, where

$$M_{\text{rec.}s}^{\text{dyn.}} = f_{\text{rec.}}^{\text{dyn.}}(K_s^{\text{agr.}}, K_{\text{thr.}s}^{\text{dyn.}}).$$

Here $\gamma_{\text{scl.}}$ is the recommended rate of skill formation (learning speed) that determines the planned trajectory of increasing the integral quality score when repeating the exercise and calculated in block 1 by the formulas: $\gamma_{\text{scl.}} = -\ln(1 - K_{\text{mast.}}^{\text{thr.}}) / N_{\text{scl.}}$ (at $K_{\text{mast.}}^{\text{thr.}} < 1$) or $\gamma_{\text{scl.}} = -\ln(1 - K_{\text{mast.}}^{\text{thr.}} + \varepsilon_{\text{mast.}}^{\text{thr.}}) / N_{\text{scl.}}$ (at $K_{\text{mast.}}^{\text{thr.}} = 1$), where $\varepsilon_{\text{mast.}}^{\text{thr.}} \in [0;1]$ is the permissible deviation from $K_{\text{mast.}}^{\text{thr.}}$ while calculating $\gamma_{\text{scl.}}$.

In block 5, after exercise completion $K_{\text{mast.}}^{\text{exp.}} = 1 - e^{-\gamma_{\text{scl.}} N_{\text{imp.}}}$ value of the integral quality index is calculated, which must be achieved after performing the exercise $N_{\text{imp.}}$ times and necessary to determine the deviation of the actual integral quality index change trajectory from the calculated one. The type and values of the learning curve parameters for the calculation $K_{\text{mast.}}^{\text{exp.}}$ can be changed based on the accumulated statistics on the operators learning outcomes using data mining techniques.

In block 8, the generation of recommendations (advisory influences) $v_{\text{rec.}}(t_z)$ in the exercise execution is made according to the dependence $v_{\text{rec.}}(t_z) = f_{\text{rec.}}(\Delta_{\text{sit.}}(t_z), T_{\text{pl.}}, M'_{\text{rec.}})$, where $M'_{\text{rec.}}$ – a set of possible recommendations, $M_{\text{rec.}} \subseteq M'_{\text{rec.}}$ – recommendations based on the results of the previous execution, $\Delta_{\text{sit.}}(t_z)$ – the deviation of the exercise (obtained in block 7) from the calculated trajectory $T_{\text{pl.}}$. Recommendations are signals indicating various modalities (tactile, audible, visual) about the moments when learner's motor actions performed, which are fed to one of the less loaded analyzers.

The mathematical model of the control system served as the basis for the architecture of TSC for the portal crane operators, functional and non-functional requirements for the software of the CTS, database structures [7, 10]. The difference of the created TSC architecture is the availability of software modules that provide the collection and processing of data on the formation of sensorimotor skills among operators.

4 EXPERIMENT

A training course for the practice of portal crane operators has been created, including a set of exercises for AnyCrane to acquire a necessary sensorimotor skills set. The training course includes exercises for improving the lifting controlling and load lowering skills, exercises for testing skills of cargo transfer controlling by changing the angle of jib rotation etc.

Each exercise in AnyCrane is implemented in the form of training tasks sequence. For the learning task, the actions performed by the learner and the observed results of actions are determined.

An experiment was conducted to prove the effectiveness of TSC application as a mean to effectively form professional sensorimotor skills.

There are two groups of trainees: experimental and control, 20 people each. Both groups successfully passed the knowledge testing at the end of the theoretical course.

The first (control) group was trained at the TSC, in which the sensorimotor skills management system was not implemented. The second (experimental) group was trained at TSC with a management system implemented in the form of program modules.

At the end of the training practical stage, during the experiment, each of the trainees moved cargo 5 times. Thus, in each group, 100 cargo transfers were carried out during control exercise execution at the TSC.

The following indicators to evaluate the performance of trainees were used:

– time taken for performing the technological operation, in seconds;

– cargo installation accuracy (cargo deviation relative to the special platform center), in percent of the platform radius;

– smoothness of the crane jib turn (angle of the load deviation from the vertical axis of the jib), in degrees.

Even under unchanged external conditions operator obviously cannot repeat exactly his actions several times in a row. We can say that the operator action with each control is random to a certain extent. Consequently, the operator quality and efficiency can only be estimated on average statistical data, and experiments under the same conditions must be repeated many times [10].

Therefore, for each of the three indicators listed above, the following data were obtained: $x = (x_1, x_2, x_3, ..., x_{100})$ – sample for the experimental group, $y = (y_1, y_2, y_3, ..., y_{100})$ –

sample for the control group, where x_i ($i = \overline{1,100}$) and y_j ($j = \overline{1,100}$) are the sampling elements, i.e. values obtained during the experiment.

To justify the difference in the control and experimental groups states, we use the Cramer-Welch statistical criterion, since the data are measured in the ratio scale.

The empirical value of the Cramer-Welch criterion is calculated on the basis of information about volumes M and N, values x and y, means $\overline{x}$ and $\overline{y}$, variances D_x and D_y of compared samples for each of the three indicators by the formula $T_{\text{emp.}} = \frac{\sqrt{M \cdot N}\left|\overline{x} - \overline{y}\right|}{\sqrt{M \cdot D_x + N \cdot D_y}}$. The results of calculations are presented in table 1.

Table 1: Results of the experiment.

Calculated value	First group	Second group
Time of the operation performing, s		
Average	61,2	49,5
Dispersion	90,8	30,9
$T_{\text{emp.}}$	10,53	
Accuracy of cargo installation, %		
Average	12,6	8,8
Dispersion	14,9	9,2
$T_{\text{emp.}}$	7,86	
Smoothness of the jib rotation, °		
Average	16,9	10,9
Dispersion	19,7	8,0
$T_{\text{emp.}}$	11,38	

As we can see in Table 1, the time of cargo transfer technological operation was reduced by 19%, the accuracy of the cargo installation increased by 30% and the smoothness of the crane turning increased by 36%.

The calculated value of the criterion in each case is $T_{\text{emp.}} > 1{,}96$, so the reliability of differences in the characteristics of the control and experimental groups after the end of the experiment is 95%. It can be concluded that the effect of changes follows from the usage of developed by the TSC with the appropriate software modules of a sensory-motor skills management system in the training process.

5 CONCLUSIONS

The most significant results of the study are:

– A mathematical model of the automated control system for developing the professional sensorimotor skills in operators of technological processes was developed. The model differs from the known ones by using new parameters when determining the control actions, taking into account repetitions of exercises, provides the necessary level of information support for the trainee and improves the efficiency of acquiring the ability to self-control the quality of performing technological operations.

– A technique of monitoring and controlling the training process is developed. The technique differs from the existing ones by using the original algorithms for trainees' informational support in the sensorimotor skills formation process.

– Models and approaches were implemented and integrated into the AnyCrane framework.

– The advantage of using TSC with implemented software modules for controlling the sensorimotor skills formation in training portal crane operators has been experimentally shown: the time of performing the load transfer technological operation has decreased by 19%, the cargo installation accuracy has increased by 30% and the smoothness of the crane turning has increased by 36%.

Further possible research in this field:

– improvement of algorithms for assessing the correspondence between the real and calculated trajectories of technological operations execution modeled in the TSC

– a recommender system to provide a trainee with suggestions and information to help him improve his performance.

ACKNOWLEDGMENTS

The research is supported by a stipend of the President of the Russian Federation to young scientists and post-graduate students (No. SP-100.2018.5), which was assigned by the grants Council of the President of the Russian Federation for government support of young Russian scientists and government support for the leading scientific schools of the Russian Federation in 2018-2020.

REFERENCES

[1] D.A. Novikov, "Collective learning-by-doing", 9th IFAC Symposium on Advances in Control Education (ACE 2012), pp. 408-412, 2012.

[2] L. Lisitsyna and A. Lyamin, "Approach to development of effective e-learning courses", Frontiers in Artificial Intelligence and Application, vol. 262, pp. 732-738, 2014.

[3] V.M. Dozortsev, "Methods for computer-based operator training as a key element of training systems (present-day trends)", Automation and Remote Control, vol. 74(7), pp. 1191-1200, 2013.

[4] V.M. Dozortsev and V.A. Nazin, "Training of fault diagnosis skills: use of heuristic diagnostic model and technical system simulation", 9th International Industrial Simulation Conference (ISC 2011), pp. 273-278, 2011.

[5] I. Veshneva, R. Singatulin, A. Bolshakov, T. Chistyakova, L. Melnikov, "Model of formation of the feedback channel within ergatic systems for monitoring of quality of processes of formation of personnel competences", International Journal for Quality Research, vol. 9(3), pp. 495–512, 2015.

[6] D. Bouhnik and G. Carmi, "E-learning Environments in Academy: Technology, Pedagogy and Thinking Dispositions", Journal of Information Technology Education: Research, vol. 11, pp. 201-219, 2012.

[7] R.A. Fayzrakhmanov, I.S. Polevshchikov, A.F. Khabibulin, F.I. Shkliaev and R.R. Fayzrakhmanov, "ANYCRANE: Towards a better Port Crane Simulator for Training Operators", Proc. of the 15th International Industrial Simulation Conference (ISC'2017), Warsaw (Poland), pp. 85-87, 31 May – 1 June 2017.

[8] S.O. Azarkasb, "An Efficient Intrusion Detection System Based on Fuzzy Genetic approaches", Life Science Journal, vol. 10(8s), pp. 6-21, 2013.

[9] Sayed Taha Mohamed, Mohamed Abdel Gawad Mostafa and Ahmed Fathi Mohamed, "A comparative study on Petri Nets in manufacturing applications", Life Science Journal, vol. 10(1), pp. 1496-1502, 2013.

[10] R. Fayzrakhmanov, I. Polevshchikov and A. Khabibulin, "Computer Simulation Complex for Training Operators of Handling Processes", Proc. of the 5th International Conference on Applied Innovations in IT (ICAIIT), Koethen (Germany), vol. 5, pp. 81-86, 16 March 2017.

The Use of Optimal Management Tasks for Verification and Adjustment of New Product Release Planning in Discrete Production Systems

Leonid Mylnikov, Dmitrii Vershinin and Daniil Fatkhullin
Perm National Research Polytechnic University, 29 Komsomolsky avenue, 614990 Perm, Russia
leonid.mylnikov@pstu.ru, nsenz@yandex.ru, danil329@yandex.ru

Keywords: Lot-Scheduling Planning, Optimal Management, Discrete Production, Project, Production System, Engineering Company, Forecast, Prediction, Business Plan, Smart Manufacturing

Abstract: The present paper investigates a modern issue of predictive models for optimal management used to enhance the performance of production system management that can be achieved by a joint consideration and synchronization of internal and external processes of an examined system. Volume planning task is considered as a task that helps verify the results of business planning and take into account the interrelation of subsystems of a production system and foreign market impact based on forecasting data. The article draws on the example of release of leading-edge vacuum pumps in an engineering company in order to define the prospects of this market and estimate manufacturing capabilities. The analysis was carried out based on the data of an approximate business plan and statistical data of vacuum pump market. As a result, it suggests production schedule for vacuum pumps that can be taken as a background for making feasibility decisions on the release of new products and adjusting production activities of an enterprise. The obtained business planning data can be used in practice by solving the tasks of production management that help perform preliminary estimates of enterprise potentiality for market needs and improve the objectivity of strategic decision making by enhancing the formalization level of describing processes and preparing objective data. The synchronization of production processes described in the paper is relevant as it is connected with the current trends, i.e. the reduction of time production, the depreciation of human factor in production processes, all that triggers increased requirements to the quality of management in production processes.

1 INTRODUCTION

Currently, we observe a trend when life cycle is reducing and the range of products is increasing. In discrete productions it leads to new problems of search and integration of new products, product line, modification selection. For all that, decision making is affected by external market environment that is not predetermined yet. The task gets more complicated because of the continuity of production process that consists in the peculiarities of a performed operation, such as in-line assembly, stamping operations, casting, pressing, etc. From the other hand, there is a trend for enhancing the versatility of production tools that sets advanced requirements to management decision making and planning as the interchangeability of industrial machinery causes the invariance of part motion and operations' performance.

Finally, production planning task is connected with the management of production resources necessary for utilizing primary products and parts for the production of final products that have best compliance with customer needs.

Planning and management tasks are considered on the tactical and strategical levels.

The target of a tactical level is to identify the most efficient way of utilizing procured resources (materials, stocks, capacities) with the planning horizon from one or several months up to two years in general (depending on the life cycle of released products). Such tasks are solved by working out operation plans for the procurement of primary products and cargo, transport routes, etc.

The tasks of prospect product portfolio are tackled on the strategical level.

The matching of operation plans with the tasks of project portfolio planning and changes connected with the performance of production systems helps increase management efficiency and decrease management and loss expenses that are caused by the non-coordination of managerial decisions.

The economic task of batch planning combines the decisions on the batch size and production planning. It is difficult to find an optimal solution for this task but it helps increase the capacities of a company.

Production planning tasks have been investigated for over 70 years as they have practical significance. Today, we have both precise methods and approximate deterministic and probability models. Along with a vast practical application, this task has scientific potential since production planning requires complex applications of approaches, methods and models.

Many authors talk about a special function of management, the extended formalization of management principles, the increase in «flexibility» within existing approaches or possibilities to develop new comprehensive methods and models.

The significance of complex estimates with high quality is increasing. Hence, accumulated volumes of collected information bring the tasks of monitoring and parameter forecasting; under current conditions, we need to apply tracking systems to control the dynamics of parameter changes and the matching of their values with the planned values [1].

At the same time, we observe shortage of methodological approaches for the formalization of project management tasks in production systems despite the fact that this area has been investigated for a long time. The methods of objective decision making with limited time factor applied for solving planning and management tasks encounter the problem of NP-completeness; therefore, many of these tasks can be solved only with help of approximate methods.

The development of the concept Industry 4.0 and IIoT that enables the collection of information about each equipment unit and operative control over production processes in production systems [2], on the one hand, presents new possibilities for developing methods of industrial engineering, on the other hand, it causes the problem of big data processing.

Furthermore, the process of achieving target indicators is not a single-stage process, yet it is a series of interdependent states. More than that, target indicators can change in time and can represent a set of differently related values [3].

2 MANAGEMENT TASK SETTING

There are studies of planning the release of new products, i.e. the determination of product costs, target market volume; as a result, business plan with target indicators is obtained. Despite the existing practice to use business planning, it should be noted, that business plans do not consider the task from the standpoint of immediate planning and production planning, procurement and sales, the allocation of production capacities and production plan adjustments in enterprises (in big production systems). Nevertheless, the application of business planning (see Table 1) (institutional management level) in the tasks on the managerial and technical levels of management can help verify the feasibility of proposed plans, sustainability of obtained results and development of alternative implementation variants.

Table 1: An example of business planning results for two types of new products.

Quarter	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Sales volume (item)																				
Turbomolecular pump	3	5	6	8	8	9	11	11	12	14	15	16	16	16	17	18	19	19	20	20
Cryogenic pump	2	3	5	6	9	11	12	14	15	16	17	19	20	20	20	21	21	21	21	22
Net profit (thou. rub.)																				
Turbomolecular pump	-3892	-2449	-1721	-150	-710	119	1854	1900	84	1687	2812	3281	3375	3559	4584	5653	6767	7029	8218	8513
Cryogenic pump	-2367	-1965	-880	-440	1487	2681	2967	3856	2888	3580	4295	5477	6243	6456	6559	7138	7252	7368	7486	8110

Applying such approach as an approach for the management on the basis of a limited parameter set and building structures of collective decision-making was shown in the works by R. Sah and J. Stiglitz.

In most cases, building a task «from nothing» is based on the application of best practices and analyst expertise.

One of the most wide-spread approaches is the application of optimization tasks.

For the mathematical formalization of optimization task, it is necessary to determine calculation parameters $Z = AX$, where Z – the parameter vector (including computable parameters), A – the incidence matrix, X- the vector of model parameters (including the parameters obtained on the basis of forecasting data with the risk $R(\varepsilon)$ and the precision ε).

$$J\{Z\} \to opt$$
$$f_i(Z) : G_i, i = \overline{1, n},$$

where - : one of the operations $>, <, \geq, \leq, =, \neq$; i – the number of forecast-based parameter imposed restrictions of a model; G_i – the values for restrictions; $f_i(Z)$ – the function with parameters. Formalization examples can be found in the references (for instance, see [4]).

From the standpoint of production system, project management task consists in the task of maximizing profits by selecting a product portfolio. However, not all economically effective goods can be produced at a certain enterprise due to its technological capabilities. In order to consider all these peculiarities, we should select feasible suggestions. Hence, we can consider the task of volume planning. The target of this task is to receive optimal production plan by maximum net profit and minimum costs. Besides, it should be taken into account, that production time depends on the time of parts' production and the time of product assembly, and the number of manufactured products should be placed within the bounds from break-even point up to the number of goods that the market is able to consume. Net profit of sold products depends on revenue and costs of manufacture. Also, it should be noted, that there should be a sufficient amount of parts at the time of manufacturing of products.

In this case, volume planning task with time restricted parameters [5] can be presented as follows:

$$\sum_n (C_n(t) - (Q_m(t) + I_n(t))) \times X_n(t) \to max$$

$$X_n^H(t) \leq X_n(t) \leq X_n^B(t)$$
$$(C_n(t) - (Q_m(t) + I_n(t))) \geq 0$$
$$D_{nm}(t) \geq Z_{nm}$$
$$S_{pc} \geq S_n(t) \geq 0$$

where X_n – the release volume of the product n, $C_n(t)$ – the revenue from the sold products n, Q_m – the costs of parts' manufacture, I_n – the assembly costs of the product n, X_n^H – the number of the products n to meet break-even requirements, X_n^B – the number of the products n that the market is able to consume, n – the type of product, m – the set of parts, t – the time, S_n – the number of available machinery, S_{pc} – the number of machinery, D_{nm}- the matrix of existing parts that can be shown as follows:

A set of parts	1	…	m
The number of parts	D_{n1}		D_{nm}..

Z_{nm} - the matrix of parts required for manufacture, that is given as follows:

A set of parts	1	…	m
The number of parts	Z_{n1}	…	Z_{nm}.

It should be considered, that the restrictions include also the functions of time. It can be explained by the fact that production system descriptions can alter in time, procurement schedule can change, and the volume of requires allotted for the implementation of a certain projects can also adjust, etc.

The interrelation with other tasks, such as power supply system management, stock control, HRM, procurement management, sales management, utilization management, management of works sequence and paralleling [6] presumes the use of total resources (production means, parts/ components/ materials, return of goods, reused parts/ components/ materials, etc. see Table 2).

By calculating weakly formalized factors it is necessary to simulate the deviations of parameter values and observe their impact on the performance of system. Such approach allows define most probable development scenarios (we take into account a joint impact of deviations on a series of interrelated tasks and parameters), estimate the probability of scenarios implementation on business planning, and define the way these scenarios are affected by certain parameters.

Table 2: An approximate structure of process chart for the production of pilot turbomolecular vacuum pumps.

№	Operation	Time	Parts and components	Amount	Measurement unit	Machinery	Outcome	Personnel
1	Processing of rotor forging	T^1_{12}	Rotor forging	Part$_{1231}$	unit	Metal-working machinery	Processed rotor forgings	Machine operator
2	Pressing and punching of rotor shaft	T^1_{21}	Cylindrical work	Part$_{2141}$	unit	Metal-forming and punching machinery	Unprocessed rotor shaft	Puncher
3	Thermal processing of shaft	T^1_{22}	Unprocessed rotor shaft	Part$_{2251}$	unit	Furnace for metal normalizing	Normalized shaft	Furnace tender
4	Mechanical processing of shaft	T^1_{23}	Normalized shaft	Part$_{2331}$	unit	Metal-working machinery	Rotor shaft (Z_{12})	Machine operator
5	Processing of stator forgings	T^1_{12}	Stator forgings	Part$_{3231}$	unit	Metal-working machinery	Processed stator forgings	Machine operator
6	Manufacture of ceramic racers	T^1_{41}	Silicon nitride	Part$_{4161}$	unit	Turning machinery	Ceramic racers	Lathe operator
7	Manufacture of ceramic bearing balls	T^1_{42}	Silicon nitride	Part$_{4271}$	unit	Processing machinery	Ceramic bearing balls	Machine operator
8	Punching of pump body	T^1_{51}	Billet	Part$_{5121}$	unit	Metal-forming machinery	Punched pump body	Puncher
9	Processing of pump body	T^1_{52}	Punched pump body	Part$_{5231}$	unit	Metal-working machinery	Pump body (Z_{15})	Machine operator
10	Manufacture of rotor forgings	T^1_{11}	Cast metal ingot	Part$_{1111}$ Part$_{1121}$	unit	Metal-cutting machinery Metal-forming machinery	Rotor forging	Machine operator Puncher
11	Manufacture of blades on rotors	T^1_{13}	Processed rotor forgings	Part$_{1311}$ Part$_{1331}$	unit	Metal-cutting machinery Metal-working machinery	Rotor disks (Z_{11})	Machine operator
12	Manufacture of stator forgings	T^1_{31}	Cast metal ingot	Part$_{3121}$ Part$_{3111}$	unit	Metal-forming machinery Metal-cutting machinery	Stator forging	Puncher Machine operator
13	Manufacture of blades on stators	T^1_{33}	Processed stator forgings	Part$_{3311}$ Part$_{3331}$	unit	Metal-cutting machinery Metal-working machinery	Stator disks (Z_{13})	Machine operator
14	Assembly of ceramic bearings	T^1_{43}	Ceramic bearing races Ceramic bearing balls	Part$_{4381}$	unit	Assembly facilities	Ceramic bearings (Z_{14})	Assembly worker
15	Assembly of cryogenic pump	T^1_{61}	Rotor shaft (Z_{12}) Pump body (Z_{15}) Rotor disks (Z_{11}) Ceramic bearings (Z_{14}) Stator disks (Z_{13})	Part$_{6181}$	unit	Assembly facilities	Ceramic bearings (Z_{14})	Assembly worker

3 MANAGEMENT TASK SOLUTION

The task presented in the previous section generally can become a multi-parameter task with nonlinear restrictions when a part of parameters is assigned by the functions of time, and they can be discrete, continuous and stochastic [7].

Depending on the task received, several approaches can be used, that help obtain both a precise and an approximate solution (see [8]). A universal solution of such tasks is the application of heuristic methods that do not impose limits on target function; this solution enables perform multiple calculations by using statistical methods for simulating deviations and discrete time (the latter one is simulated in accordance with the principle Δt or by allocating special states – completion of certain works and commencement of other works [9]). Evolutionary programming is one of the most wide-spread methods (for example, genetic algorithm). A significant advantage of these methods is a possibility to assign calculation time that can be allotted for calculation performance. It becomes especially important by the necessity of conducting multiple calculations when time for decision making is limited. That is why, industrial engineering task is solved with help of the library rgenoud for the language R (see Listing 1).

Listing1: The function that implements optimization task solutions in the language R.

```
opt<-function(v1_, v2_, pump1, pump2,
ss, niz1, niz2, verh1, verh2, n1, n2){
  n2<-1
  n1<-1
  niz1<-floor(rnorm(1, mean = 8, sd=1))
  niz2<-floor(rnorm(1, mean = 9, sd=1))
  verh1<-floor(rnorm(1, mean = 20,
sd=8))
  verh2<-floor(rnorm(1, mean = 18,
sd=7))
  y<-NA
  if (v1_*n2>v2_*n1) {f<-
genoud(function(y) v1_*y[1]+v2_*y[2],
nvars=2, max.generations = 10,
wait.generations = 10, Domains =
matrix(c(min(niz1,min(verh1, pump1,
ss))-0.1, min(verh1, pump1, ss),
min(niz2,min(verh2, pump2, ss-
min(verh1, pump1, ss)))-0.1, min(verh2,
pump2, ss-min(verh1, pump1, ss))),
ncol=2, byrow=TRUE), max = TRUE,
data.type.int = TRUE)}
  else {f<-genoud(function(y)
v1_*y[1]+v2_*y[2], nvars=2,
max.generations = 10, wait.generations
= 10, Domains =
matrix(c(min(niz1,min(verh1, pump1, ss-
min(verh2, pump2, ss)))-0.1, min(verh1,
pump1, ss-min(verh2, pump2, ss)),
min(niz2,min(verh2, pump2, ss))-0.1,
min(verh2, pump2, ss)), ncol=2,
byrow=TRUE), max = TRUE, data.type.int
= TRUE)}
  return(f)
}
```

In the examined case business planning data can alter depending on the unfolding situation. That is why, we need to consider for planning data possible deviations of their values within confidence interval (if we assume, that planning data are valid (true), the deviations will obey the normal distribution law); hence, the algorithm for task solution will be as presented in the Fig. 1.

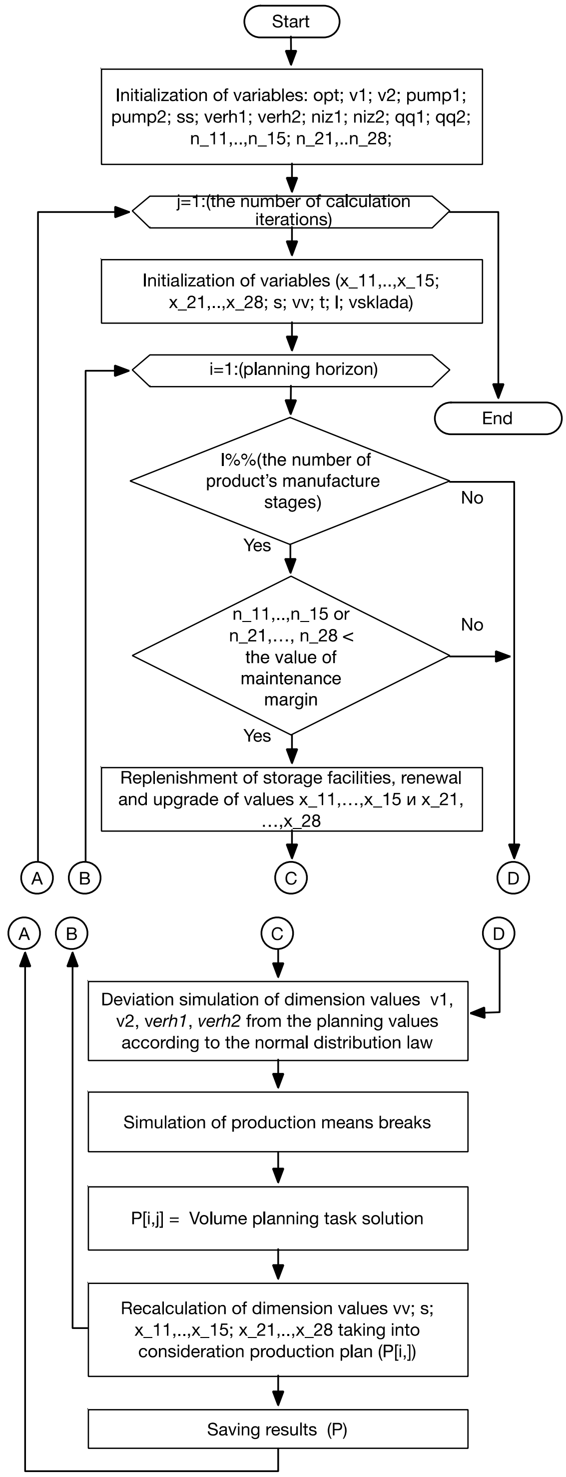


Figure 1: The algorithm of volume planning task solution taking into account possible deviations of parameter values (*v1*, *v2* – the values of net profit from the products 1 and 2; *pump1*, *pump2* – the number of the products 1 and 2, that can be manufactured from the procured number of parts/

components/ materials; *ss* – the number of available production means; *verh1*, *verh2* – the restriction value for market capacity for the products 1 and 2; *niz1*, *niz2* – the value of break-even point for the products 1 and 2; *qq1*, *qq2* – the pump production time of the products 1 and 2; n_11,…,n_15 – the number of different parts/ components/ materials required for the manufacture of the product 1; x_11,…,x_15 – the number of different procured parts/ components/ materials for the manufacture of the product 1; x_21,…,x_28 – the number of different procured parts/ components/ materials for the manufacture of the product 2; s – the matrix of available production means; vv – the matrix of production means that become available on the next stage; vsklada – the size of storage facilities. n_21,…,n_28 – the number of parts required for the manufacture of the pump 2).

4 THE ANALYSIS OF OBTAINED RESULTS

By solving optimal management tasks taking into consideration time factor and a certain discrete time step, the solution will present a tabular function. According to Bayesian theorem, the probability of an effective transition into a new state (to a new solution) will depend on the previous state (the state that we are placed in). At the same time, the simulation of parameter deviations will help obtain a set of solutions for each step (taking the total probability of transition to the next step by 1 taking into account iterations of possible solutions); hence, occurrence probabilities of each of possible states $s_1, s_2, \cdots, s_n$ – are equal to:

$$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{12} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nm} \end{pmatrix},$$

where n – the maximum amount of states on the steps m.

In this case, management task adds up to the selection of a desired state from the set of possible states and to path determination (a series of intermediate states) for its achievement. As a result, we can define the probabilities for obtaining solutions. Therefore, a solution represents a set of project development paths (decision tree) (see Fig. 2) that can be considered as a Bayesian network, that can be used for selecting most probable development path on the basis of the method of dynamic programming (Bellman's method) and matched with planning results for their verification or disproof.

Decision tree nodes in the Fig. 2 (for the task of volume production planning) match to the combinations of release volumes of products (for example, for 20 time steps and 100 calculations in each time station for a joint release of two types of products see Table 3).

The obtained values will have a different number of iterations (in the examined examples from 3 up to 38), which means, that the probabilities of different states will be also different. As shown in Table 4, in several cases invariant states are encountered. Besides, if we match most probable states with business plan dates (Table 4), they most likely do not match; this observation is demonstrated by the results of multiple calculations, which proves that business planning does not consider production system peculiarities, that show up in planning tasks as optimization tasks in their restrictions.

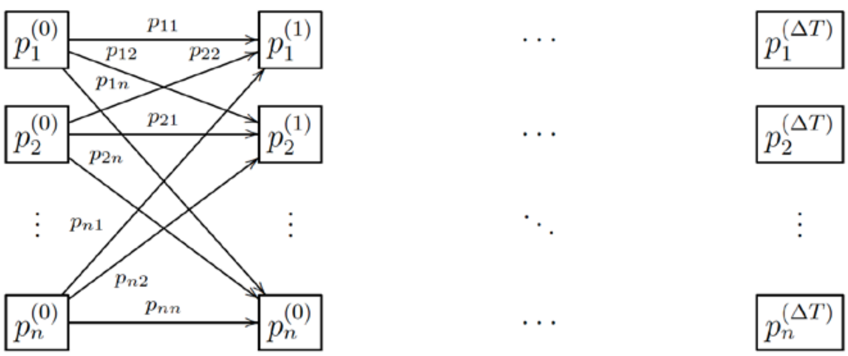


Figure 2: The decision tree for selecting the development path of production system or project implementation.

Table 3: Production volume plan for two products taking into account restriction deviations in market size and net profit by step-by-step increase of allotted production funds according to the business plan.

Calculation number	Time (number of planning month)								
	1		**2**		**3**		**…**	**20**	
1	5	0	5	3	4	5	0	5	3
2	5	0	5	3	8	5	0	5	3
3	5	0	5	3	8	5	0	5	3
4	5	0	5	3	8	5	0	5	3
5	5	0	5	3	8	5	0	5	3
6	5	0	5	3	8	5	0	5	3
7	5	0	0	5	6	5	0	0	5
8	5	0	5	3	8	5	0	5	3
…	…	…	…	…	…	…	…	…	…
100	5	0	8	0	11	5	0	8	0

The calculation results can lead to phenomena when upon removing restrictions on production funds the release of production units changes (for example, first only one type of goods is released and afterwards we start the release of another type of goods), and by the combination of restrictions the release of only one type of goods is deemed optimal.

Such phenomenon justifies the significance of a proper implementation of an adjustment process and market entry taking into account not only state of market, but also the processes that run inside production system. Finally, such verification can indicate the possibility of a backward approach – draw a plan for market entry based on the volume planning task, perform market studies on its basis and set a task.

Table 4: The matrix of state probabilities for time steps without taking into account their interdependence in time (most probable states are marked dark grey, the results received by business planning are marked light grey) with production funds restraint according to the plan of production volume growth that was obtained as a result of business planning.

Time (number of a planning month)																			
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
The probability of system to enter the examined state s_i																			
0,54	0,41	0,50	0,52	0,41	0,40	0,46	0,02	0,47	0,43	0,01	0,43	0,01	0,38	0,29	0,37	0,01	0,29	0,01	0,03
0,45	0,48	0,48	0,48	0,01	0,49	0,41	0,34	0,01	0,01	0,44	0,27	0,01	0,25	0,42	0,21	0,28	0,01	0,26	0,02
0,01	0,01	0,01	-	0,47	0,01	0,01	0,43	0,42	0,31	0,01	0,01	0,32	0,02	0,02	0,01	0,28	0,01	0,02	0,04
-	0,02	0,01	-	0,02	0,01	0,01	0,03	0,01	0,01	0,31	0,01	0,01	0,05	0,02	0,01	0,02	0,26	0,25	0,22
-	0,02	-	-	0,01	0,01	0,02	0,01	0,01	0,01	0,02	0,02	0,01	0,02	0,01	0,01	0,04	0,01	0,01	0,22
-	0,01	-	-	0,01	0,01	0,01	0,02	0,02	0,01	0,02	0,03	0,01	0,02	0,02	0,01	0,02	0,01	0,02	0,01
-	0,01	-	-	0,02	0,01	0,01	0,03	0,01	0,01	0,02	0,02	0,27	0,01	0,01	0,01	0,03	0,01	0,01	0,02
-	0,01	-	-	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,01	0,01	0,01	0,03	0,02	0,02
-	0,01	-	-	0,03	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,03	0,01	0,01
-	0,01	-	-	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,02	0,01	0,02	0,02	0,03	0,02
-	0,01	-	-	-	0,01	0,01	0,01	0,01	0,03	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,01
-	-	-	-	-	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,03	0,02	0,02	0,01	0,02	0,01	0,02	0,01
-	-	-	-	-	-	0,01	0,01	-	0,02	0,01	0,02	0,02	0,01	0,01	0,03	0,02	0,01	0,03	0,01
-	-	-	-	-	-	0,01	0,02	-	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,02	0,01	0,01
-	-	-	-	-	-	-	0,01	-	0,01	0,01	0,01	0,03	0,01	0,01	0,02	0,01	0,01	0,02	0,02
-	-	-	-	-	-	-	0,01	-	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,01	0,02	0,01	0,02
-	-	-	-	-	-	-	0,01	-	0,01	0,01	0,02	0,02	0,01	0,01	0,01	0,02	0,02	0,02	0,04
-	-	-	-	-	-	-	0,01	-	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,02
-	-	-	-	-	-	-	-	-	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,03	0,01	0,03	0,01

-	-	-	-	-	-	-	-	-	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
-	-	-	-	-	-	-	-	-	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,02	0,01	0,04
-	-	-	-	-	-	-	-	-	0,01	0,01	0,01	0,03	0,01	0,01	0,01	0,01	0,01	0,01	0,01
-	-	-	-	-	-	-	-	-	0,01	0,01	0,01	0,02	0,02	0,01	0,01	0,01	0,01	0,02	0,02
-	-	-	-	-	-	-	-	-	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,01	0,01	0,01	0,01
-	-	-	-	-	-	-	-	-	-	-	-	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01
-	-	-	-	-	-	-	-	-	-	-	-	0,02	0,01	-	0,01	0,01	0,01	0,01	0,01
-	-	-	-	-	-	-	-	-	-	-	-	0,01	0,01	-	0,01	0,01	0,01	0,01	0,02
-	-	-	-	-	-	-	-	-	-	-	-	0,01	0,01	-	0,02	0,01	0,01	0,01	0,01
-	-	-	-	-	-	-	-	-	-	-	-	0,01	-	-	0,02	0,02	0,01	0,01	0,01
-	-	-	-	-	-	-	-	-	-	-	-	0,01	-	-	0,01	0,01	0,01	0,01	0,01
-	-	-	-	-	-	-	-	-	-	-	-	0,01	-	-	0,01	-	0,02	0,01	0,01
-	-	-	-	-	-	-	-	-	-	-	-	0,01	-	-	0,01	-	0,01	0,02	0,01
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,01	-	0,01	0,01	0,01
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,01	-	0,01	0,01	0,01
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,01	-	0,01	0,02	0,01
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,01	-	0,01	-	0,01
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,01	-	0,01	-	0,01
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,01
The number of possible states n on the step m																			
3	11	4	2	10	12	14	18	12	24	24	24	32	28	25	37	30	37	35	38

The selection of a certain solution should be based on path choice by the consideration of probabilities of their implementation (see Table 4 and Fig. 3). The probability of each solution achievement is defined by chain rule: $P(s_1, s_2, \dots, s_n) = \prod_{j=1}^{n} P(s_j|s_{j-1}, \dots, s_1)$ (for most probable sequence of states from Table 5 we receive the following probability values - 0,54; 0,2592; 0,1296; 0,0674; 0,0317; 0,0155; 0,0071; 0,0031; 0,0014; 0,0006; 0,0003; 0,00012; 0,00004; 0,000014; 0,000006; 0,0000022; 0,00000062; 0,00000018; 0,00000005; 0,00000001). Obtained values will be smaller than those obtained in Table 4. Small values show, that project implementation involves a set of factors that are not considered.

The table given above demonstrates that the data obtained by solution selection differ from those of economic planning. It is connected with the optimization processes and the matching of processes in each production system. At the same time, in optimal planning tasks we need to consider a minimum quantity of goods of each type to be produced; such tasks involve a group of other tasks to be tackled, i.e. management, social function, production peculiarities, break-even for each type of goods, etc.

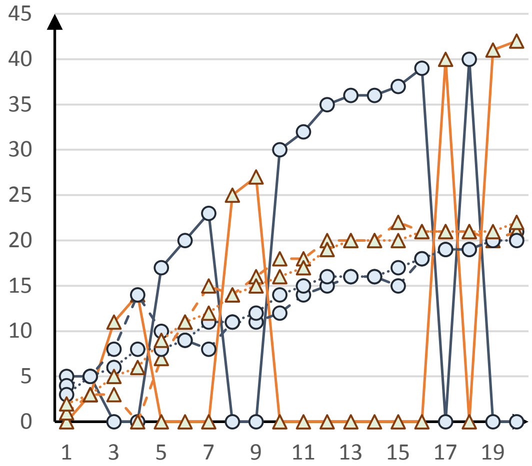


Figure 3: Alternative volume plans for two products (orange line – first product, blue line – second product; solid line - results of lot-scheduling planning; dotted line - results from business planning; dashed line - result from the pool of lot-scheduling planning results which near business planning result).

5 CONCLUSIONS

The application of forecast data in the tasks of optimal management and information systems, based on data and models, creates new possibilities to investigate processes that run in production systems. In particular, we can consider probabilistic nature of processes that run outside the examined production system, estimate the risks of planning and production activities, time factor [11].

Obtained solutions are approximate solutions; the task applies also planning data. Therefore, the task has statistical setting, but the model itself can superpose different formalization types that leads to the combination of different simulation types.

By the consideration of time factor, we examine system performance as a string of states. These states can take place and cannot take place (the system will take an alternative path then), that is why, calculation and evaluation of a state should be iterated on each time step.

Despite the possibility to perform multiple calculations, the decision is eventually made by a person; however, the processing of subjective data and their workable representation increases the objectivity of made decisions which is crucial in the situations when most promising markets are the markets of innovation goods; on such markets we cannot talk about the stability of production processes due to the short life cycle of such goods, big amount of modifications and components, increased power and resource intensity of goods.

Corresponding libraries and modules can be integrated into information systems of production and business process management, i.e.: ERP, MRP, CSRP.

Production systems where management tasks are solved are considered inertial objects of management. They cannot momentary adjust running processes; besides, the adjustment of technological processes requires the use of additional resources, i.e. time, financial resources, staff expertise, organization resources. The studies of such issues and the consideration of their impact on production system development paths can become the subject of further investigations; the data of the analytical agency Gartner verify the necessity of such research directions («Road map of production information systems» dating from March 28th, 2017).

ACKNOWLEDGMENTS

The author thanks the government of Perm Krai for the support of the project for "Development of software and economic and mathematical models for supporting innovation project management processes in production systems", which is being implemented in accordance with decree №166-п of 06.04.2011.

The reported study was partially supported by the Government of Perm Krai, research project No. C-26/058.

REFERENCES

[1] C. Kaiser, S. Schlick, and F. Bodendorf, "Warning system for online market research ? Identifying critical situations in online opinion formation," *Knowl.-Based Syst.*, vol. 24, no. 6, pp. 824–836, Aug. 2011.

[2] C. Arnold, D. Kiel, and K.-I. Voigt, "How the industrial internet of things changes business model in different manufactoring industries," *Int. J. Innov. Manag.*, vol. 20, no. 08, pp. 1640015-1-1640015–25, Dec. 2016.

[3] L. Mia and L. Winata, "Manufacturing strategy and organisational performance: The role of competition and MAS information," *J. Account. Organ. Change*, vol. 10, no. 1, pp. 83–115, Feb. 2014.

[4] L. Mylnikov, "Particularities of Solving the Problems of Support for Managerial Decision Making in Production and Economic Systems Using the Statistical Data," *Int. J. Econ. Financ. Issues*, vol. S8, no. 6, pp. 1–11, 2016.

[5] L. Mylnikov, "Risk Evaluation in Manufacturing Organization Tasks for Product Technological Projects and Establishment of Project Portfolio for Production Systems," in *Proceedings of the 2016 International Conference on Applied mathematics, simulation and modelling*, Beijing, China, 2016, pp. 399–402.

[6] M. Cheng, N. J. Mukherjee, and S. C. Sarin, "A review of lot streaming," *Int. J. Prod. Res.*, vol. 51, no. 23–24, pp. 7023–7046, Nov. 2013.

[7] H. Jalali and I. V. Nieuwenhuyse, "Simulation optimization in inventory replenishment: a classification," *IIE Trans.*, vol. 47, no. 11, pp. 1217–1235, Nov. 2015.

[8] F. T. S. Chan, N. Li, S. H. Chung, and M. Saadat, "Management of sustainable manufacturing systems-a review on mathematical problems," *Int. J. Prod. Res.*, vol. 55, no. 4, pp. 1210–1225, Feb. 2017.

[9] M. Díaz-Madroñero, J. Mula, and D. Peidro, "A review of discrete-time optimization models for tactical production planning," *Int. J. Prod. Res.*, vol. 52, no. 17, pp. 5171–5205, Sep. 2014.

[10] J. A. Anderson, *Discrete mathematics with combinatorics*. Upper Saddle River, N.J: Prentice Hall, 2001.

[11] L. Mylnikov and M. Kuetz, "The risk assessment method in prognostic models of production systems management with account of the time factor," *Eur. Res. Stud. J.*, vol. 20, no. 3, pp. 291–310, 2017.

Development of an Algorithm for Transition from an Electric-power System Visual Representation to a Parametric Representation

Anatolii Kusakin, Anton Petrochenkov, Dmitrii Leizgold and Ivan Luzyanin
[1]Electrotechnical Department, Perm National Research Polytechnic University
Komsomolsky Ave. 29, 614990, Perm, Russia
kav@msa.pstu.ru, pab@msa.pstu.ru, ldu@msa.pstu.ru, lis@msa.pstu.ru

Keywords: Electric-Power System, Oil Field, Object Model, Calculation Model, T-list, Algorithm, Equivalent Circuit, Branch, Node.

Abstract: Precise estimation of the oil field electric-power system (EPS) parameters is an important task when estimating energy efficiency of oil well equipment. One possible solution for this task is development a software module that is capable both to graphically represent grid topology and to calculate the parameters of EPS on the basis of electrical parameters of equipment. The currently used methods for EPS modeling are based on theoretical approaches that do not completely meet modern requirements on precision and convergence. In addition, the interaction of EPS elements is considered insufficiently in the existing models. Implementing Smart Grid systems require development of new modeling tools. To create them, it is necessary to describe two models and develop transition algorithm between the visual representation of the electric grid and the parametric model of the EPS which determines the main parameters of the system. The paper considers the object model that allows to construct virtual electric scheme using objects (elements) and the calculation model that provide electric-power system parameters on the basis of calculation Kirchhoff's circuit laws. The development of a transition algorithm between the object model of the EPS into a calculation model. To implement this algorithm the LabVIEW is used. Both models are included in a software package which are designed to simulate stationary and quasi-stationary operating regimes and calculate the EPS parameters for further operation analysis. Since the existing solutions do not provide an optimal models' transition technique, the authors propose the based on T-list algorithm. The research results will be useful for developing efficient EPS control techniques in various regimes and improve energy efficiency both on existing oil fields' EPS and new ones.

1 INTRODUCTION

Oil field electric-power system (EPS) is a distributed object with a complex structure. It is an emergent system that has a multilevel structure of rather large dimension with complex energy, technological, information interconnections [1].

In EPS there is actually no time period when the system works in stationary conditions [2][3].

The set of processes that occur in the EPS and determine its state at a specific point of time or within a certain time interval is called the EPS regime.

There are the following EPS regime types:

- Stationary regime when the EPS was designed and its main indicators were defined;
- Quasi-stationary regime when the system moves from one operating state to another;
- Emergency regime when the relay protection devices operation settings are determined, and the parameters change over allowable values;
- Post-emergency regimes when the possibility of further EPS operation with degraded performance indicators is revealed.

In this paper stationary regimes are considered.

The EPS regime has the following parameters: currents flowing through EPS elements and voltages at the clamps of the elements. The EPS regime parameters are interdependent to one another, they include the parameters of the grid elements with connection schemes being taken, power supply parameters and load characteristics. The EPS regime parameters can be determined by computing or experimental methods.

In case of EPS, the full-scale experiment is difficult to conduct due to safety requirements and technological process restrictions, therefore a mathematical model development is considered [4] [5].

Any EPS include 2 models: visual and calculated. The visual model is the EPS scheme, the calculated one is the system of equations that most fully describing the EPS structure and processes taking place in it. Often these 2 models are in no way connected with each other and fulfil different tasks. Once creating the relationship between them, a software package is created. It allows to practice methods of effective EPS management in various regimes, and to develop measures to increase the electrical equipment (EE) energy efficiency in operating existing oil fields and designing new ones.

The aim of the paper is to develop an algorithm for the transition from the visual EPS representation to the parametric one.

To achieve the goal, many tasks must be performed which related to:

- EPS representation in the form of an object model;
- EPS representation in the form of a calculation model which allows to calculate the main EPS regime parameters. The calculation is realized on the matrix-topological method basis [6];
- The development of an interface that provide a one data transition from the object model to the calculation model, and back another one.

2 SOFTWARE PACKAGE DESCRIPTION

The software package is designed to simulate stationary and quasi-stationary EPS operation regimes and calculate currents in branches and voltages in the nodes of the electric grid in all regimes.

The package under development consists of three interconnected elements (Fig. 1).

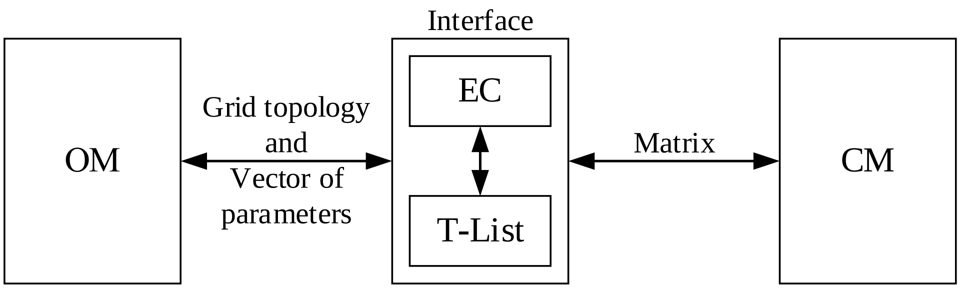


Figure 1: System elements interaction scheme. OM - Object Model; EC - Equivalent Circuit; CM - Calculation Model.

2.1 EPS Object Model Implementation in LabVIEW

Object Model (OM) is a virtual system similar to a real system and capable of reproducing properties and processes occurring in a real system with a certain accuracy. The OM should reflect those concepts and objects of the real world that are important for the developing system.

The structural unit of the OM is an object. A one model objects group is united by a hierarchy (typology). Each object includes a set of parameters that uniquely define this object and become sufficient for the calculations, a set of operations involving the object, as well as interactions with other objects.

The task of OM development is to provide a virtual representation of the real EPS in computer.

The created OM represents an objects tree in LabVIEW [7]. A small fragment is shown in Fig. 2.

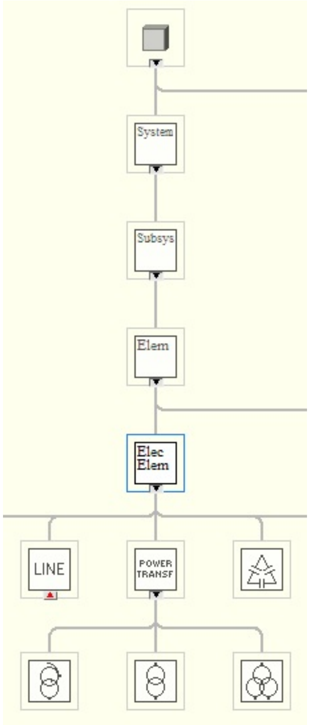


Figure 2: OM fragment in LabVIEW.

The figure shows the inheritance concept, according to which the descendant can inherit the ancestor parameters (properties) and the ancestor functions (methods) [8].

2.2 EPS Calculation Model

Calculation model (CM) is a matrix equations system [9] for nodes and branches; the model's input data are the objects parameters. This system is used to calculate the currents in the branches and voltages at the EPS nodes. The power consumed by the electric receivers is calculated and the losses in the system's elements are estimated in response to these parameters.

The process of creating a mathematical model is considered as a sequence of two stages: creating models of individual elements and creating a model for their interaction.

Approaches to the study of this type systems are subdivided into methods aimed at simplifying or reducing the dimensionality of the model in question and decomposition methods that enable one to solve a large high-order problem in which individual small problems are considered and the solutions obtained are combined [10].

The interaction of electrically connected elements is most naturally organized by means of electrical parameters such as currents and voltages. The calculations of the electric regimes are based on the solution of equations formulated on the basis of one of the two methods derived from Kirchhoff's circuit law: the method of the node voltage equations or the method of loop currents equations.

The nodal voltage matrix (NVM) is used to calculate the EPS regime parameters, the calculation is based on the formula:

$$\mathbf{Y}^{-1} \times \mathbf{J} = \mathbf{U} \tag{1}$$

where **Y** is the matrix of known nodal conductivities (**Y**-matrix); **U** is matrix-column of node voltages (**U**-matrix); **J** is matrix-column of the driving currents (**J**-matrix).

For the EC's *i*-th node, there is a *Usti* value (the reduced basic voltage of the transformation stage, kV) according to the formula:

$$U_{sti} = U_b \cdot \prod_{a=1}^{z} K_{T\mathbf{Z}a} \tag{2}$$

Where U_b is the basis voltage, kV; $K_{T\mathbf{Z}a}$ is a transformation coefficient of the *a*-th branch transformer (if there is no transformer in the branch, $K_{T\mathbf{Z}a} = 1$); *z* is the number of branches from the basic to the *i*-th node, equal to the number of **Z**-vector elements; **Z** is a vector column containing the ordinal branches numbers of the smallest path from the basic node to the node *i*.

The reduced base branch voltage is equal to the value of its starting node:

$$Usti_{SN} = U_{stj} \tag{3}$$

where $Usti_{SN}$ is the reduced basis voltage of the branch *j* starting node, kV; U_{stj} is the reduced base voltage of the branch *j*, kV.

Active and reactive conductivities are calculated and brought to the basis voltage through the known equivalent resistances and reactances of the composed circuit elements, then it forms the conductivity matrices (**G**- and **B**-matrices). After that, the conductivity matrices form an **Y**-matrix.

$$\begin{pmatrix} \begin{bmatrix} G_{11} & \cdots & G_{1k} \\ \cdots & \cdots & \cdots \\ G_{k1} & \cdots & G_{kk} \end{bmatrix} & \begin{bmatrix} B_{11} & \cdots & B_{1k} \\ \cdots & \cdots & \cdots \\ B_{k1} & \cdots & B_{kk} \end{bmatrix} \\ -\begin{bmatrix} B_{11} & \cdots & B_{1k} \\ \cdots & \cdots & \cdots \\ B_{k1} & \cdots & B_{kk} \end{bmatrix} & \begin{bmatrix} G_{11} & \cdots & G_{1k} \\ \cdots & \cdots & \cdots \\ G_{k1} & \cdots & G_{kk} \end{bmatrix} \end{pmatrix} \to \begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ Y_{m1} & Y_{m2} & \cdots & Y_{mm} \end{pmatrix} \tag{4}$$

where k is the number of the EPS EC nodes, m = 2·k.

The p-th element of the **J** column-vector is determined by the formula:

$$Jp = \begin{cases} \sum_{j=1}^{n} \left(\frac{E_{gj} \cdot R_j}{R_j^2 + X_j^2} \cdot \frac{U_{stj}}{U_b} \right), \text{while } p = 1..k \\ \sum_{j=1}^{n} \left(\left[-\frac{E_{gj} \cdot X_j}{R_j^2 + X_j^2} \right] \cdot \frac{U_{stj}}{U_b} \right), \text{while } p = k+1..m \end{cases} \quad (5)$$

Where E_{gj} is the value of Electromotive Force (EMF) of the branch i EC equivalent voltage source (VS), kV; R_j - equivalent resistance of the j-th branch EC, Ω; X_j - equivalent reactance of the j-th branch EC, Ω; TNj is the j-th branch's daughter node which is determined from the T-list.

The **U** column-vector is calculated by the formula:

$$\begin{pmatrix} U_1 \\ U_2 \\ ... \\ U_m \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} & ... & Y_{1m} \\ Y_{21} & Y_{22} & ... & Y_{2m} \\ ... & ... & ... & ... \\ Y_{m1} & Y_{m2} & ... & Y_{mm} \end{pmatrix}^{-1} \times \begin{pmatrix} J_1 \\ J_2 \\ ... \\ Jm \end{pmatrix} \quad (6)$$

The voltage drops of the **ΔU** column-vector are calculated from the formula, based on the known nodal voltages:

$$\Delta U_q = \begin{cases} U_{SNja} - U_{TNja} + E_{gja}, \text{while } q = 1..v \\ U_{SNjr} - U_{TNjr} + E_{gjr}, \text{while } q = v+1...w \end{cases} \quad (7)$$

where ΔU_q is the q-th element voltage drop (branch j) of the **ΔU** column-vector, kV; U_{SNja} - the active component of the j-branch starting node voltage, kV (U_{SNjr} - reactive); U_{TNja} - the active component of the j-branch daughter node voltage, kV (U_{TNjr} - reactive); E_{gja} - the active component of the j-branch's EMF (E_{gjr} - reactive); v is the number of the EPS EC branches, w = 2 · v.

According to the received voltage drops, the **I** column-vector currents are calculated from the formula:

$$Iq = \begin{cases} \frac{\Delta U_q \cdot R_j}{R_j^2 + X_j^2}, \text{while } q = 1..v \\ -\frac{\Delta U_q \cdot X_j}{R_j^2 + X_j^2}, \text{while } q = v+1..w \end{cases} \quad (8)$$

where Iq is the q-th element of the **I** column-vector (branch j), A.

The branches' currents and the nodes' voltages were calculated as reduced to the base voltage, so it is necessary to inverse transformation according to the formulas:

$$I_{aj} = \begin{cases} I_q \cdot \frac{U_b}{U_{stj}}, \text{while } q = 1..v \\ I_q \cdot \frac{U_b}{U_{stj}}, \text{while } q = v+1..w \end{cases};$$
$$U_{ai} = \begin{cases} Up \cdot \frac{U_{stj}}{U_b}, \text{while } q = 1..v \\ Up \cdot \frac{U_{stj}}{U_b}, \text{while } q = v+1..w \end{cases} \quad (9)$$

where I_{aj} is the value of the j-th branch's current, A; U_{ai} is the voltage value which belongs to the i-th node, kV.

2.3 OM-CM Transition Algorithm

The interface is used for the interaction of OM and CM. It performs the function of organization the T-list based on the EC and technical data of scheme elements. The T-list provides the initial data for the EPS regime parameters calculation.

The T-list formation algorithm can be presented in the form of a block diagram (Fig.3).

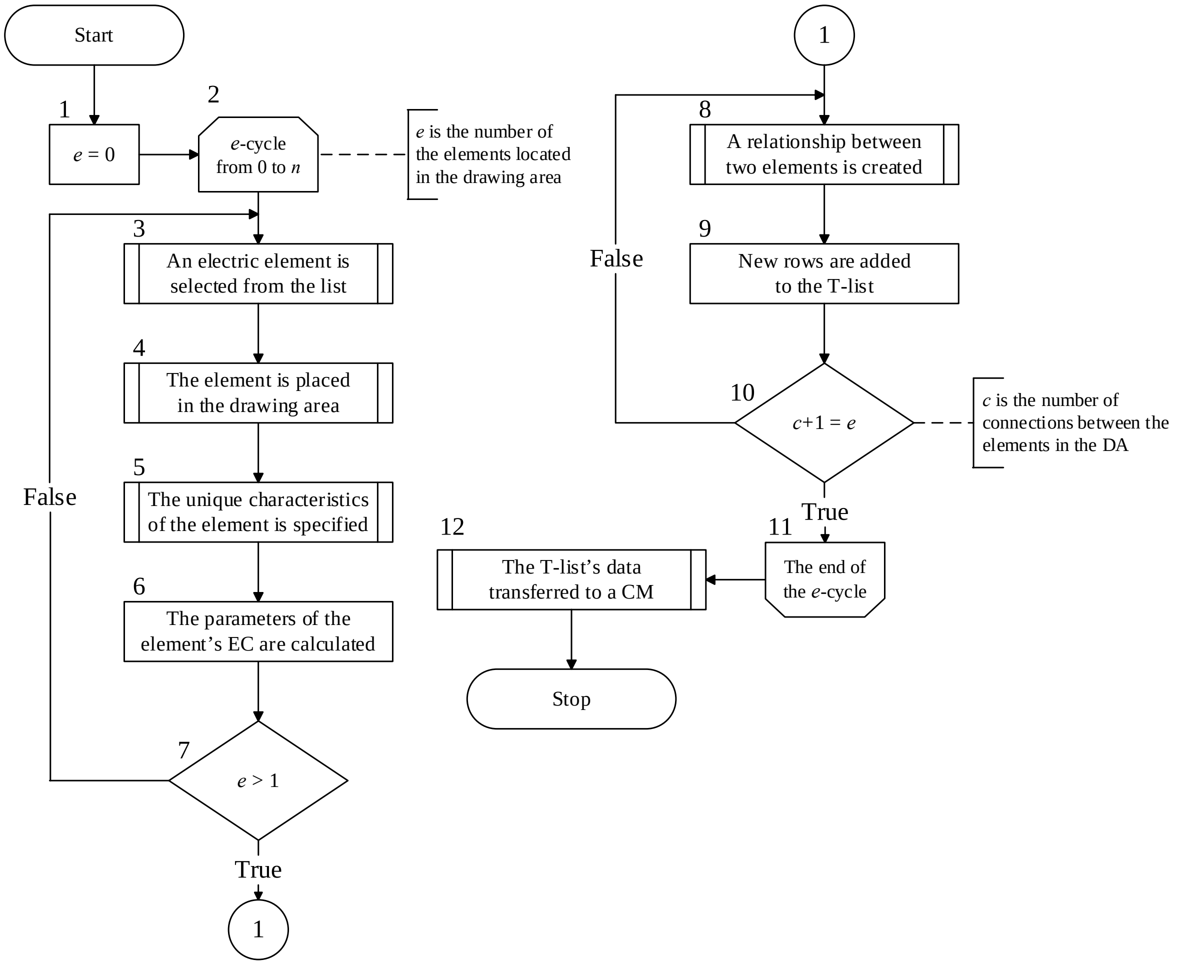


Figure 3: The interface algorithm in the form of a block diagram.

Block 1. At the first stage the number of elements (e) in the drawing area (DA) is zero. The drawing area is a field intended for a visualization of an electrical circuit and its modification by interaction with graphic primitives. An element is an object stored in the OM.

Block 2. A cycle adds new elements and a scheme formation is initiated (begin).

Block 3. The electric element is selected from the list. This list includes transformers, lines, generators, motors, switching devices and compensating devices [11].

Block 4. The selected element is placed in the DA.

Block 5. At this stage, the element parameters are entered and recorded into a cluster of private data. The EE parameters and the input power (IP) voltage are defined. Then the transformation coefficients of all the system transformers are determined, and the IP voltage is automatically taken as the basis voltage.

Block 6. Equivalent resistances and reactances, the EMF of the equivalent source and the current of each EC branch element are calculated according to the given technical data. The EC is a model of the real circuit, which has at least two nodes and one branch. Each branch connects two nodes, each node is able to contain two or more branches. Each branch is characterized by an equivalent impedance, an equivalent voltage source and an equivalent current source. The impedance is calculated by formulas with the equipment's technical data. The EC is stored in the cluster of element's private data in the form of an array and its filled when specifying the element technical data. The EC of each element has input and output nodes. The exception is IP (it has only an output node) and electric consumer (it has only an input node). The number of element EC branches depends on the type of this element:

1) If the IP's EC is considered (Fig.4), then the number of branches is 1;

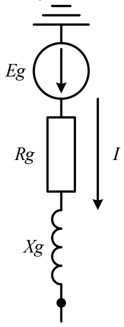


Figure 4: Input power equivalent circuit.

2) For the EC of all other elements, the number of branches is 2, for example, overhead transmission line (OL) EC (Fig.5);

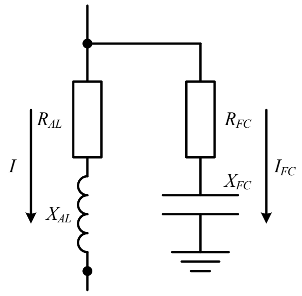


Figure 5: OL equivalent circuit.

Block 7. The condition of existence in the DA at least two elements are checked.

Block 8. A link between two electrical elements is created in the DA. To create a link, the output node of one element (the parent element) is connected to the input node of the second element (the child element).

Block 9. The elements EC arrays and the string, corresponding to the connection between them, are entered in the topological list (T-list). The T-list is a table with a fixed number of columns and a variable number of rows (Table 1). T-list is used for recording, storing and transferring the system EC characteristics to the CM. An alternative to the T-list is the incidence matrix, however, in the framework of this study it is not used because it has a weak fullness. This leads to an excessive expenditure of computing time and the computer's storage space.

The main advantage of the T-list is the free order of the lines numbering, i.e. entering the information in any order into the T-list is possible.

Interaction with the T-list is realized as follows:

1. Once a new element or/and a new link between elements (branch) is added to the DA, new rows with information about the element are added to the T-list. To write a new electrical element, one or two rows are allocated to the T-list, depending on the element's EC (Fig. 4, 5).

2. Each branch and each node of the EC acquires a unique identifier (ID) in the T-list.

There are three types of nodes are distinguished:

1) Generator node - no superior node is defined for it;

2) Load node - it is not superior to any other node;

3) Intermediate node - it has superior nodes and is a superior node to other nodes;

3. The rows are removed from the T-list when corresponding branch or element is excluded;

4. The element parameters are recalculated when the technical data of electrical elements is changed, or/and the branches higher and lower nodes are overridden when system is reconfigured.

Table 1: T-list example.

Branch ID	*Starting Node ID*	*Daughter Node ID*	*R*, Ω	*X*, Ω	*E*, kV	*Transformation coefficient*	*J*, A
1	1	2	5	3	0	1	0
…	…	…	…	…	…	…	…
n	b	c	d	e	f	g	h

The T-list includes the following positions:

1) The first column contains the unique identifier of the graph's branch (*Branch ID*);

2) The second column contains the unique identifier of the graph's branch starting node (*Starting Node ID*);

3) The third column contains the unique identifier of the graph's branch daughter node (*Daughter Node ID*);

4) The fourth column contains the value of the branch's equivalent resistance (R);

5) The fifth column contains the value of the branch's equivalent reactance (X);

6) The sixth column contains the branch's equivalent IP EMF value (E);

7) The seventh column contains the magnitude of the branch equivalent driving current (J).

Block 10. The condition of all elements connection that placed in the DA is checked.

Block 11. A cycle adds new elements and a scheme formation is initiated (end).

Block 12. The matrix of nodal conductivities (4) is organized, the base voltage and transformation coefficients of the transformers (2) are formed, as well as the equivalent IP EMF values of each EC branch for the driving currents calculation (5) are formed based on the information contained in the T-list.

3 CONCLUSIONS

The paper discusses the question of the relationship between the EPS graphical model and the EPS parametric model. The existing methods of transition from one model to another are not optimal. An algorithm for the OM-to-CM transition through the T-list was developed in order to decrease the computation time and reduce the computer's storage space. It was necessary to create an EPS object model in LabVIEW, create a calculation model in LabVIEW, and create an interface that allows for transferring one data from the object model to the calculation model, and the other data back, to implement the algorithm. The key element of the interface is the T-list, which is convenient for storing and transmitting the grid topology, and it also has a number of advantages over other frequently used structures, for example, the incidence matrix. This approach is consistent with the work of other authors who use the matrix-topological approach and graph theory in the EPS calculation [10]. Unlike known methods, the method of reduction to a single baseline level is used [2]. Based on the calculation of stationary and quasi-stationary EPS regimes, it will be possible to proceed to the calculation of the dynamic EPS regimes.

REFERENCES

[1] L. Cochrane, “Smart Energy Reference Architecture,” Microsoft Worldwide Utilities Industry, March 2013.

[2] A. Petrochenkov, I. Luziyanin, “Practical Aspects of Software Developing for the System of Structural and Functional Analysis of Power Supply Systems in Oil Companies,” Proceedings of International Conference on Applied Innovation in IT, vol. 1, pp. 65-69, March 2017.

[3] A. Khabdullin, Z. Khabdullina, G. Khabdullina, S. Tsyruk, “Development of a software package for optimizing the power supply system in order to minimize power and load losses,” Energy Procedia, v. 128, pp. 248-254, September 2017.

[4] E. Kremers, “Modelling and Simulation of Electrical Energy Systems through a Complex Systems Approach using Agent-Based Models,” KIT Scientific Publishing, 2013.

[5] A. Herbst, F. Toro, F. Reitze, E. Jochem, “Introduction to Energy Systems Modeling,” Swiss Society of Economics and Statistics; vol. 148 (2), pp.111-135, 2012.

[6] A. B. Petrochenkov, “An Energy-Information Model of Industrial Electrotechnical Complexes,” Russian Electrical Engineering, vol. 85, No. 11, pp. 692–696, 2014.

[7] N. A. Wahab, M. S. Gaya, Y. Md Sam, U. Jeppsson, R. Katebi, “A LabVIEW-Based Simulator for the Activated Sludge Process,” IFAC Proceedings Volumes, vol. 45, 2012, pp. 942-947, 2012.

[8] J. Kl´ımek, M. Necaskˇy., “On Inheritance in Conceptual Modeling for XML,” Procedia Computer Science, vol. 10, pp. 54–61, 2012.

[9] L. Liu, X. Zhao, “Application of Improved Adjacency Matrix Multiplication in Distribution Network Flow Calculation,” Energy Procedia, vol. 14, pp. 983-989, 2012.

[10] B. V. Kavalerov, A. B. Petrochenkov, K. A. Odin, V. A. Tarasov, “Modeling of the Interaction of Structural Elements of ElectricPower Systems,” Russian Electrical Engineering, vol. 84, pp. 9–13, January 2013.

[11] M. Khakimyanov, I. Shafikov, F. Khusainov, “Electric Submersible Pumps in Oil Production and Their Efficiency Analysis,” Proceedings of International Conference on Applied Innovation in IT, vol. 1, pp. 35-38, March 2016.

Direct Machine Translation and Formalization Issues of Language Structures and Their Matches by Automated Machine Translation for the Russian-English Language Pair

Anna Novikova
Perm National Research Polytechnic University, Komsomolsky Ave. 29, Perm, Russia
novikova@yandex.ru

Keywords: Machine Translation, Semantic Match, Structural Match, Formalization Of Language Structures, Formalization Language, Distributive Principles, Surface Language Structures, Syntactic Positions, Skeleton Language Structure, In-Depth Model, Pre-Processing Of Text, Referential Text Corpus.

Abstract: The present paper introduces a formalization language for sentences and text corpora that helps tackle the acute problem of formalizing semantic and structural matches of different language systems by direct machine translation. In the paper, a detailed look is taken at the elements of reference and situation-based analysis of the representations of surface and in-depth semantics of semantic contexts in the sphere of business communication relying on the meaning-text theory for automated formalization of language structures and their matches by machine translation. The study is aimed at working out an algorithm that will enhance the quality of machine translation excluding intermediary natural languages and post-editing of translated texts. A distinguishing feature of the suggested approach is structurization and formalization of language structures on the stage of text pre-processing. Such «input filter» of information will enable decoding system create literate messages in compliance with lexical and grammatical distributive principles both in a foreign language and native language, and use them for proofreading of texts in a native language and building structures that can be further translated by existing machine translation systems with high quality.

1 INTRODUCTION

Mankind's history cannot be imagined without intercultural communication that is increasing over decades. Strong evidence of this process is the penetration of a vast number of adopted words in different languages and the appearance of well-known global terms and expressions. Despite that, the issues of automated translation were firstly addressed only in the 20-s of the XX century. The pioneers of this movement are considered to be Estonian (A. Vakher), American (W. Weaver, H.P. Edmundson, P.G. Hays), French (G. Artsrouni), and Russian (P.P. Smirnov-Troyanskii) scientists [1]. The introduction of computing machines contributed to further development of this scientific movement. In the middle of the XX century, the achievements of Warren Weaver of the Rockefeller Foundation and RAND in computerized translation became world-renowned [1]; those forefront achievements laid down a fundamental translation principle when human translation was replaced by computer as a mediator. Since that time, machine translation has distinguished into a separate science intensive direction.

However, despite the rapid development of information technologies, the appearance of artificial intelligence, machine learning, the development of mathematical statistics and the enhancement of computer technics there is still no solution for the problem of machine translation that is able to fully replace a human mediator [2].

Modern implementations in the considered area help solve the problem of translating actual content of simple sentences from 8 up to 30 words, and adjust classifiers for selecting term matches in certain subject areas [3].

Today we have free online solutions from Microsoft, Google, Yandex and Baidu that provide users with medium-quality translation, that in most cases can be post-edited by an attracted expert. Current situation creates high competitiveness on the market of machine translation, where the main efficiency indicator of machine translation systems adds up to the solution of the problem of semantic and structural matches in different language systems.

2 LANGUAGE STRUCTURES AND APPROACHES FOR THEIR FORMALIZATION

In modern machine translation systems distorted meanings are placed, as a rule, beyond the bounds of the 8th word, when the quality of translation deteriorates due to search errors [3][4][5]. Firstly, this trend can be explained by different syntactic structures of languages – for instance, rigid linear order in English language and frame structure of tenses in German language, - and, secondly, by an increasing word distribution distance. In 2012, the corporation Google made attempts to tackle the problem of language structure matching by integrating a method of cross-lingual word clusters, that performs a direct transfer of kindred language structures; according to the developers, this method helped slightly increase the quality of translation by 26% [6] (see also the Russian system «Crosslator 2.0», that uses a mediator language in order to achieve multilingual translation of terms).

The existing phrase methods of statistical text processing «see» only the local distribution environment and do not consider the lexical-grammatical context of the whole phrase. That is why, scientists continue work on the models and methods that will expand the diapason of hypothetic prediction of machine translation system without the distortion of meanings: for example, an enhanced method for the prediction and estimate of future actions of machine translation system based on algorithm modifications applied in Pharaoh [7], the method of dynamic selection by Arianna Bisazza and Federico Marcello, when the reordering of data is not restricted by decoding by the closest lexical environment (like in the Moses system) but extends to a long distance and, hence, expands the process of hypothetic decision making by translation up to the 15th word [8] (refer also to [9], [10], [11]). One of new approaches for solving the problem of ambiguity by translation is the architecture Transformer, announced by the working group of Google on August 31st, 2017. Underlying this architecture is a mechanism of self-attention, that uses deep neural learning mechanism, lines up and analyzes the interconnection points of all word representations in the context of a sentence. According to the BLEU scores (see [12] and [13]), such decoding algorithm improves the quality of translation by overcoming the difficulty of choosing the single accurate meaning representation of an ambiguous word for a certain semantic context with help of weighted values.

In the paper [14] the problem of machine translation was examined from the standpoint of modelling a speaker's linguistic competence, and suggested a complex functional method for business text translation based on the analysis of semantic features and basic frames. By the approbation of the suggested method for the machine translation from Russian into English the level of semantic match increased in comparison with the translations of the existing machine translation systems. However, this method requires the formalization of language structures (the description of matches among the structures of different languages). The drawback of this methods consists in the necessity to build a separate data base with language structures and their matches for different language pairs. Labor-intensive character of this method can be eliminated by the automation of language structure formalization. To achieve that, we need to classify the types of written forms of meaning representations and formalize a way to present them taking into account the skeleton-semantic and context-dependent components.

Underlying the analysis of language structures is the meaning-text theory [14], that is structurally implemented on the example of the analysis and classification of business communication speech acts. Reference- and situation-based method includes the analysis of surface and in-depth semantic representations (read more about case grammar by Ch. Fillmore in [15]), and namely: the analysis of 1) skeleton-language structures that actualize this or that speech act of business communication 2) valence «core» components, that guide the vector of thought unfolding, and 3) the components of semantic variability to have a possibility to generate *n*-phrases, not depending on the subject content of a phrase. From the standpoint of surface semantics, the analysis of speech acts helps obtain a limited set of surface language structures utilized by speakers in the sphere of business communication in order to describe a certain scenario (for example, invitation, appeal, recommendation, etc.) as well as design the variability of a semantic context-dependent description of this scenario in form of situations and, finally, describe the possibilities of stylistic tint for the actualized situations.

To test the hypothesis about the possibility to model semantic contexts and formalize language structures we carried out an experiment on the basis of over 500 speech messages in business communication (personal correspondence, analysis of scripts of English-speaking movies and series and their Russian-speaking translated versions, tutorials); the experiment results helped single out 3 big groups of speech acts, i.e. speech acts with pragmatic description, speech acts with emotional (perlocutionary) description, and speech acts with structural-logical description. Each group was further classified from the standpoint of the actualized scenarios. The analysis included the method of structural analysis, skeleton-semantic and context-dependent components of business communication messages.

The structural analysis revealed a group of most frequent scenarios used to implement speech acts with pragmatic description, and namely: appeal, prohibition, request, recommendation, suggestion (offer), invitation, reminding, conviction (persuasion), gratitude, congratulation, hope, wish, apologies).

At the same time, the speech act of a suggestion subdivides into the following types of situations: suggesting an alternative solution, suggesting a way to solve a problem/ problematic situation, suggestion with identifying a desired outcome; and the speech act of congratulations implements the following situations: congratulations with holidays/seasons, congratulations on receiving an award.

The analysis of skeleton-semantic and context-dependent (variable) content of speech acts with pragmatic description helps formalize lexical-grammatical distributive environment of phrases and determine their components.

Let us consider an example of the Russian-English match for the English phrase «**Please** **go to the following link** **in order to** print your card» and the Russian phrase «**Для** печати вашей карточки, **пройдите, пожалуйста, по следующей ссылке»**. The structural analysis of language structures used to implement the meaning of an appeal and a speech act of an appeal consequently, is presented in the Table 1.

An important observation is the difference in the in-depth syntactic positions in the phrase segmentation in the source and target languages, see Table 2. In this case it is necessary to pay attention to a «free» theme - rheme based character of thought unfolding in certain languages (for example, Russian).

Table 1: Structural analysis of language structures.

Components of structural analysis	**English model**	**Russian model**
Skeleton language structure	Please do something in order to do something.	Для чего-то сделайте, пожалуйста, что-то
Valence core component	go to the following link	пройдите по следующей ссылке
Semantic variability component	print your card	печати вашей карточки

Table 2: The matching analysis of in-depth syntactic positions in a phrase segmentation.

<table>
<tr><td colspan="4">Model of English sentence</td></tr>
<tr><td>1</td><td colspan="2">2</td><td>3</td></tr>
<tr><td>Please</td><td colspan="2">go to the following link</td><td>in order to print your card.</td></tr>
<tr><td colspan="4">Model of Russian sentence</td></tr>
<tr><td>1=3</td><td>2=2</td><td>3=1</td><td>4=2</td></tr>
<tr><td>Для печати вашей карточки</td><td>пройдите</td><td>пожалуйста</td><td>по следующей ссылке.</td></tr>
</table>

In such cases, independent of the order of the train of thoughts in a Russian language the process of sentence decoding should analyze the in-depth semantics of word representations and identify a match for in-depth syntactic positions taking into account the grammar system of the target language (for example, English).

By the consideration of certain stylistic genres of communication, it is possible to narrow a number of linguistic types and structures that actualize emotional tints. Due to a rigid and laconic character of lexical-grammatical standards of language structures that have historically put together for the stylistic code of business communication, the category of emotionality is poorly represented in this stylistic code and is conventionally represented by 4 basic speech acts, i.e. joy, sadness, bewilderment (feeling confused), respect.

The analysis of speech acts from the standpoint of structural-logical information resulted in 7 basic speech acts: 1) initiating a message, 2) action's plan, 3) attached information, 4) appearance of questions, 5) comments, 6) finalizing, 7) finishing a message. The speech act that describes the beginning of a message characterizes the following sub-situations: writing for the first time, resuming a communication after a break in correspondence, a recurrent reply (ongoing correspondence), a recurrent reply with giving the prehistory of previous replies, a recurrent reply with identifying the goal of an email.

The speech act that realizes action's plan divides into the following sub-situations: 1) a series/ consequence of actions, 2) the current state of business, 3) immediate actions. The speech act that indicates the presence of attached information/materials in an email can actualize 1) a message about the attached information, 2) a message about additional email recipients, 3) a message about forthcoming (consecutive) information, 4) a message that states where additional information can be found. The finalizing speech act can actualize 1) positive attitude to the forthcoming situation, 2) negative attitude to the forthcoming situation, 3) final notes. The speech act that describes the situations with commenting actualizes the following situations: 1) confirmation, 2) clarifying, 3) consent (agreement)/ non-agreement, 4) an intention to understand the reason of a current state, 5) causes and effects, 6) difficulties. In the speech act of questions reasons are described, i.e. if 1) the solution fits, 2) the situation is understood correctly; 3) what is the reason of the current situation; 4) how to solve this or that issue.

Multivariable realizations of speech acts with logical-structural description indicate a relevant character of this subgroup of speech acts for business communication.

Therefore, the presence of a direct match in surface semantics and in-depth syntactic positions in the segmentation of a phrase is evidence of the possibility to formalize the matches of language structure without the introduction of additional conditions.

The determination of skeleton-semantic and context-dependent variable semantic content of business communication messages with stylistic tint description (polite, neutral and informal (rough)) helps classify language expressions and work out a language for describing matches among the structures of the source language and the target language based on the functional method for automated text translation [14].

3 THE FORMALIZATION AND USE OF LANGUAGE STRUCTURES

The following symbols are used for the description of language structures:

() – sentence type definition;
= - clarifying the concept;
{} – merge;
<> - mandatory part of a sentence;
[] – optional part of a sentence;
| - or;
"" – rigidly given context;
\ - clarifying a new variable.

Text description consists then of two stages: sentence type definition and the description of sentence structure.

To describe the types of sentences we use the following classifiers and labels:

1. Classifier of sentence purpose: declarative (1A), interrogatory (1B), imperative (1C).
2. Classifier of mood: indicative (2A), conditional (2B), imperative (2C).
3. Classifier of voice: active (3A), passive (3B).
4 Classifier of tenses: past simple tense (4A), present simple tense (4B), present in process (4C), past in process (4D), present with result (4E), past with result (4F), present with result and process (4G), past with result and process (4I), future simple (4J), future in the past (4H), future in the past with result (4K), near future (4L), near future in the past (4M), immediate future (4N), immediate future in the past (4P).
5. Classifier of the connection between the object of speech and the fact that is reported about the object: assertion (5A), negation (5B).
6. Classifier of predicate type: simple predicate (expressed by notional verb or auxiliary verb) (6A), predicate with modal tint (modal verbs) (6B), predicate with aspect tint (aspect verbs) (6C), predicate with modal and aspect tint (modal verbs + aspect verbs) (6D), predicate with a notional estimate (verbs with meanings of supposition, suggestion, desire, recommendation, advice) (6E),

predicate with modal and notional estimate (modal verbs + a verb with the meaning of supposition, suggestion, desire, recommendation, advice) (6F).

7. Additional types of classifiers can be introduced depending on the language system (for instance, a classifier of words that can change the reordering of a phrase can be set for English language).

The description of sentence structures requires the introduction of such basic variables like: noun phrase (subject) (NP1), object (complement) (NP2), auxiliary verb of a notional verb (always conjugated) (Aux), Gerund (Gerund), infinitive of a notional verb (V), object (complement) (NP3), adjective (Adj), determinative (Det), definite article (Art/def), indefinite article (Art/indef), etc.

These variables can be used to form a new variable or adjust their meanings (see Listing 1), for the specification of sentence types see Listing 2.

Listing 1: An example of describing new variables via basic variables.

```
\AdvP1 = [AdvP1.2][AdvP1.1x]
\NP2v={<Det><N>}|{<Prep><Det><N>}|{<Pro
n/pers>}|{<Adv><Adj>}
\NP2v1g={<Adj>}|{<Det><N>}|{<Prep><Det>
<N>}|{<Adv><Adj>}|{<Pron/pos>}
\AdvP2 =
{"slowly"}|{"quickly"}|{"rapidly"}
\AdvP3 =
{"here"}|{"there"}|{"outside"}|{"inside
"}|{"somewhere"}|{"everywhere"}
```

Listing 2: An example of describing general cross-functional sentence structures.

```
(1A,2A,3A,4B,5A,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1][AdvP2][AdvP3]<".">
(1A,2A,3A,4B,5B,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1][AdvP2][AdvP3]<".">
(1B,2A,3A,4B,5B,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1][AdvP2][AdvP3]<"?">
(1A,2A,3A,4C,5A,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1]<".">
(1A,2A,3A,4C,5B,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1]<".">
(1B,2A,3A,4C,5A,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1]<"?">
(1B,2A,3A,4C,5B,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1]<"?">
(1A,2A,3A,4E,5A,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1]<".">
(1A,2A,3A,4E,5B,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1]<".">
(1B,2A,3A,4E,5A,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1]<"?">
(1B,2A,3A,4E,5B,6A)=<NP1><VP><NP2>[NP2e
xt][NP3][AdvP1]<"?">
```

The introduction of formal description helps utilize these structures to find matches among languages (use the formalization language as an intermediate language), and develop constructors for building phrases that due to the restrictions, artificially installed by developers into phrase structures, will serve as an input filter and instrument to prepare phrases for translation (see Fig. 1).

The suggested description helps structure any types of sentences. For instance, for English language see the examples in the Listing 3, and the corresponding examples in Russian language are presented in the Listing 4.

Listing 3: An example of structuring sentences for English language.

```
(1A,2A,3A,4B,5A,6A,7A)                =
{[NP3]<’’,’’><NP1>[AdvP1.2]<VPvf1><NP2>
[NP2ext][AdvP2] <".">}
(1B,2A,3A,4E,5A,6A,7C)                =
<VPAuxE><NP1>[AdvP1.3/4]<VP3><NP2>[Np2e
xt][AdvP1.3/2*]<’’?’’>
```

Listing 4: An example of structuring sentences for Russian language.

```
(1A,2A,3A,4B,5A,6A,7A)                =
{<NP1>[NP3][AdvP1.2][AdvP2]<VPvf1><NP2>
[NP2ext] <".">}|
{[NP3]<NP1>[AdvP1.2]<VPvf1><NP2>[NP2ext
][AdvP2] <".">}|
{[NP3]<NP1>[AdvP1.2][AdvP2]<VPvf1>|<VPv
f1.2><NP2>[NP2ext] <".">}|
{<NP1>[AdvP1.2][AdvP2]<VPvf1>[NP3]<NP2>
[NP2ext] <".">
(1B,2A,3A,4E,5A,6A,7C)      =      {<NP1>
AdvP1.3/2*][AdvP1.3/4]<VP3><NP2>
[Np2ext]<’’?’’>}|{[AdvP1.3/2*]<NP1>[Adv
P1.3/4]<VP3><NP2>[Np2ext]<’’?’’>}
```

The utilization of structures in the source language helps increase the quality of translation by excluding idiomatic expressions, expletive words from the texts, etc. The use of matches to reorder a phrase before using machine translators in many cases helps eliminate errors and distorted meanings (see Table 3) taking advantage of already existing solutions.

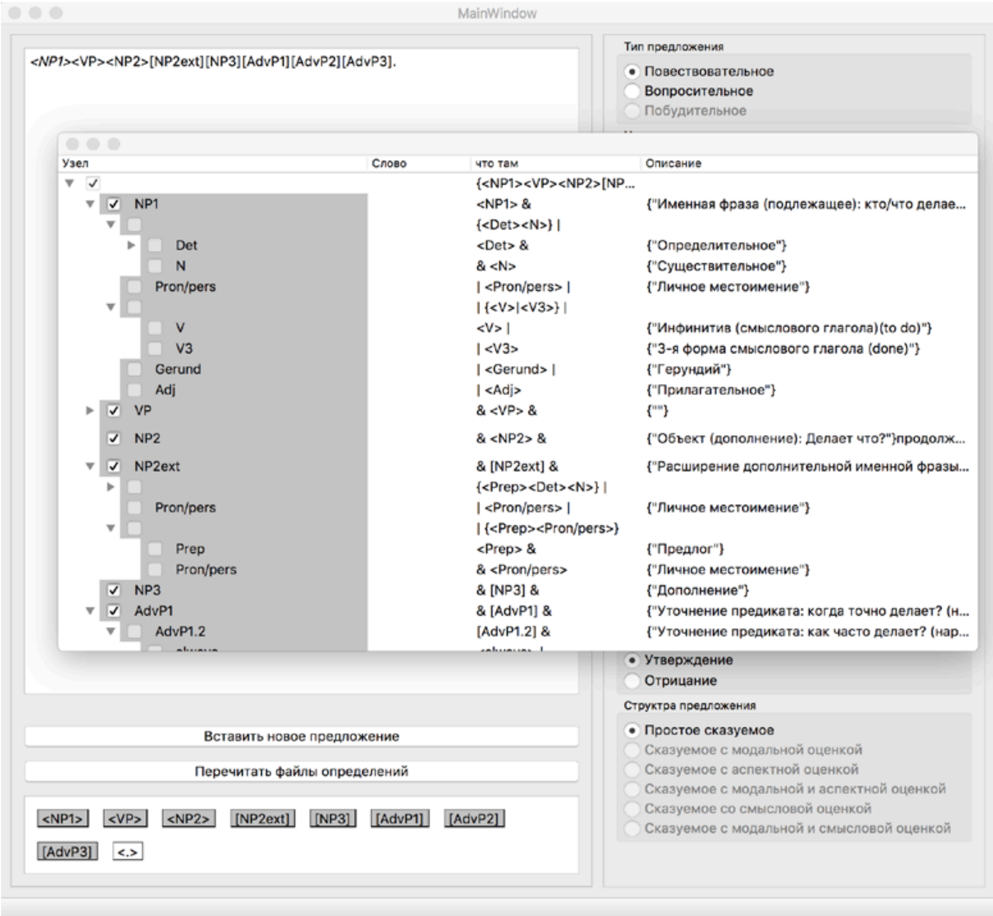


Figure 1: Interface example of phrase construction system.

Table 3: The examples of translation from Russian into English with help of the system Google translate without modifying a sentence structure in an initial sentence and with the modification of sentence structures.

	Sentence structure	**Translation**	**Amount of errors**
Example 1 – «Они в компании всегда быстро проводят обновление программного обеспечения»			
Initial sentence	<NP1>[NP3][AdvP1.2][AdvP2] <VPvf1> <NP2>[NP2ext] <".">	They always update the software in the company.	3
Modified sentence	[NP3]<’’,’’><NP1>[AdvP1.2] <VPvf1><NP2>[NP2ext][AdvP2] <".">	In the company, they always carry out software updates quickly.	0
Example 2 – «Раньше ваша компания когда-либо обновляла программное обеспечение для переводчика?»			
Initial sentence	[AdvP1.3/2*]<NP1>[AdvP1.3/4] <VP3> <NP2>[Np2ext]<’’?’’>	Did your company ever update the software for an interpreter?	3
Modified sentence	<NP1>[AdvP1.3/4]<VP3><NP2> [Np2ext][AdvP1.3/2*]<’’?’’>	Has your company ever updated the software for an interpreter before?	0

Besides, unlike the existing solutions the suggested approach is neither aimed at correcting the grammar which leads to distorted meanings by grammatically correct structures nor requires post-editing of received results.

4 CONCLUSIONS

The use of the described approach helps exclude intermediary natural languages by machine translation (it is known, that by the translation to/from Hebrew German is used as an intermediary language) having replaced them by an artificial language that helps 1. add restrictions into the lexical diversity (if necessary) 2. perform pre-editing of texts for translation having replaced by this operation the attempts to correct errors of automated translation systems (this approach is more attractive as it is easier for the user to make required adjustments in a native source language rather than changing the target language) 3. carry out automated proofreading of texts for their preparation quality. The main method of this approach is based on structural analysis and formalization of texts. At the same time, this approach is not sensitive to the subject area and the prehistory of use.

Further developments can be connected with such directions as language learning, proofreading of the source language, automated translation, text constructor for building phrases in a foreign language, that are implemented in accordance with the scheme presented in the Fig. 2.

Despite all the evident advantages, there are also difficulties connected with the necessity to describe each natural language separately with help of the developed language, which requires high expertise.

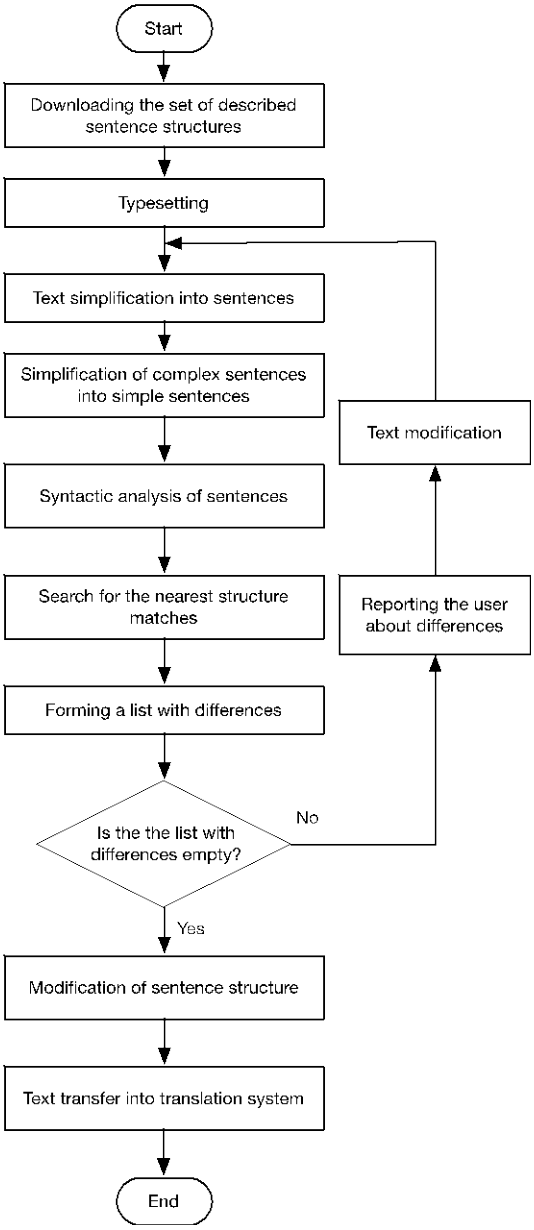


Figure 2: The algorithm of using the description language for language structures.

REFERENCES

[1] B. Henisz-Dostert, R. R. Macdonald, and M. Zarechnak, Machine translation. The Hague ; New York: Mouton, 1979.

[2] P. Sojka, Ed., Text, speech and dialogue: 13th international conference, TSD 2010, Brno, Czech Republic, September 6-10, 2010: proceedings. Berlin; New York: Springer, 2010.

[3] M. Aiken, K. Ghosh, J. Wee, and M. Vanjani, “An Evaluation of the Accuracy of Online Translation Systems,” Commun. IIMA, vol. 09, no. 04, 2009.

[4] A. Kazemi, A. Toral, A. Way, A. Monadjemi, and M. Nematbakhsh, “Syntax- and semantic-based reordering in hierarchical phrase-based statistical machine translation,” Expert Syst. Appl., vol. 84, pp. 186–199, Oct. 2017.

[5] J. B. Mariño et al., “N -gram-based Machine Translation,” Comput. Linguist., vol. 32, no. 4, pp. 527–549, Dec. 2006.

[6] O. Taeckstroem, R. McDonald, and J. Uszkoreit, Proceedings of the 2012 Conference of the North American Chapter Of. Saintoudsburg: Association for Computational Linguistics, 2012.

[7] R. Moore and C. Quirk, “Faster Beam Search Decoding for Phrasal Statistical Machine Translation.,” Proc. 11th Mach. Transl. Summit Cph., pp. 321–327, 2007.

[8] A. Bisazza and F. Marcello, “Efficient Solutions for Word Reordering in German-English Phrase-Based Statistical Machine Translation.,” Assoc. Comput. Linguist., pp. 440–451, 2013.

[9] N. Durrani, H. Schmid, A. Fraser, P. Koehn, and H. Schütze, “The Operation Sequence Model—Combining N-Gram-Based and Phrase-Based Statistical Machine Translation,” Comput. Linguist., vol. 41, no. 2, pp. 185–214, Jun. 2015.

[10] E. Hasler, A. de Gispert, F. Stahlberg, A. Waite, and B. Byrne, “Source sentence simplification for statistical machine translation,” Comput. Speech Lang., vol. 45, pp. 221–235, Sep. 2017.

[11] F. J. Och and H. Ney, “Discriminative training and maximum entropy models for statistical ma- chine translation.,” Proc. 40th Annu. Meet. Assoc. Comput. Lin- Guistics ACL, pp. 295–302, 2002.

[12] Association for Computational Linguistics, P. Isabelle, and Association for Computational Linguistics, Eds., Proceedings of the conference, 40th annual meeting of the Association for Computational Linguistics: Philadelphia, [6 - 13] July 2002, University of Pennsylvania, Philadelphia, Pennsylvania. Hauptbd. ... San Francisco: Morgan Kaufmann, 2002.

[13] P. N. Astya et al., Proceeding, International Conference on Computing, Communication and Automation (ICCCA 2016): 29-30 April, 2016. 2016.

[14] A. V. Novikova and L. A. Mylnikov, “Problems of machine translation of business texts from Russian into English,” Autom. Doc. Math. Linguist., vol. 51, no. 3, pp. 159–169, Jun. 2017.

[15] S. Mazarweh, Fillmore Case Grammar Introduction to the Theory. München: GRIN Verlag GmbH, 2010.

Automation of the Process of Regional Development Management Based on the Use of Smart Benchmarking Technology

Julia Dubrovskaya[1], Elena Kozonogova[1], Igor Nosov[2] and Yuliya Karpovich[1]
[1]*Economics and Finances Department, Perm National Research Polytechnic University, 29 Komsomolsky prospekt, Perm, Russia*
[2]*Department of Applied Mathematics and Informatics, Perm National Research Polytechnic University, 29 Komsomolsky prospekt, Perm, Russia*
uliadubrov@mail.ru, elenaa.semenovaa@gmail.com, dobby1706@gmail.com, karpushki@mail.ru

Keywords: Process Automation, Regional Development, Smart Benchmarking, Distance Matrix.

Abstract: The process automation is one of the effective approaches to process management based on the use of information technology at the present stage of the society development. This approach allows to increase productivity, reduce the process execution time, reduce costs, increase the accuracy and stability of operations. The automation of processes has currently covered both industries and services fields of economy. This article presents an algorithm for automating the procedure for regional benchmarking with a view to its further use in a software environment for statistical data processing. The tasks of regional benchmarking are the selection of the leading region for comparison in a certain area of development, the identification of differences between it and the region being analyzed, as well as an analysis of the key success factors of the leading region. A special feature of strategic planning for regional development based on smart benchmarking is the preliminary identification of structurally similar territories. The authors revealed the essence and advantages of this approach, defined criteria for comparing regions and made their systematization. The algorithm of benchmarking procedure is developed and its testing is carried out in the regions of Russia. The practical significance of the algorithm developed by the authors of the benchmarking procedure is that it can be used by government authorities to justify the priorities of regional development. And further automation of the proposed model will allow the software product to be applied to any government authority, regardless of the initial parameters of its development.

1 INTRODUCTION

Realization of modern tasks, projects, management programs and business-engineering is difficult to imagine without the use of specialized software products, i.e. without automation. Automation of processes allows the management of operations, data, information and resources through the use of software that reduces the degree of human participation in the process, or completely excludes it. The main purpose of automation is to improve the quality of the process.

Questions about the need for continuous improvement and automation of processes are actual and widely discussed. Scientists came to the conclusion that the existing methods of economic evaluation are cumbersome and difficult, and it leads to huge losses of time and resources, therefore achievement of effective work results requires the automation of these processes.

K. S. Braunwarth, M. Kaiser and A. L. Müller [1] actualize the issue of implementation of information systems in the processes of economic analysis. The scientists concluded that automation increases the speed of searching and analyzing information, as well as increasing the efficiency and quality of the assessment.

Speaking about the automation of regional management systems, we should note the work of P. V. Stankevich, M. V. Kopnov, A. V. Kudinov and A. I. Finko [2]. They describe the opportunities for improving regional management systems by automating applied management tools in detail.

One of the modern tools for regional management is territorial benchmarking. Benchmarking in the broad sense is understood as the process of improving the activity of the analyzed object by transplanting the best practices. Initially, the term "benchmarking" was related to the concept of "dantotsu", meaning "effort, concern of the leader to become even better leader". It is believed that this definition was formed in the late 50th in Japan, after the Japanese experts visited the leading companies of the USA and Western Europe in order to study and subsequently use their experience in management. The name of the method "benchmarking" comes from the English words "bench" (level, height) and "mark" (a mark), the combination of which is treated as "reference mark", "elevation mark", "reference comparison", etc.

Initially, the methodology of the benchmarking procedure was developed to improve business processes in various fields of activity: marketing, assortment policy, personnel management, logistics, pricing policy, etc. At the same time, benchmarking is widely used at present to solve regional problems with the aim of the identification of unique opportunities in the territories [3] - [5]. In this case, the term "regional benchmarking" refers to "interregional comparisons of types of activities, processes, practices, policies and the use of this information to improve regional development" [6].

So, in the majority of the analyzed works of Russian authors in the field of regional benchmarking research the authors used a method based on the selection of "best samples" by comparing individual criteria for socio-economic and innovative development. At the same time, the regions were compared with those that showed the best indicators of socio-economic development, regardless of whether they have similar characteristics or not.

At the same time, the latest developments of foreign regional scientists have indicated the objective shortcomings of this approach, which they call "simplified benchmarking" [7]-[9]. The essence of the criticism lies in the absence of a preliminary analysis of the reasons and prerequisites for achieving leadership by the best regions. Indeed, the initial conditions for the development of territories are determined by formal and informal institutions that have developed historically and are essential for innovation [10]. Simply put, for objective reasons, not all indicators of the development of leading regions can be achieved by outsider regions in practice in the foreseeable future. Therefore, the results of "simplified benchmarking" represent nothing more than regional ratings, and therefore cannot be an acceptable basis for developing an effective development strategy for the territories.

It should be noted that the advantages of using another type of benchmarking, called "smart / system benchmarking" were described in detail in 2001 [11]. The meaning of this type of benchmarking is based on a preliminary analysis of the baseline development conditions of the compared subjects. Thus, smart benchmarking of regional systems involves the creation of a development strategy based on comparison with territories that have similar institutional conditions and development indicators.

This article presents an algorithm for automating a regional management system based on smart benchmarking technology for its further use in a software environment for statistical data.

2 DATA AND METHODOLOGY

The methodology developed by the Basque Institute of Competitiveness [12] and used by regions and countries of Europe to create effective innovative strategies for territorial development was used as a methodological basis for conducting a benchmarking procedure. As the basic factors used to compare the regions, those factors were selected that, firstly, reveal the strengths and weaknesses of the territory to the best way, and secondly, do not tend to change in the short term.

The authors singled out seven criteria for regional development: the geo-demographic criterion, the criterion for the formation of the population, the criterion for the development of innovations, the criterion of the sectoral structure of the economy, the criterion of business conditions, the criterion of openness in foreign economic activity, the criterion of institutions and values. Each criterion includes a system of statistical indicators (a total of 34 indicators), on the basis of which the territories are compared.

3 CONCEPTUAL MODEL OF SOFTWARE PACKAGE

In order to automate the "smart benchmarking" procedure for further use in software environments for statistical data processing, you need to create a model. For more visual visualization, we divided it into several stages (Figure 1).

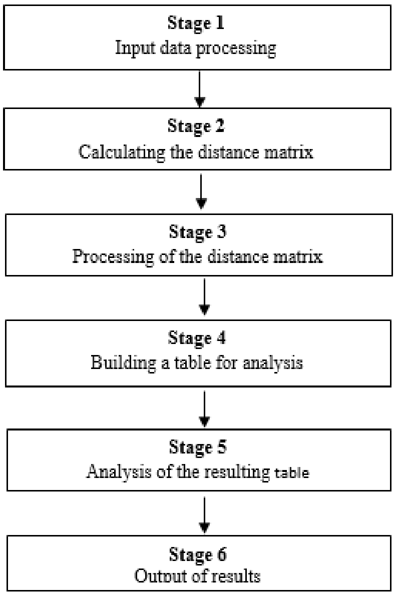


Figure 1: Stages of automation of the process of managing regional development based on the use of smart benchmarking technology.

The first stage is responsible for the mobilization and processing of input data that will be used in the software product. At the second stage, the structural distances indices are calculated to construct the distance matrix. Distance matrices are the basis for conducting a smart benchmarking procedure. At the third stage, the distance matrix is processed to determine identical areas for the analyzed region. At the fourth stage analytical tables of data on identical regions are formed. The fifth stage analyzes the regional data obtained. During the sixth stage, the user can make the necessary visualization of the data in the form of convenient tables in Excel and graphs. Let's take a closer look at each stage of automation.

3.1 Input Data Processing (Stage 1)

To start the program it is necessary to download statistical data from the website of the Federal State Statistics Service, so the user specifies the year in which he wants to conduct the benchmarking procedure. Then the data is loaded into a multidimensional array. After that, the average Russian values in each criterion will be calculated and placed in a separate array, which will be used in the analysis in step 4. It is also necessary to bring the values of all statistical indicators to a single scale.

Step 1.1. Evaluation of the asymmetry of the each indicator distribution.

The importance of asymmetry characterizes the degree of asymmetry in the distribution of the statistical indicator relative to the average value of the indicator for the country. If the asymmetry value is greater than 0.5, then to smooth out the "emissions" (extreme values), each value of the indicator is transformed by the formula (1):

$$x_{ij} = \sqrt[k]{x_{ij\_0}} \quad (1)$$

where x_{ij} – transformed value of the j-index of the i-region;

x_{ij_0} – initial value of the j-index of the i-region;

k – degree of asymmetry (takes values from 2 to 4, depending on the magnitude of the asymmetry coefficient).

Step 1.2. Reduction of values to one scale.

The indicators expressed in percentage remain unchanged, the rest ones are converted into percentages relative to the sum of the values of the variable of the corresponding indicator (2):

$$\overline{x_{ij}} = \frac{x_{ij}}{\sum_j x_{ij}} \cdot 100\% \quad (2)$$

Where $\overline{x_{ij}}$ – normalized value of the j-index of the i-region.

After that, it is possible to go to the next stage.

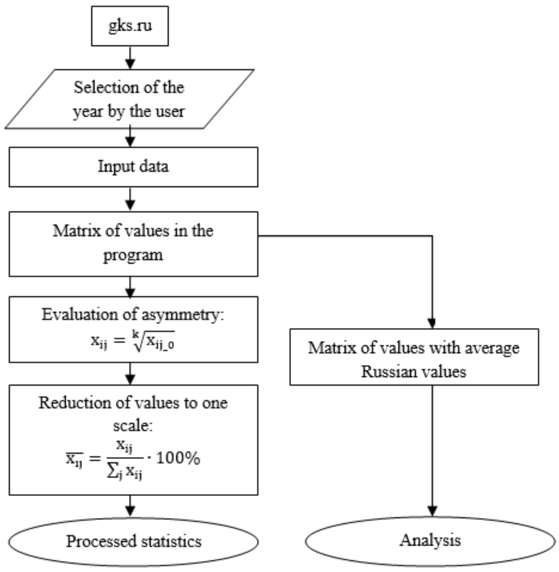


Figure 2: Input data processing.

3.2 Construction of a distance matrix (Stage 2)

At this stage the user selects the analyzed region and the program calculates the distance matrix based on the statistical data grouped in the criterion. It is assumed that each statistical indicator makes the same contribution to the formation of the criterion value. This assumption is based on the results of the study (Navarro et al., 2014), which proved that the use of different weights does not give significant changes in the final results. At the same time, the use of different weights substantially increases the subjectivity of the method as a whole. In view of this, each of the 7 criteria for comparing regions is given an equal weight, which is distributed in equal parts between the variables (3) that enter into it:

$$m_j = \frac{1}{7} / k \quad (3)$$

where m_j – weight coefficient;

k – number of indicators characterizing the criterion.

Further, the index of the structural distance is calculated using the formula (4), along which the distance matrix is constructed:

$$d(i,i') = \sum_{j=1}^{k} m_j (\overline{x_{ij}} - \overline{x_{i'j}})^2 \quad (4)$$

where $d(i,i')$ – index of structural distance of i-region;

$\overline{x_{ij}}$ – value of the j-value of the i-initial region;

$\overline{x_{i'j}}$ – value of the j-value of the i – «other» region.

After the distance matrix is constructed, it is stored in a separate array of values.

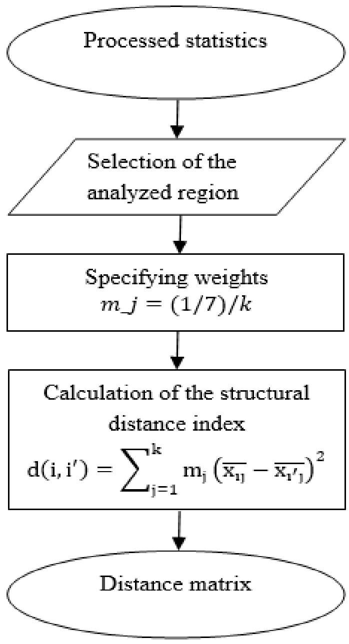


Figure 3: Construction of a distance matrix.

3.3 Processing of the distance matrix (Stage 3)

At this stage the received values are sorted for the region selected by the user. Sorting is done in ascending order with reference to the names of regions. After that the user will be asked to enter, with how many identical regions the program will need to work. Next, the program will display the required number of regions in order of increasing the index of the structural distance. In addition, the region closest in terms of structural characteristics will be singled out separately.

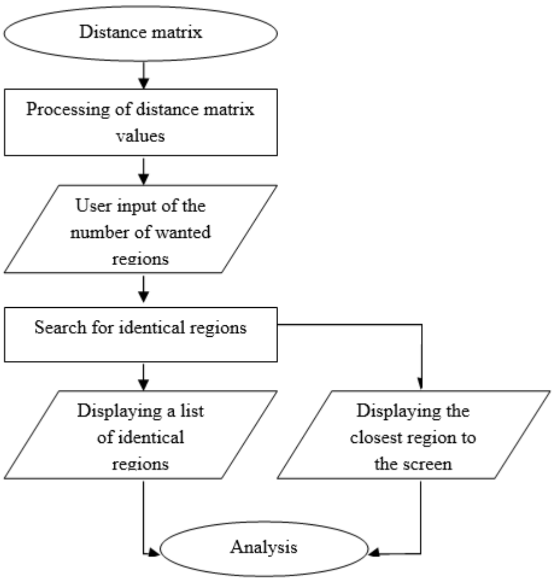


Figure 4: Processing the distance matrix.

3.4 Formation of analytical data tables for identical regions (Stage 4)

At this stage, a new multidimensional array of values from the original data will be created. This array can be imported into Excel tables. The tables will contain statistical data for a given number of identical regions.

At the end of the resulting table the program calculates the average and maximum values of each column. The last line records the average Russian values that have been received in stage 1 (Table 1).

Table 1: Example of a table for the analysis of the selected number (n) of identical regions.

	Criterion x_1	Criterion x_2	...	Criterion x_l
Analyzed region	76,6*	672	…	84125
The nearest region	76,1	668	…	28632
Region (2)	77,3	659	…	26385
...	…	…	…	…
Region (n)	76,9	671	…	39712
Maximum value	79,3	675	…	84125
Average value	73,2	670	…	42748
The average Russian value	70,4	625	…	27985

* - values are random

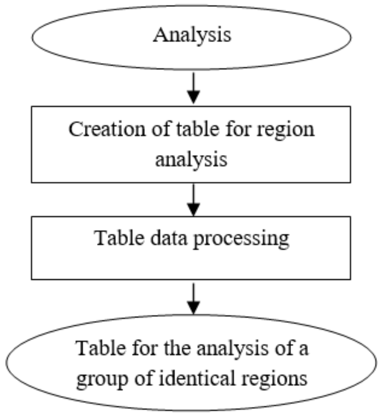


Figure 5: Formation of analytical tables of data by identical regions.

3.5 Analysis of the obtained data (Stage 5)

At the beginning of this stage, the best indicators are determined among identical regions for each of the 7 criteria (column). To do it, a maximum value was found in step 4 in each of the columns representing the best region indicator by this criterion.

After that the strengths and weaknesses of the region analyzed by the user are determined in relation to the average values for Russia, as well as for a group of identical regions (these values were also found in step 4). To determine the strengths and weaknesses of the analyzed region in comparison with the average Russian values for each criterion, the value of R_i is calculated using formula (5):

$$R_i = \frac{x_{i_{\text{chosen region}}}}{x_{i_{\text{average Russian value}}}} \quad (5)$$

To determine the strengths and weaknesses relative to the average value among identical regions, the value is calculated in the following way I_i (formula 6):

$$I_i = \frac{x_{i\text{region chosen}}}{x_{i\text{identical regions}}} \geq 1 \quad (6)$$

If the obtained ratios $R_i < 1$ and $R_i < 1$, then the i-criterion refers to the weak side of the region in comparison with the average Russian value and average value among the identical regions values, respectively. Conversely, if the value of

$R_i \geq 1$ and $I_i \geq 1$ refers to the strong side of the analyzed region.

The received data is stored in a multidimensional array, which will be presented to the user in the form of a table 2:

Table 2: Multidimensional array, reflecting the strengths and weaknesses of the analyzed region (a schematic distribution of criteria in the cells of the table).

	Strengths		Weaknesses	
Comparison of the analyzed region with the average Russian indicators	**A**	The criterion x1 The criterion x2 The criterion x3	**B**	The criterion x4 The criterion x5
Comparison of the analyzed region with identical regions	**C**	The criterion x2 The criterion x4 The criterion x5	**D**	The criterion x1 The criterion x3

As can be seen from Table 2, some criteria in the columns in the comparison of the indicators of the selected region with the average Russian indicators and with the indicators of identical regions coincide. Thus, the identification of the priorities for the development of the territory on the basis of "smart" benchmarking should be carried out by a consistent double comparison of the average all-Russian indicators and the average indicators of identical regions.

For this purpose, the program will have a separate cycle to understand whether it is necessary to use the direction of development of the leading regions among identical regions, or whether the program of the analyzed region should be left unchanged. The symbols of the location of the criteria are entered in the table, which reflects the strengths and weaknesses of the analyzed region (see Table 2). For example, A is a cell in which a strong criterion of the analyzed region is located in comparison with the average Russian values. To determine whether it is necessary to change the development priorities of the analyzed region, it is required to check each available criterion from the table. If the "Criterion xi" is simultaneously presented in cells "A" and "C" or in cells "B" and "C", then priorities should not be changed in the development direction of the analyzed region, expressed by the criterion xi, since this is the strength of the chosen region. However, if the "Criterion xi" is not repeated in cells "A" and "C" or in cells "B" and "C", then for the analyzed region in the direction of development expressed by the criterion xi, it is necessary to adopt the best experience in the leading region for this criterion among identical regions.

Also, for greater visibility, a table of priorities for the development of the analyzed region is proposed. This table will be compiled from the statistical data of the selected region, the average for identical regions, the best indicator for identical regions, the average Russian indicator, the leader region among identical ones, and also the column with the results of the criterion check. At the same time, if the criterion is tested positively, there will be a "+" (it means that the directions of development of the leading region should be taken into account), with negative - "-" (it means that this criterion should remain unchanged, since this is the strength of the chosen one region). An example of the priority table is shown in Table 3.

Table 3. Priorities for the development of the analyzed region.

	Selected region	Average for identical regions	Best among identical regions	Average for Russia	Leading regions among identical	Is it worth taking the direction of development
The criterion x_1	76.6*	73.2	79.3	70.4	Kaluga region	+
The criterion x_2	667	670	675	625	Arhangelsk region	+
The criterion x_3	2230	1763	1305	1681	Kirov region	+
...	…	…	…	…	…	…
The criterion x_n	884	427	841	279	Perm region	-

* - values are random

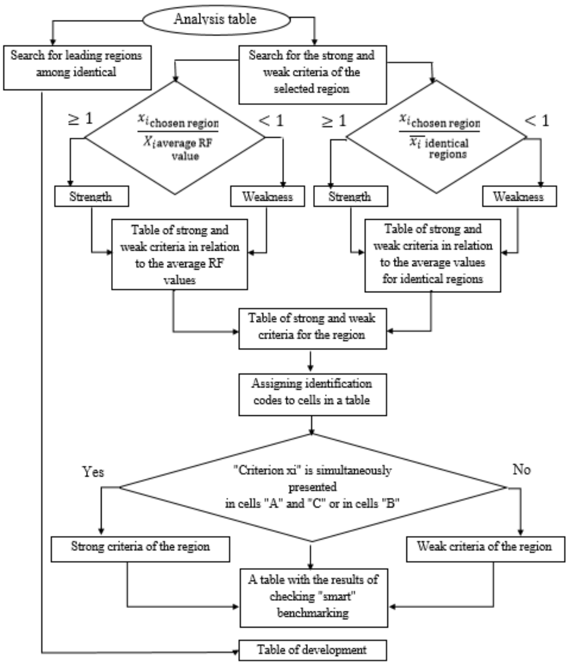


Figure 6: Analysis of received data.

3.6 Output of results and visualization (Stage 6)

At this stage, the user can perform the necessary visualization of the data in the form of selected tables in Excel and graphs according to the data of the multidimensional array.

The user selects the table numbers in the dialog box himself.

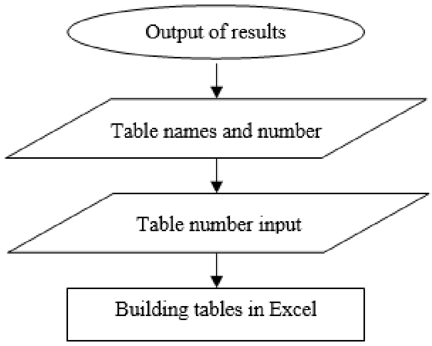


Figure 7: Output of results and visualization.

4 APPROBATION OF METHODOLOGY

The automation of the "smart benchmarking" procedure was carried out in accordance with the steps described in section 3.

Statistical data of Russian regions for 2016, published by the Federal Service for State Statistics, were used as an experimental base. The Republic of Crimea and Sevastopol did not participate in the regional benchmarking because of the lack of data for comparison. As a result, a complete matrix of distances was constructed for the regions of Russia. The Perm region was chosen as the analyzed region.

To determine the regions identical to the Perm region, a common matrix of distances was constructed for the regions of Russia, the elements of which are structural distance indices. Figure 8 shows a fragment of the matrix containing regions identical to the Perm region.

According to Figure 8, Kaluga Region has the smallest value of the structural distance index with the Perm region.

Region	Perm region	Kaluga region	Arhangelsk region	Republic of Tatarstan	Ulyanovsk region	Kirov region	Vologda Region	Vladimir region	Tver region
Perm region		0,428	0,476	0,598	0,642	0,714	0,828	0,865	0,953
Kaluga region	0,428		0,655	1,528	0,591	0,655	1,443	0,311	0,845
Arhangelsk region	0,476	0,655		0,863	0,998	0,516	1,470	1,210	0,628
Republic of Tatarstan	0,598	1,528	0,863		1,180	0,886	0,935	1,897	1,040
Ulyanovsk region	0,642	0,591	0,998	1,180		0,621	0,667	0,776	0,688
Kirov region	0,714	0,655	0,516	0,886	0,621		1,021	1,205	0,076
Vologda Region	0,828	1,443	1,470	0,935	0,667	1,021		2,151	1,015
Vladimir region	0,865	0,311	1,210	1,897	0,776	1,205	2,151		1,503
Tver region	0,953	0,845	0,628	1,040	0,688	0,076	1,015	1,503	

Figure 8: Matrix of distances of identical regions for the Perm region.

Further, the characteristics of the Perm region development were determined, which have low criteria values in comparison with the average and maximum indices of identical regions, and also in comparison with the average country indicators. The systematization of the obtained statistical data is given in Table 4.

Table 4: Systematization of statistical data characterizing the development of the Perm Territory.

	Strengths		Weaknesses	
Comparison of the analyzed region with the average Russian indicators	**A**	- **The volume of innovative products**; - Number of enterprises and organizations; - Foreign economic activity	**B**	- **Age structure of the population;** - Transport accessibility - The level of education **- Crime level**
Comparison of the analyzed region with identical regions	**C**	- Number of enterprises and organizations - Foreign economic activity	**D**	**- Level of culture, recreation and tourism** **- Transport accessibility** **- The level of education** **- The volume of innovative products**

Table 4 highlights with bold the criteria for which attention must be paid. Thus, the indicator "the volume of innovative goods, works, services", the value of which for the Perm region is above the national average, at the same time, in comparison with identical regions, shows a significant lag. Thus, the identification of the priorities for the development of the territory on the basis of "smart" benchmarking should be carried out by means of a consistent double comparison of the average countrywide indicators and indicators of identical regions.

On this basis the main priorities for the development of the Perm region were identified. They include increasing the transport accessibility of the region, raising the level of education, increasing the volume of innovative goods, works and services, as well as raising the level of culture and leisure in Perm region. The identified priorities are the basis for identifying and analyzing best practices among identical regions in terms of their development programs and the application of innovative policy tools.

5 CONCLUSIONS

A given algorithm for the automation of the process of the benchmarking procedures could make a meaningful contribution to the process of developing, implementing and monitoring innovative strategies of territory development by identifying weaknesses and competitive advantages of the region.

This method of benchmarking is universal and can be applied to any territorial units of both regional and municipal levels.

In the future we plan to register a specialized software tool that allows us to identify structurally similar subjects of the Russian Federation in order to model the optimal development for a particular territory.

As the most significant results of the study we note the following.

First, we justified that the technology of smart benchmarking is a successful decision to enhance the effectiveness of territorial development. The purpose of using this concept is to inform the authorities about the main directions, conditions and opportunities for further development of the region. At the same time, the key objectives of benchmarking remain: the selection of a region-leader for comparison in a certain area of development, the identification of differences between it and the region being analyzed, and an analysis of key success factors for the region-leader.

Secondly, we have empirically proven that identification of territorial development priorities based on smart benchmarking should be carried out by means of a consistent double comparison of the average countrywide indicators and indicators of identical regions.

Thirdly, for the first time, the algorithm of the benchmarking procedure for territorial units was presented and tested.

We believe that the development and improvement of the regional benchmarking procedure with respect to territorial entities requires further research and is of scientific interest in creating an interactive tool that synthesizes statistical data in order to create an innovative national economy.

ACKNOWLEDGMENTS

This work is supported by the Russian Foundation for Basic Research (project "Modeling of the regional development strategy based on smart benchmarking" № 18-310-00263).

REFERENCES

[1] K.S. Braunwarth, M. Kaiser, A. L. Müller (2010), "Economic Evaluation and Optimization of the Degree of Automation in Insurance Processes", Business & Information Systems Engineering, vol 2, 1st ed, pp. 29-39.

[2] P. V. Stankevich, M. V. Kopnov, A. V. Kudinov, A. I. Finko (2014), "Regional Information System Development Based on Web-technologies", Available at: http://portal.tpu.ru/SHARED/k/KUDINOVAV/Tab2/Tab/ifost%20paper.PDF

[3] N. Groenendijk (2010), “EU and OECD Benchmarking and Peer Review Compared”, The EU and Federalism: Polities and Policies Compared, Ashgate, pp. 181–202.

[4] L. Iurcovich, N. Komninos, A. Reid, P. Heydebreck and Y. Pierrakis (2006), “Mutual Learning Platform: Regional Benchmarking Report - Blueprint for Regional Innovation Benchmarking”, European Commission, Brussels 2006.

[5] R. Huggins (2010). “Regional Competitive Intelligence: Benchmarking and Policy-making”, Regional Studies, vol 44, No 5, pp. 639-658

[6] C. Koellreuter (2002), “Regional Benchmarking as a tool to improve regional foresight”, European Commission-Research DG -Directorate K, April 2002, pp. 14.

[7] J. Navarro, J.P. Smart (2017), “Specialisation benchmarking and assessment: pilot study on wind energy”, Available at: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC106035/2017_03_09_wind_pilot_regions_final.pdf (accessed 01 November 2017).

[8] M. Navarro, J.J. Gibaja, S. Franco, A. Murciego (2012), “Territorial benchmarking methodology: The need to identify reference regions”, Innovation, global change and territorial resilience, pp. 99-133.

[9] M. Navarro (2014), “Regional benchmarking in the smart specialisation process: Identification of reference regions based on structural similarity”, S3 Working Paper Series, no 03, pp. 9-10.

[10] C. Nauwelaers, R. Veugelers, B. Van Looy (2003), “Benchmarking National R&D policies in Europe: Lessons from Belgium”, Available at: http://www.stis.belspo.be/docs/papers/pdf/Nauwelaers.pdf (accessed: 01 November 2017).

[11] M. Tomlinson, B.A. Lundvall (2001), “Policy learning through benchmarking national systems of competence building and innovation – learning by comparing”, Available at: http://www.academia.edu/24150494/Policy_learning_through_benchmarking_national_systems_of_competence_building_and_innovation-learning_by_comparing (accessed: 01 November 2017).

[12] Smart Specialisation Platform (2016), Available at: http://s3platform.jrc.ec.europa.eu/s3-platform (accessed: 01 November 2017).

Specifics of Project Management on Industrial Innovation

Stepan Mezhov[1] and Leonid Mylnikov[2]
[1]Altai State University,
Lenin Avenue 61, 656049, Barnaul, Russia
[2]Perm National Research Polytechnic University,
Komsomolskiy Avenue 29, 614990, Perm, Russia
megoff@mail.ru, leonid.mylnikov@pstu.ru

Keywords: Innovations, Modelling, Strategy, Multiplier, Funding Of Innovative Processes, Net Present Value Adjustment.

Abstract: The paper offers a scientific and methodological approach to assessing the efficiency of innovative projects which develops and supplements the theoretical bases of innovative projects analysis through the accounting of recursive dependence of investments on such factors of innovative process as (1) the constructive complexity of the product, (2) the number of stages of the innovative process and (3) the innovative capacity of the enterprise. It is also proposed to enter an adjustment factor (multiplier) to reflect the causal connection between the net present value of innovative projects and the abovementioned factors of innovative process and to improve the accuracy of calculation of cash flows under the project linking the assessments with the certain stages of the project and the corporate capacity. This will reduce the overall integral risk of project implementation for potential investors. The author's concept and the system of models for planning the production and innovations allow revealing quantitative correlation between investments, innovations, and sources of financing and can be used for a reasoned strategic decision making by enterprises implementing innovative development model.

1 INTRODUCTION

Global industrial development demonstrates the strengthening role of innovative production in mature economies. Innovations become the key factor of competitive capacity [4]. Taking into account the increasing backlog of our industry from many other countries we need an intensive and scientifically grounded systemic modernization of the national economy. The economic growth through oil and gas industry is at the limit, and the oil prices no longer determine the dynamics of GDP growth.

The transition of Russian economy to the innovative path of development occurs under the lack of theoretical elaboration and the violation of principles of systemic approach. In particular, the program documents do not provide a clear description of the mechanism and concept of «material media» of innovative industrial transformations. They address side issues like intellectual property rights, venture funds, the support for young scientists and some others which are important but do not help solving the core problem of organization of mass knowledge-intensive production [3][9].

One of the vital problems in managing a modern corporation is the construction of its economics in the line of innovative production. The science lacks the methodological substantiation of the concept of building the innovative economy. Therefore, there is a need of formation of theoretical and methodological foundations for the organization and management of modern innovation-oriented corporations in Russia.

Methodological aspects of this problem are reflected in the works on instrumental methods of research, i.e., on the assessment of innovative potential, the economic and mathematical modelling of innovation planning [1][2][5][6]. Still, there is no clear understanding that large industrial companies are the core of development strategies. Not enough attention is paid to the problem of efficiency of innovations from the point of coordination of production and innovation processes in an industrial enterprise. Methods of corporate modelling require clarification and development to become the efficient planning tools.

2 METHODOLOGY

Production program is the most important element of corporate planning. At that, this task allows rather clear formalization and application of programming tools of decision-making. A production program stipulates optimum availability of resources and calculates all the technical, economic, and financial indicators and parameters. However, production planning is much more complicated for an enterprise which competitive capacity is based on innovations and constant launch of new products. The authors take as a premise that (1) production plans should include the planning of innovations and investments in innovations; (2) assessment of investment efficiency requires a simultaneous and coordinated forecast of cash flows from all activities; and (3) the production planning shall be a long-term prospective forecast.

The research of production structures with innovative potential that allows covering all the stages of innovation process on the basis of constant renewal of products and receiving the rent during a long period of time methodologically relies on the introduced notion of innovation-oriented corporation. This notion summarizes and systemizes those variants of defining the enterprises oriented on innovations which are most popular in scientific literature.

3 CORPORATION AND THE ECONOMIC MECHANISM OF ITS SUSTAINABLE DEVELOPMENT

At the end of XX - beginning of XXI century, the leading industrial corporations underwent a large-scale reconstruction following the new paradigm of organization of business and competition: a modern enterprise is a multilayer structure that integrates in time and space the flows of resources which evolve at a different speed. At that, the category of «resource» is significantly expanding and is supplemented with the notions of «key competences», «dynamic abilities», and «routines» [8][10].

Such corporations successfully combine the tactical and strategical aspects of their activity. What is the foundation for the economics of a leading western or Japanese company? It definitely possesses enough assets, the technology relevant to the industry development level, the R&D facilities, the market share and cost structure that ensure balanced production of goods and innovations with a pre-set standard return on assets.

Our hypothesis is that such an innovation-oriented corporation shall follow certain functional relations between key parameters such as equity, output and sales, production and innovation costs, return on investment (ROI), and etc. Speaking about innovations, we have to understand how they are financed, how the strategic stability is ensured, and how the balance between the initial costs and return is reached.

Our study and research are focused on the following concept: speaking about innovations, we see the main outcome in the form of cash income. ROI is the return on capital: intellectual, technical, and managerial. How this capital shall work in a corporation?

A competitive production of a modern corporation consists of the processes of production (operations) and innovations. Operational process means solving of current tasks of production and sales of products. It is the source of financial resources for all forms of investments including innovations. The process of innovations solves the prospective tasks for future production (transfer of competition from the sphere of production to the sphere of innovations). The processes of operations and innovations have a sequential and parallel logic of interaction and can be formally presented as a set of life cycles. Consequently, a modern corporation has certain proportions between the processes of operations and innovations which shall be reflected in the production plan.

A company invests in both processes, operations and innovations, but it looks like the added value is generated only by the capital invested in operations. In fact, in innovation-oriented companies the investments in innovation provide added value with a certain time delay. Taking into account the probabilistic nature of this process, we have to mention that this is the most complicated task of corporate planning of innovations.

4 MODEL OF PLANNING OF NEW PRODUCTS IN CONDITIONS OF CONSTANT INVESTMENT IN INNOVATIONS

We offer a concept of Operations and Innovations Program (OIP) as a unified and balanced in the selected time interval plan of production and sales of products and innovative works on the preparation of production of promising products substituting phase out products. The main task of such a program is in the synthesis of models of innovative and production processes in a way ensuring rational combination of limited resources for production of goods and innovations with an increase or, at least, the maintenance of competitive capacity in the strategic perspective.

For this purpose, such planning shall resolve the following tasks:

- Preparation of an optimum production plan taking into account the demand, the available resources, and the development strategies;
- Definition of suitable investment funding strategies and the assessment of return in the form of rent and profit;
- Definition of time of the launch of new products and the withdrawal from production of the old ones;
- Balancing of operations and innovations in the program through the calculation of financial indicators, forecasting of balance and preparation of data for the start of the next iteration in the planning interval.

This model includes forecasting blocks of the operational plan, the choice of innovation process implementation strategy and the assessment of investment efficiency. Purpose of the model: determining the optimum structure of production oriented on market demand, the assessment of investment funding strategies including the assessment of innovative potential and stage of RTD funding. It is also used to forecast the payback of a new product, calculate the income and rent using such data as forecasted live cycles of each product, the financial and economic characteristics of the enterprise and some others. The balance of model calculations between the operational and innovational parts of the OIP is ensured by the Higgins model [7] which shows how the assets and a number of other financial indicators shall change under changing sales and costs.

The demand for products is forecasted by N application of Monte-Carlo method to the life cycle interval, with the subsequent averaging. At that, marketing department shall define the forecasted demand for a new product within developed market development strategy. This demand will represent exogenous parameters of the simulation.

The “Functions Costs-Sales” block sets the sales (based of forecasted demand), costs and profit functional parameters for each product while the integral characteristics may be obtained from corporate reports. These functions are the basis for criteria, constraints on costs, output, and capacity of the optimization model.

The optimization model is formed automatically. The results are first processed by the “Calculation of financial indicator” block, then – by the “Calculation of innovation and investment indicators” block to obtain NPV, payback period, NPV adjustment, and rent assessment, define the final parameters, and deliver the variants of OIP.

When modelling the OIP (Fig. 1), each planning stage shall deliver a separate optimization model with its own parameters and constraints defined by the current state of corporate resources at the moment *t*.

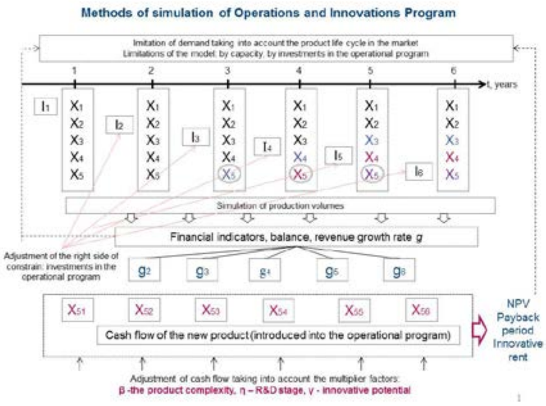


Figure 1: OIP simulation approach.

Balance sheet indicators are simulated using the optimization model. Financial indicators are calculated using balance sheet calculations for each forecasting year.

The block of calculation of investment characteristics solves the following tasks: the assessment of R&D investment efficiency, the calculation of NPV, the adjustment of NPV taking into account the current innovative potential (using of expressions (1) and (4)), the consideration of changes in the innovative potential depending on investments, the calculation of rent from production and sales of a new product.

Innovative potential assessed through processing the results of questioning the experts is not a constant value. This potential is a dynamic parameter which varies depending on the volume and intensity of investments in innovations. According to our methodology, the increase in innovative potential reduces the adjustment factor and the payback period of an innovative project for launching a new product. The development and future production of a new product suggest the receipt of Schumpeterian (business) rents what should be taken into account in the simulation by changing the price of the new product. Price of a new product depends on the competitive influence on the market. The rent to be received from production and sales of the new product is calculated for the strategies defined by the marketing department.

Our methodological statements allow adjusting the volume of investments in innovations taking into account the payback periods and the balanced growth rates.

5 ASSESSMENT OF THE VOLUME OF INVESTMENTS IN OIP INNOVATIVE PROJECT

The prepared OIP model addresses rather important theoretical and practical issues related to the dependence of the volume of investments on the characteristics of the innovation process and the level of corporate innovative potential. Technological parameters and the innovative potential in general have a definite but not a direct influence on the efficiency of innovative projects. An accurate assessment of investments, efficiency and payback period of innovative project based on NPV requires a methodical approach to the investment analysis of innovative projects which would take into account the technological and organizational characteristics of the corporation.

More accurate definition of investments in innovations and the assessment of ROI require clarifying the basic NPV concept from the point of dependence of the result from the complexity of the product, the depth of the innovative process and the innovative potential of the enterprise by including these characteristics in the NPV calculation formula.

The analysis of scientific literature on the issue and assessment of efficiency of innovative projects shows that the existing methods do not take into account a number of essential factors that will be mentioned below.

Total expenses on innovations in an industrial corporation include expenses on purchasing the scientific equipment, apparatus, software, wages of scientists and engineers, patent and informational support, outsourcing of works and financing of certain projects. Expected income of a classical investment project mainly depends on the amount of investments, rate of return (ROR) and risk, while an innovative project has a complex systemic dependence mainly from technical and economic, or technological factors.

Formally, the forecasted investments in an innovative project I_n depend on a certain set of parameters and factors:

$$I_n \leq \Psi(c_{nj*}, x_{nj*}, \delta, \beta, \eta, \gamma, T_n) \quad (1)$$

where c_{nj*} is the expected price of a developed innovation product $j*$ substituting the “old” product j; x_{nj*} is the expected total output of the new products; δ is the planned ROR including the risks and the required ROR defined by the corporate management; Ψ is the type of function; β is the structural complexity of the product which may be assessed by an expert or analytically in comparison to an analogue; η is the depth of scientific study (the development of theory, survey, technology) which reflects the level of spending on R&D stages; γ is the integral criterion of scientific and production potential which reflects the availability of key competences, routines, the level of equipment of laboratories, and the similar depth of penetration of the industry. I.e., the corporation as an average representative of the industry has its history, a stable market share, the innovations, and etc.; T_n is the payback period defined for a fixed level of investment.

We offer the procedures for assessing the technological parameters and innovative potential. Such parameters as the structural complexity of the product, the depth of innovative process, the innovative potential of the enterprise can be assessed by questioning the qualified experts in the fields of science and production.

Our concept is based on the following assumptions:

- A corporation with a high innovative potential, competences and abilities will require less investment in R&D to create the same product than a corporation with a smaller potential. In the current term, it will spend less on preproduction and research equipment and training as it has more qualified staff, a developed scientific and production base, an information base and essential reserves in many directions of development. The corporate experience also has a great meaning.
- In the absence of potential ($\gamma \rightarrow 0$), the costs of innovations are very large; and vice versa, the higher is the potential, the lower is the cost of innovations.
- A corporation can become innovative at any stage: the basic research, the applied research, or can get a patent and start developing a concept product and the design. From the strategic point of view, its decision-makers have to decide on their intentions: do they want to start the research at an early stage with

higher I_n, to create key competences and routines for this area and to gain additional income (rent), or they want to save I_n but to lose the rent as the competitors will also be able to deliver this product.

Some properties of exponential function allow building models of accounting and analysis of investments in innovations, especially, through defining the α argument by regression analysis on the basis of the available dynamic ranks.

Let's see how the parameters of innovative potential affect NPV. We have defined that the higher is the potential γ, the less initial capital is required, and vice versa. Also, the deeper is the innovative process η, the more funds are required for its implementation. The same refers to the product complexity: the more complex is the product (the greater the value of the coefficient β), the more investment is needed.

Complex influence of these factors characterizing the innovative project and the enterprise itself can be accounted for by introducing a special multiplier (adjustment factor), for example, by using an exponential function,

$$\sigma = e^{\frac{b\beta\eta}{\gamma}} \quad (2)$$

where b is the coefficient taking into account the cost forecasting accuracy in development of complex products.

The parameter b can reflect the degree of inaccuracy of the planned investment decisions and can be defined by an expert. I.e., if 6 out of 10 planned investment decisions with the pre-set parameters in average have significant deviations of actual parameters from the planned ones, than b is equal to 0.4. The b value shall be defined by experts from corporate planning department as an adjustable parameter.

The properties of expression (2) depend on the following restrictions of the function indicator σ:

$$0 \le \beta \le 1;\ 0 \le \eta \le 1;\ \gamma d \le \gamma \le 1.$$

1. σ = 1, if β = 0 v η = 0; σ > 1; 1 if β ≠ 0 v η ≠ 0. Here, "v" means logical operator "or". But b can be equal to 0, so the planned system absolutely precisely defines the efficiency of investments as $-\sigma = e^0 = 1$. Therefore, the correcting coefficient has no influence on the investment project efficiency.
2. σ → ∞, at γ → 0. So, a decrease of innovative potential γ results in a significant growth of σ.
3. As the coefficients of complexity and depth are in the range of 0 ≤ βη ≤ 1, the variable $\frac{b\beta\eta}{\gamma}$ has the maximum value at γ = γd, βη =1, that is, when the product has the highest complexity and the lowest level of preparedness, and the potential γ is at the extremely low level. As σ has the minimum value of 1 under the provided restrictions on dimension parameters, the values of β and η (characteristics of a certain innovative project) at various values of γ with the adjustment factor b are needed to calculate the usual *NPV* (without adjustment).
4. The parameter b is used to define the multiplier σ for a certain enterprise considering its technical and organizational characteristics and the statistics of investments in different projects.

Innovative project investment strategies will be integrally assessed using the adjusted NPV calculation procedure. We have used the innovative potential of the enterprise, the product complexity, the depth of R&D and the multiplier to obtain an equation for assessing the amount of initial investments in an innovative project depending on the characteristics of the innovative process:

1. ρ_{nj*} is the return on current costs on production x_{nj*}, and is the sum of the return on costs Rc_{nj*} and the additional yield from the Schumpeterian rent Ren_{j*} :

$$\rho_{nj*} = Rc_{nj*} + Ren_{nj*} \quad (3)$$

2. The product price is the production of yield and costs S_{nt} in the period t:

$$c_{nj*} = S_{nt} + \rho_{nj*}S_{nt} = (1 + \rho_{nj*})S_{nt} \quad (4)$$

With the adjustment, the NPV calculation formula will look as follows:

$$NPV = -I_n^{п} + \sum_{t=1}^{T} \frac{(1-\tau)(Rc_{nj*}\, S_{nj*}\, x_{njt*} + Ren_{nj*}\, S_{nj*}\, x_{njt*} - S_{cnsj*})}{(1+\delta)^t} \quad (5)$$

Let us transform the formula (5) into the equation for the case when NPV = 0 to receive the equation for assessing the amount of initial investments in an innovative project depending on the characteristics of the innovative process:

$$NPV = -I_n^{п} e^{\frac{b\beta\eta}{\gamma}} + \sum_{t=T1+1}^{T} \frac{(1-\tau)(Rc_{nj*}\, S_{nj*}\, x_{njt*} + Ren_{nj*}\, S_{nj*}\, x_{njt*} - S_{cnsj*})}{(1+\delta)^t} \quad (6)$$

where *T1* is the time spent on R&D, τ is the income tax rate.

Here, the production $Rc_{nj*} S_{nj*} x_{njt*}$ represents the operating income, and $Ren_{nj*} S_{nj*} x_{njt*}$ is the rental income from the competitive advantage of the corporation. In competitive environment, the corporations follow up with promising products and the yield gradually decreases to reach the industry average.

So the amount of investments in complete development of an innovative product including the launch of its production is within the interval bounded from below by the planned output, and from the top – by the adjusted amount of investments taking into account the structural complexity of the product, the number of stages of scientific development, and the scientific and production potential of the enterprise. The top bound of the interval mostly depends on the innovative potential which structure depends on the technological level, the human capital, the R&D, and etc. The higher is the innovative potential, the shorter is the interval; than the forecast of the actual investments in innovations is more accurate.

Preparation of raw data, calculation of such parameters as the innovative potential, the depth of the innovative process, the forecasted values of demand, prices and costs of the products, and formation of strategies of launching the new product are followed by the assessment of strategies of the OIP models and the selection of optimum variant by certain criteria.

6 CONCLUSION

We have shown the causal relationship of net present value innovative projects with the factors: the structural complexity of the product, the number of stages of the innovation process, innovative potential of the enterprise. To assess the factors of the innovative project, expert-analytical method.

We proposed to reflect this relationship a adjustment factor (multiplier). The introduction of a multiplier allows for more to correctly calculate the cash flows of the project, reducing the integral risk of the project to potential investors.

In fact, we formulated the theoretical position that the actual value of the investment in innovation lies in the interval, the lower boundary of which is a target value, and upper boundary adjusted by the coefficient (multiplier) the amount of investment.

This interval defines the problem of choosing the optimal investment strategy. The upper limit of the interval depends on the innovative potential: the higher the level of innovative capacity, the interval is narrowed, which improves the accuracy of the forecast investment. It develops and refines the theoretical framework of the estimation of efficiency of investments.

Further studies are interesting from the point of view of key problems of organization of innovative processes in the companies – the unresolvable dilemma of the insider. Why many private companies are reluctant to develop and invest on some limits? This will increase transparency, which will lead to the loss of control over a stable stream of income.

Therefore, the focus on domestic funding sources is a survival strategy, is futile from the point of view of overcoming the problems and difficulties faced by the Russian economy. The expansion and understanding of this phenomenon needs more research, content analysis, and the development of new models and tools.

REFERENCES

[1] M. A. Al-Fawzan, A. Al-Hargan, “Promoting techno-entrepreneurship through incubation: An overview at BADIR program for technology incubators”, Innovation: Management, Policy and Practice, 16(2), pp. 238-249, 2014.

[2] A. Baniak, I. Dubina, “Innovation Analysis and Game Theory: A Review”, Innovation: Management, Policy and Practice, 14(2), pp. 178-191, 2012.

[3] E. Carayannis, Y. Goletsis, E. Grigoroudis, “Multi-level multi-stage efficiency measurement: the case of innovation systems”, Operational Research: An International Journal, 15(2), pp. 253-274, 2015.

[4] H. Chesbrough, J. West, W. Vanhaverbeke, Open Innovation: Researching a New Paradigm, Oxford University Press, Oxford, 2006.

[5] A. Chuvaev, M. Khayrullina, O. Kislitsyna, “Production systems continuous improvement modelling”, Quality Innovation Prosperity, 19 (2), pp. 73-86, 2015.

[6] O.G. Golichenko The National Innovation System: From Concept to Research Methodology // Problems of Economic Transition, 58 (5), pp. 463-481, 2016.

[7] R. Higgins, Analysis for Financial Management, 10th edn, the McGraw-Hill/Irwin series in finance, insurance and real estate), New York, NY 10020, 2012.

[8] A. Rosiello, M. Mastroeni, M. Teubal, G. Avnimelech, “Evolutionary Policy Targeting: Towards a Conceptual Framework for Effective Policy Intervention”, Technology Analysis & Strategic Management, vol. 25(7), pp. 753-772, 2013.

[9] J. Schumpeter, Capitalism, Socialism, and Democracy, Taylor&Francis e-library publishing, First published in the UK in 1943, 2003.

[10] Teece, D. J, A dynamic capabilities-based entrepreneurial theory of the multinational enterprise, Int Bus Stud (2014) 45: 8. https://doi.org/10.1057/jibs.2013.54.

A Concept for a High-reliability Meteorological Monitoring System Using AMQP

Anna Kostromina[1], Eduard Siemens[1] and Yurii Babich[2]

[1]*Future Internet Lab Anhalt,*
Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Köthen, Germany,

[2]*Department of Telecommunication Networks,*
Odessa National O.S. Popov Academy of Telecommunications, Odessa, Ukraine
anna.kostromina@student.hs-anhalt.de, eduard.siemens@hs-anhalt.de, y.babich@onat.edu.ua

Keywords: AMQP, Rabbit MQ, Meteorological Monitoring, Reliable Message Queues, Availability

Abstract: This paper describes a concept for a resilient meteorological monitoring system for reading data from sensors by using AMQP in order to increase reliability of the data acquisition system. A set of sensors is connected to the Beaglebone Black and is located in the mountains of north of Thailand. Gathered data are queued on a local SoC and sent to a server located in Germany whenever a network connection is available. Further in the work implementation and test of such a system in Thailand is discussed. Special challenges in the implementation of the system is the presence of frequent thunderstorms and outages caused by them. To improve the reliability of data transmission free AMQP implementation is used. The protocol has been studied, tested, and programs have been created for transmitting data from sensors to a server in Germany using the Rabbit MQ to store the data in the case of connection failures. Besides that, memory usage problems were raised when using AMQP in single-board computers, such as Beaglebone Black. The main task of this work is to propose the most stable and reliable operation of the data transmission system.

1 INTRODUCTION

Monitoring systems are nowadays used in a very wide range of technical and technological systems. This paper presents a concept of a meteorological monitoring system running in a rural area under harsh weather conditions in the north of Thailand. The goal of this research was to design and to deploy a monitoring station in an indigenous village of a Karens tribe. The station shall be involved in long-term monitoring of weather conditions in order to obtain the data required for simulations of energy availability and consumption minimization in the case of energy-autarkic cooling houses, designed for the ASEAN region.

This work was part of the preparatory work for the Silaa Cooling project where the team of the Future Internet Lab Anhalt (FILA), along with three industrial partners from Germany and one Thailand company are going to develop a coffee-cooling system for long-term coffee storage. Figure 1 shows a typical village, at which such a coffee cooling system has to be built and operated.

The main technical challenges under the given conditions are as follows:

- The system must have a battery buffered energy supply due to frequent power outages at the facility
- The system needs a solid weather protection but also a flash protection of all the electronic and electrical components due to frequent thunderstorms
- All the data must be gathered and reliably transmitted to a cloud environment at the facilities in Germany via a GSM modem. Since the internet connectivity at the location is very unstable and drop-offs of the internet connection can last for days or even weeks, some robust data queuing and delivery must be implemented, which stores data locally and, delivers queued data whenever the internet connectivity to the domestic cloud is available.

The queueing system shall have a persistncy of data which prevents data loss if a power outage occurs at any stage of data acquisition and queueing

- The data must be kept in a quite generic way since at the time of gathering of the data the exact needs of later data processing is not yet completely known. So, it is advisable to keep the data at the receiving site in two formats – in a non-growing data-base with data aggregation and compression, primarily for visual representation of the data which must be available in semi-real-time. Along with that, in a raw data format without data reduction is needed, for later in-depth analysis of the particular weather parameters.

Figure 1: Village in the Karens area at which the weather data have to be gathered.

2 RELATED WORK

The development of self-sustained and energy-autarkic systems significantly grows in the importance during recent years. Especially in the development aid as well as in the implementation of decentralised energy production. The chain between information and communication is crucial and promises to enable implementation of economically reasonable systems for assuring a subsistence of people in rural, mostly off-grid, areas. One of the approaches similar to SilaaCooling is the ColdHubs system. That project aims at supporting people in Africa to store the food in solar-powered cold rooms [1]. However, the technical approach of ColdHubs is quite static, the energy provision is secured by some overprovisioning of the amount of photovoltaic modules and of the lead batteries capacity, which makes hard to keep the deployment costs reasonably low. In contrary to ColdHubs, the SilaaCooling concept includes heuristic and proactive energy management, based on weather and energy demands prediction, operated on a computer on-side, equipped with a collection of sensors of ambient condition. The idea and technical realization of SilaaCooilng bases on a concept of the implementation of energy-autarkic decentralised energy supply for small villages in Siberia, developed in one of previous projects [2].

However, reliable metrological and weather data acquisition and delivery is also used in other application cases. So in [3] a data acquisition system consists of a set of wireless sensors for measuring meteorological parameters. While the focus in that research is on the local system design for connecting wireless sensors, our focus is mostly on the implementation of a robust resilient data delivery system.

To avoid data loss on data delivery to the cloud, existing transmission and queuing protocol called Advanced Message Queueing Protocol (AMQP) [4] has been investigated and tested. Among several AMPQ implementations, the Rabbit MQ system has been compared with Active MQ. A good comparison of the said implementations is given in the article “The analysis of the performance of RabbitMQ and ActiveMQ” [5]. Furthermore, in [6] and [7] an in-depth performance analysis for distributed message delivery using RabbitMQ is given. However, due to a quite low message load in our target scenario, performance is not considered as a bottleneck of our system. The mentioned resilience and persistency of data is in the foreground of our investigations.

A good and detailed tutorial of use of RabbitMQ is given in [8]. Here, a lot of important information about the work and writing programs for messaging using RabbitMQ is given.

3 THE PROJECT ELEMENTS

3.1 Hardware of the project

3.1.1 Components of monitoring system

In this work we mostly focus on the choice of a robust queuing and message delivery system able to work under the harsh conditions described above.

To build a self-sufficient food storage we need to collect the following weather data: ambient temperature, humidity, wind gust, solar irradiance. The latter parameter is crucial for the later dimensioning of the solar panel size and also for the size of battery buffer for the SilaaCooling system, to be developed.

For gathering data in Thailand, in Mai Lae, a Beaglebone Black (BBB) has been installed, running Debian 8 Linux. The cheapest and easiest way to get necessary weather parameters was to get an off the shelf weather station with a USB interface:

- WH1080 Radio weather station [9] is connected to the BBB, which gathers the ambient temperature, humidity, wind data, air pressure and amount of rain in a 10 minutes interval to the weather station via an UHF radio link. Up to 14 hours of data can be buffered on the weather station without data loss in case of downtimes of the BBB.
- For redundancy and for verification of the data gathered by the weather station, we have decided to add an additional sensor set to the system. For this an AM2302 temperature-humidity sensor, placed at a JeeNode board has been placed on a pillar next to the sensors of the weather station, which deliver their data via UHF to a JeeLink, plugged via a USB hub to the BBB.
- For solar irradiance metering a Si-RS485TC-T sensor, manufactured by Ingenieurbüro Mencke & Tegtmeyer GmbH [10], is used, which is connected to the BBB via an USB/RS 485 adapter. A daemon running on the BBB is reading the irradiance data in a 30 seconds cycle. In contrary to temperature and humidity, this sensor was not doubled, for cost reasons and since the sensor is a calibrated one with certification.

AM2302 is a temperature and humidity sensor used to verify temperature and humidity data, measured by the weather station. The sensor is mounted on a JeeNode which transmits data via UHF band to the Jeelink board, connected to the BBB, see Figure 2.

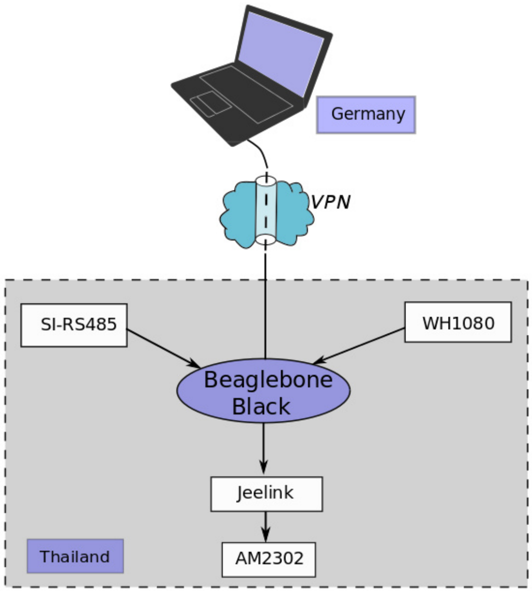


Figure 2: Components of monitoring system.

3.1.2 The data acquisition system

Data from the three different sources – the weather station, irradiance meter and JeeNode connected sensors are read by a tiny Python program. For reading data from weather station WH 1080, the API of "pywws" [11] is used by our Python program. The main loop of iteration through the different sensor data is running in a daemon fashion, which is monitored by the Supervisor daemon of Linux. The JeeNode sensor data are also written by a python script, which has been developed in a preparatory work for the SilaaCooling project. Figure 8 shows how output data from producers sending on Rabbit MQ and then on servers.

3.2 The software concept

The main idea in the implementation of the system is to use publicly available open-source code wherever it is possible for keeping the development effort for the system low. For gathering the weather data from the weather station, the open source python package "pywws" [11] has been used.

The main technical challenge of the system is however the message queuing and delivery system, which shall on the one side work on a low-performance SoC like BBB and on the other hand be very robust with avoidance of data loss even in the event of abrupt power outages, connection drops or link failures to the sensors. For this, the Advanced

Message Queuing Protocol (AMQP) has been deployed on the system [3]. Though this system is originally known as a message broker in banking transactions, the robustness shall be used in our design. While a wide range of AMQP implementations are known on the market, there are some lean implementations which promises to run on a low-performance PC platform.

Rabbit MQ is an open source AMQP suite implementation which provides an exhaustive Python API for Linux. Reasons to choose Rabbit MQ are the following [8]:

- RabbitMQ is the only one open source implementation of the AMQP standard besides Qpid, whereby Qpid doesn't provide a Python API

- Essential features like federation, clustering, persistency are implemented in RabbitMQ
- Clustering became simpler because of Erlang
- RabbitMQ is more reliable and crush resistant that it competitors

Clustering connects multiple machines together to form a single logical broker. Communication is via Erlang message-passing, so all nodes in the cluster must have the same Erlang cookie. The network links between machines in a cluster must be reliable, and all machines in the cluster must run the same versions of RabbitMQ and Erlang [12].

By default, each Rabbit MQ instance delivers its messages from the queue to the consumers using the Round-Robin algorithm. Also by default, each message it is deleted from the queue on delivery. To change this behaviour were used flag auto_delete=False

3.2.1 Implementation

The instantiation of a RabbitMQ broker consists of the following steps [8]:

1. Connect to Rabbit MQ
2. Obtain a channel
3. Declare an exchange
4. Create the message
5. Publish the message
6. Close the channel
7. Close the connection

3.2.1 The publish/subscribe model

In the application of weather station gathering two independent data destinations are necessary – one which delivers to a fixed size data base for monitoring purposes and another one with ever growing raw data. The both subscribers (consumers) can then be co-located at the same machine or at geographically remote locations.

For implementing this, a channel publishes two queues and two clients, which run AMPQ consumers, consumes the messages from the respective queue and stores the sensor data to each data base. This pattern is known as publish/subscribe pattern (Figure 3).

- Producer (P) is a user application that sends a message to an exchange
- Exchange(X) receives messages from producer and push them to the queues
- Queue is a buffer that stores the messages
- Consumer(C_1, C_2) is a user application that receives messages v

In figure 3 is shown as a working publish/subscribe algorithm.

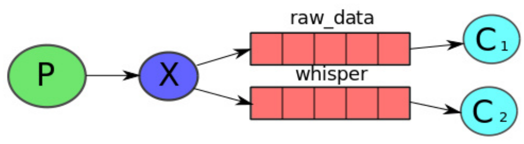


Figure 3: Publish/subscribe.

In case of publish/subscribe using the exchange type as fanout. From exchange data were sent into queues raw_data and whisper.

The following customization has to be applied to the default configuration of RabbitMQ to meet the requirements of the weather data gathering for SilaaCooling. The data gathering has to be reliable even in cases of frequent abrupt and long-term power outages. The first measure is making the message queues **persistent** and **durable**. With persistency, each data chunk written to the queue is copied to a SD card. Also the file system on the sd-card is not buffered, so at each instance the queue is mirrored to a non-volatile memory. This approach for sure reduces the transport performance of the queue to the read/write speed of the SD card. However, this must be pretty enough for delivering a few messages per minute with some hundreds of bytes per read/write. With the durability, the AMQP is saving **the configuration** of the published queues to a non-volatile memory so, on a re-start of the system and of the AMQP daemon (which provides the exchanges, channels and queues) the queues of the messages are automatically published and so the clients can re-subscribe to the queue and continue the message reception from the queue.

Between reading the data from the queue by a consumer and bringing it to save harbour in a data base, some network errors, data base errors or file system errors can occur. So, a message shall be deleted from the queue only if the data base confirms the storage of the data item. For this an acknowledgment of messages has been configured in the RabbitMQ system.

3.2.2 Error resilience tests

Before the system can be deployed in the field in the mountains of Thailand, a series of scenarios of power disruptions, network errors and other failures has been defined, with which message integrity checks has been performed. The system for testing is shown on Figure 4.

The Rabbit MQ server and producer which read the sensor data are running locally on a Beaglebone Black.

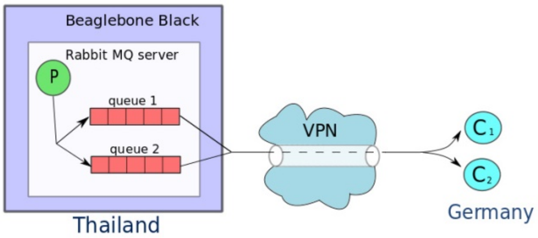


Figure 4: System for sending a data by using Rabbit MQ.

A first test scenario contains the following steps of tests:

1) abrupt disconnect of consumers (Figure 6)
2) crash of the BBB
3) stop a consumer and re-start after several minutes to hours
4) power on of the BBB
5) re-connect the producer (while the consumer is already running)

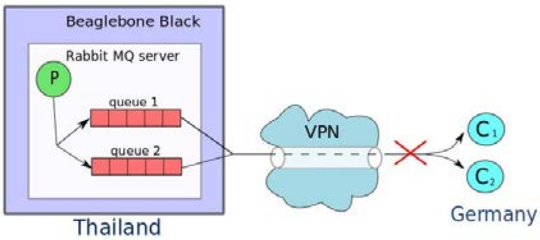


Figure 5: Disconnect of consumers.

In all the test iteration, no messages where lost. All of them has been delivered and stored in the respective data bases at the consumer's site. n this case messages were successfully delivered after the disconnect.

In a second series of tests, the behaviour of the system in case of producer crash or BBB crash including the Rabbit MQ server is investigated. For that the following sequence has been executed:

1) Disconnect of consumers
2) Killing Rabbit MQ server and producer (Figure 6)
3) Restarting Rabbit MQ
4) Reconnect consumers
5) Restart of the producer

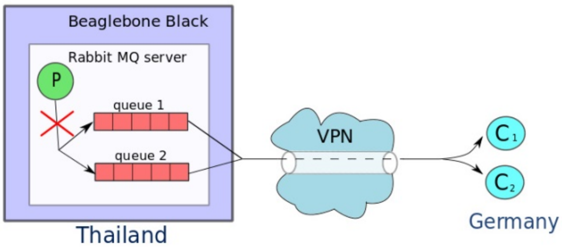


Figure 6: Crush of Rabbit MQ producer and server.

All messages were delivered after restart. Testing were successful.

3.3. Final concept of system for data transmission

Finally we have implemented the concept as follows. Figure 7 shows how data from producers are sent to consumers. For each program a producer (solar_producer, pywws_producer, silaacooling_producer) is instantiated and then data are written to the queue through the AMQP exchange. For every program two queues are created – the raw_queue and whisper_queue. All data from raw queues are going to the consumers on the cloud server and data from the whisper queues are going to the consumers on the other server. These servers are located in Germany.

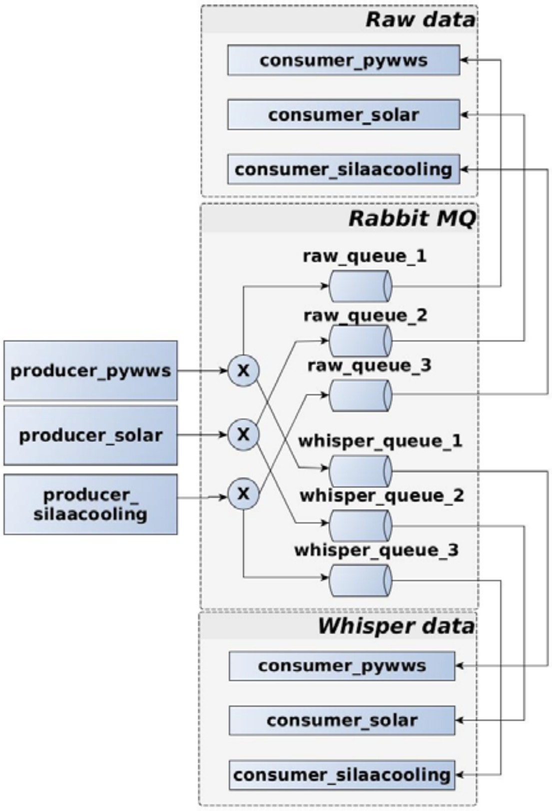


Figure 7: Sending data on Rabbit MQ.

4 PREDICTIVE MONITORING OF MEMORY USAGE FOR RABBIT MQ

Rabbit MQ provides a basis for a robust and error resilient message delivery system in presence of a variety of network and system errors in the field, which is lean enough to run on a single-board computer with rather limited CPU and memory resources. However, crashes due to memory overruns can lead to message loss, which must be avoided by additional means. To control memory usage we propose to implement a predictive monitoring approach described in [15]. In this case, the monitored parameter is the value of memory in use marked y. It makes sense to use only one threshold value corresponding to available memory denoted as y_{kP}. If the value of memory in use reaches the threshold value we consider having emergency situation, which must be prevented by means of the predictive monitoring. In order to minimize load on single-board computer resources (memory, processor) we use only one forecast horizon out of three specified in [15]. The following expression corresponds to the normal operation of the system.

$$\hat{y}\left(n_{\min}^{C}+L_{C}\right)<y_{kp}, \quad (1)$$

where $\hat{y}$ denotes a forecasted value of the y parameter estimated using the minimum required series of the $n_{\min}^{C}$ size with the time horizon of L_C .

If the term (1) is violated memory usage predictive monitoring system alerts about the expected emergency situation in order to let take actions and prevent the situation. If it is impossible to take any actions due to lack of connection, than the predictive monitoring system deletes some data from the memory in order to prevent system crash.

One message of *pywws* program contains 98 bytes on average, one message of the *solar_producer* program (for reading Irradiance) contains 93 bytes and one message of *silaacooling_producer* program contains 362 bytes. Consequently, for 24 hours without connecting to consumers the data will use 41.376 Kb of memory.

In the case of Beaglebone Black or other type of SoC it is very important to check if there is enough memory or not. In particular case is good decision to use some mechanism for removing old log files from the system without creating problems when reading data. In case of using a similar system a calculation, how long it will take to fill the entire free space on the computer's disk is inevitable.

5 CONCLUSIONS

A system design, for data delivery from the acquisition part to different consumers positioned thousands of kilometres away, interconnected via narrowband erroneous GPRS links is shown in the paper. Stable and resilient data delivery without message losses, even in cases of network outages for several days has been reached with the system. It has turned out as very flexible, so the location of the consumers can be easily migrated from one location to another one. Using two different queues leads us to a very convenient decoupling of different kinds of use of the data for post-processing. The memory usage predictive monitoring system has been implemented in order to prevent system crash caused by the lack of available memory, which is crucial for the Rabbit MQ implementation on a single-board computer.

REFERENCES

[1] Project Coldhubs, http://www.coldhubs.com/

[2] S. Zinov, E. Siemens. "The Smart Lighting Concept". Workshop on Problems of Automomous Power Systems in the Siberian Region. Koethen, October 2013.

[3] M.Benghanem, "Measurement of meteorological data based on wireless data acquisition system monitoring", Applied Energy, pp. 2651-2660, December 2009

[4] S. Vinoski, "Advanced Message Queuing Protocol", IEEE Internet Computing, vol. 10, Issue 6, pp. 87-89, November 2006

[5] V. M. Ionescu, "The analysis of the performance of RabbitMQ and ActiveMQ", IEEE, Romania, pp. 132-137 , October 2015

[6] Jones, et. al., "RabbitMQ Performance and Scalability Analysis", project on CS 4284: Systems and Networking Capstone, Virginia Tech 2011

[7] M. Rostanski K. Grochla,A. Seman "Evaluation of highly available and fault-tolerant middleware clustered architectures using RabbitMQ", IEEE, Poland, pp 879-884, October 2014

[8] A. Videla and J. Williams, "RabbitMQ in action. Distributed messaging for everyone." Manning, April 2012.

[9] "Radio Weather Station with USB and Touchscreen" [Online]; http://www.produktinfo.conrad.com/datenblaetter/650000-674999/672861-an-01-ml-FUNK_WETTERSTATION_MIT_USB___TOUCH_de_en.pdf

[10] "Digital silicon Iradiance Sensor Si-rS485-TC-T" [Online], http://imtsolar.com/wp-content/uploads/2013/02/Si-Sensoren_RS485_E.pdf

[11] J. Easterbrook, "pywws" [Online] https://github.com/jim-easterbrook/pywws

[12] "Distributed RabbitMQ brokers" [Online], http://www.rabbitmq.com/distributed.html

[13] "RabbitMQ documentation" [Online], http://www.rabbitmq.com/documentation.html, accessed 21.01.2014.

[14] "Advanced Message Queuing Protocol. Protocol Specification" [Online], https://www.redhat.com/f/pdf/amqp/amqp_0-8_specification.pdf

[15] Y.O. Babich and L.A. Nikityuk, "Functional improvement of monitoring the dynamic characteristics of information and communication networks", vol. 4/9 (76), Eastern European Journal of Enterprise Technologies, Kharkiv, 2015, pp.9-14.

Experiences Implementing QoS Aware Routing on Off-the-shelf SDN Switches

Jannis Ohms, Olaf Gebauer,
Nadiia Kotelnikova, Marina Arikova and Diederich Wermser
Research Group IP-Based Communication Systems, Ostfalia University of Applied Sciences, Salzdahlumer Str. 46/48, D-38302, Wolfenbüttel
jannis.ohms2@ostfalia.de, ola.gebauer@ostfalia.de, nadiia.kotelnikova@hs-anhalt.de, m.arikova@ostfalia.de, d.wermser@ostfalia.de

Keywords: SDN, OpenFlow, Routing, QoS

Abstract: This paper provides an overview of the Quality of Service (QoS) capabilities defined in the OpenFlow specification. Several vendor documentations from off-the-shelf products are compared with the OpenFlow specification. This research reveals inconsistencies between the specification and the vendors implementation. Queues for examples are not implemented by all vendors. This gap can lead to interoperability problems in a network while using hardware from different vendors. The research also shows, that the majority of vendors provide a port statistic function which gives information about incoming and outgoing bandwidth about each port of a switch. Based on this function a QoS aware routing application for off-the-shelf switches is proposed. This concept can be used to change the flow of traffic in an OpenFlow network based on the utilization of the interfaces. Based on the conducted research, the application can be used with hardware from multiple vendors. This paper does not contain a quantitative evaluation of the implemented application.

1 INTRODUCTION

Software-Defined Networking (SDN) is a new concept for the implementation of computer networks. The most commonly used realization of the SDN concept is the OpenFlow Protocol [1] which is specified by the Open Networking Foundation (ONF). SDN Networks consist out of multiple SDN switches and at least one SDN controller. The controller has an overview of the whole topology and creates flow rules for the switches. The controlled switches forward packets according to the flow rules they received. The use of SDN in large networks creates a requirement for QoS aware routing. The goal of QoS aware routing is to choose one of the multiple possible paths in the network under consideration of the QoS requirements of the transported traffic. Classical routing algorithms were developed for autonomous systems (AS) [2] where each router makes its own routing decisions. In SDN the routing decisions are provided by the centralized SDN controller. This form of centralized control is visualized in Figure 1 and Figure 2. The new form of centralized control makes it necessary to rethink QoS aware routing in the context of OpenFlow. In this paper, the authors will take a look at the QoS mechanisms provided by the OpenFlow specification, and compare it with the QoS mechanisms implemented in commercially available off-the-shelf OpenFlow hardware. This paper also describes a concept to provide QoS aware routing which uses the QoS mechanisms implemented in commercially available off-the-shelf OpenFlow hardware. The concept has been implemented as an SDN application.

2 RELATED WORK

OHMS et al. [3] showed that it is possible to use the queueing mechanism of an off-the-shelf OpenFlow switch to provide QoS for Voice over IP (VoIP) streams. The used topology consists of a single switch. GUCK et al. [4] analysed a set of routing algorithms in the context of SDN. The analysis focused on resource consumption and efficiency. The algorithms have only been simulated and not implemented on top of real OpenFlow hardware. ZHANG et al. [5] compared the performance of OpenFlow with routing protocols like Open Shortest Path First (OSPF). Their results show that OpenFlow can react much quicker to topology changes when compared to OSPF. JINYAO et al. [6] proposed a QoS aware routing concept which uses queueing and queue statistics. The concept has been imple-mented and tested on a single computer using a network simulator. The related work indicates that QoS aware routing is a research topic in the scientific community. Our focus on off-the-shelf OpenFlow hardware makes this paper unique when compared to the related work.

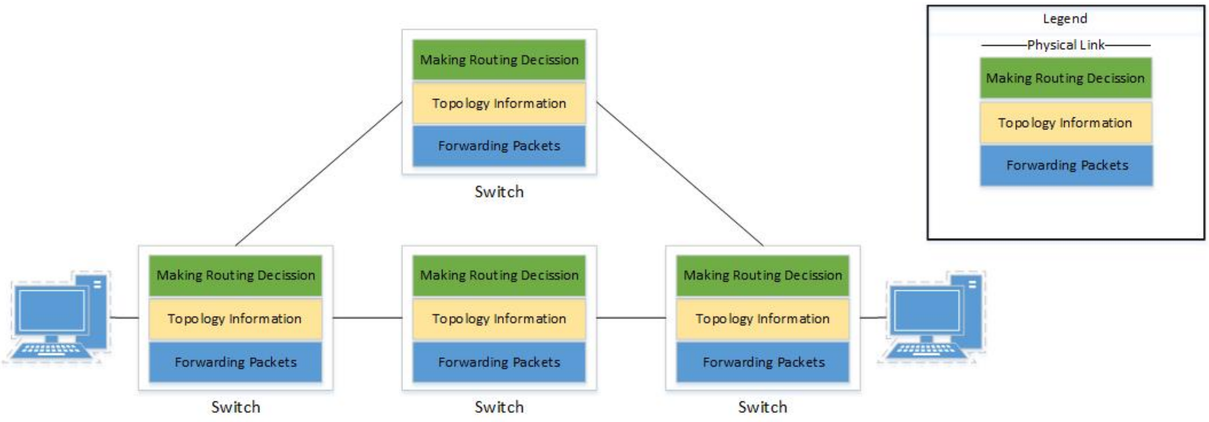


Figure 1: Routing in a classical network.

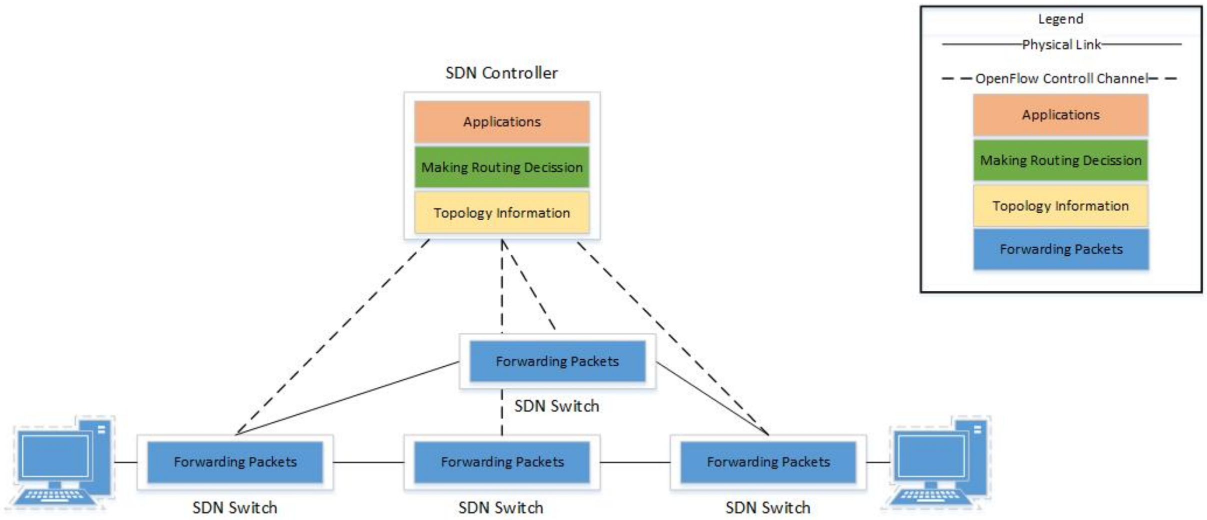


Figure 2: Routing in an SDN network.

3 QOS CAPABILITIES DEFINED IN THE OPENFLOW SPECIFICATION

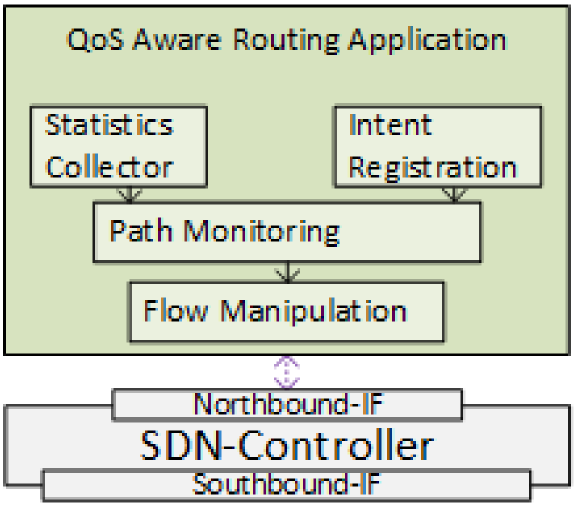


Figure 1: Architecture of the developed QoS aware routing application.

All statements made are based on the OpenFlow specification 1.3 which is the most commonly implemented version regarding the hardware available on the market [1]. OpenFlow 1.3 provides three mechanisms to enable QoS.

Queueing

OpenFlow 1.3 provides a class-based queueing algorithm on the egress port, as a not mandatory part of the specification. The queues provide a guaranteed bandwidth and a maximum bandwidth. The standard specifies OpenFlow messages to configure queues and to get statistics about bandwidth and errors for each queue. Each flow table entry can use an enqueue action to insert a matched packet into a queue. The standard does not specify a precedence between the queues or a required number of queues. The standard also does not specify any queueing algorithm.

Metering

The OpenFlow specification provides a metering mechanism which limits the bandwidth of a given flow. When the limit of a meter is exceeded, all traffic which goes through the meter gets dropped. Alternatively, the Type of Service (ToS) bits of the IPv4 header can be rewritten. The specification contains OpenFlow messages to configure meters and to get statistics about bandwidth and errors for each meter. Each flow table entry can use an action to assign a matched packet in a given meter. The specification does not specify the number of meters which have to be implemented on the switch.

Port Statistics

The specification provides port statistics which measure the sending and receiving bandwidth for each port of the switch. The OpenFlow specification contains a message which enables the controller to collect port statistics.

4 QOS CAPABILITIES OF THE OPENFLOW IMPLEMENTATION FROM DIFFERENT VENDORS

The authors compared the OpenFlow capable product families of four different vendors. Namely HP, Brocade, Juniper, and Pica8. This comparison focuses on the QoS capabilities described in chapter 3. This comparison is based on documentation provided by the vendors [10][11][12][13].

Queueing

Not every vendor implements queues. If they do, they use between 4 to 8 queues per port with a hierarchical precedence. The queue configuration messages are not implemented by any vendor, the queues are usually configured over proprietary CLI interfaces. Most implementations use the HTB scheduling algorithm [7]. The queue statistic messages are only implemented on the Brocade Netiron switches.

Metering

Metering is implemented by all vendors. The meter statistics are not always implemented. The meters are configured through a proprietary CLI interface.

Port Statistics

Port statistics are implemented by all hardware vendors used in this comparison.

Based on this results port statistics seem to be the only commonly available mechanism for our application. The use of port statistics has a significant drawback. The application cannot detect full queues. This can lead to packet loss if a port seems to be idle regarding overall bandwidth consumption while one or more queues exceed their upper bandwidth limit. This problem can be solved using queue statistics.

5 CONCEPT AND ARCHITECTURE OF A QOS AWARE ROUTING APPLICATION

The application has been developed as an internal module for the open source SDN controller Project Floodlight [8]. The architecture of the SDN controller is visualized in Figure 5. The controller provides basic functionality implemented as modules, for example, information about connected hosts, current network topology, etc. Some functions are used for the development of the application. The application for QoS aware routing consists of four logical components.

Statistics Collector

The Statistics Collector collects the port based statistics from the switches. This component contains a background task which gets executed periodically at a fixed rate.

Intent Registration

The Intent Registration component allows the registration of intents with QoS requirements. An intent is a desire of a terminal endpoint to communicate with another terminal endpoint. It consists of a set of header fields which identifies a set of packets. The Intent registration uses the Routing Service of the SDN controller to find every possible path to fulfil the desired intent. All paths get observed by the Path Monitoring.

Path Monitoring

The Path Monitoring component has a background task which receives data from the Statistics Collector to evaluate the bandwidth consumption for every possible port on every path for every registered intent. If the bandwidth capacity of a path which is currently in use by an intent gets exceeded, the Path Monitoring uses an alternative path if possible. This component uses the Flow Manipulation component to change the flow of traffic. The evaluation of every path might not be possible in large topologies. In this case, a preselection is necessary to reduce the set of paths to a size which can be handled by the application. When an intent gets unregistered, the monitoring of the possible paths is canceld.

Flow Manipulation

This Flow Manipulation component uses the Switch Service of the SDN controller to push flow table entries for every switch on a given path. This component opens the possibility to create end-to-end flows for a given path. There is no need to manipulate each flow table of every switch directly.

This architecture is visualized in Figure 3. The routing process can be separated into the following steps.

- If a new intent gets registered, the application looks up every possible path which connects the terminal endpoints of the intent.
- If a path has been found, the application assigns one path to the intent.
- Every packet which matches the header fields (which are specified in the intent) gets forwarded through the assigned path.

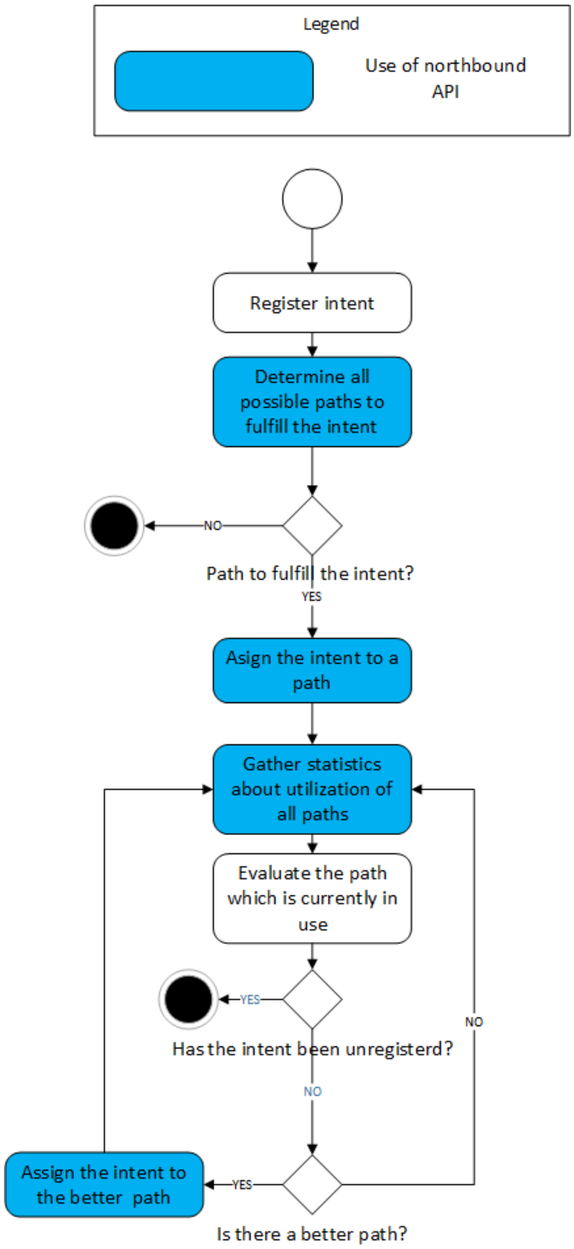


Figure 2: Concept for QoS aware routing presented by the authors.

- The application collects statistical data about the utilization of every port on every path which is currently tracked.
- The statistical data is used to compare the utilization of the current path with every possible path.
- If a better path is available, the application reassigns the intent to the better path.

The concept of the application is visualized in Figure 4.

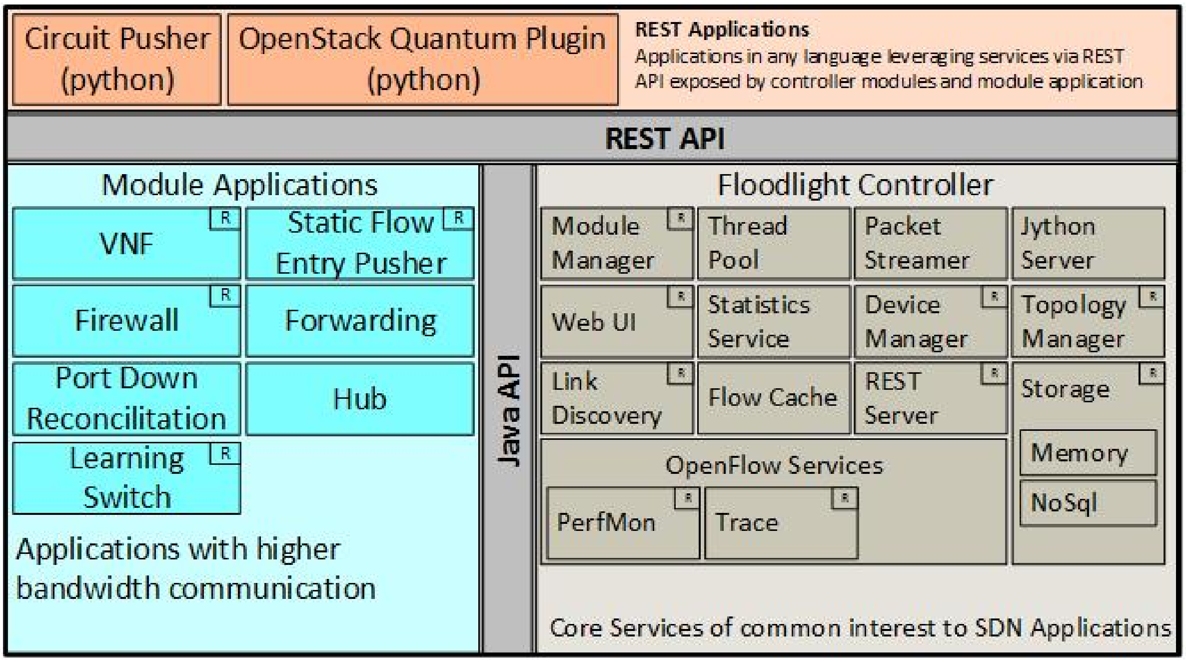


Figure 3: Architecture of the Floodlight SDN Controller.

Table 1: Comparison of QoS capabilities of OpenFlow product families of different vendors.

	Pica8	Juniper	Brocade	HP
Queues	Yes	Yes	Yes	No
Metering	Yes	Yes	Yes	Yes
Queue Statistics	No	No	Not in all products	No
Port Statistic	Yes	Yes	Yes	Yes

6 CONCLUSIONS

There are inconsistencies between features of the OpenFlow specification and the vendor implementations of hardware switches (see Table 1). Depending on the vendor this gap can cause interoperability problems. The ONF specifies new OpenFlow versions every year, which results in outdated hardware. Most OpenFlow features are implemented in the application specific integrated circuit (ASIC) of a switch. In case of a new performance demanding OpenFlow functionality (e.g. queueing), new ASICs and switches have to be developed and deployed. This results in skipping certain version and features. An alternative way to implement new features in existing hardware is the use of P4 [9]. This is a domain specific language which enables software-based packet processing. P4 programs are compiled into hardware. This turns the static switch ASIC into a chip which can be dynamically reprogrammed after it has been deployed as part of a switch. This can be compared to a Field Programmable Gate Array (FPGA). By using P4, vendors can update their hardware after it has been deployed. The proposed concept shows that the routing behaviour of an OpenFlow network can be changed based on the utilisation of the hardware interfaces. A quantitative evaluation of the application is not within the scope of this paper. The proof-of-concept implementation uses the current bandwidth utilisation to determine the quality of a given path. In future more parameters and models should be evaluated in order to provide guaranteed QoS. This is indispensable in the context of Industry 4.0 and wide area SDN networks.

ACKNOWLEDGMENTS

The research presented in this paper is funded by the BMWi (Bundesministerium für Wirtschaft und Energie) within the ZIM-Program (Zentrales Innovationsprogramm Mittelstand), project INAASCA (Integrated Network as a Service Solution as Part of Cloud IT Application Portfolio) [14]. Additionally this work was supported by the Ministry for Science and Culture of Lower Saxony as part of SecuRIn (VWZN3224), which is funded by the funding initiative "Niedersächsisches Vorab" of Lower Saxony.

REFERENCES

[1] Open Networking Foundation, „OpenFlow Switch Specification Version 1.3.5 (Protocol version 0x04)," [Online] Available: https://3vf60mmveq1g8vzn48q2o71a-wpengine.netdna-ssl.com/wp-content/uploads/2014/10/openflow-switch-v1.3.5.pdf.

[2] A. S. Tanenbaum und D. Wetherall, Computer networks, Prentice hall, 1996.

[3] J. Ohms, O. Gebauer, N. Kotelnikova, D. Wermser und E. Siemens, "Providing of QoS-Enabled Flows in SDN Exemplified by VoIP Traffic," 5th International Conference on Applied Innovations in IT, 2017.

[4] J. W. Guck, A. van Bemten, M. Reisslein und W. Kellerer, "Unicast QoS routing algorithms for SDN," IEEE Communications Surveys & Tutorials, 2017.

[5] H. Zhang und J. Yan, "Performance of SDN routing in comparison with legacy routing protocols," In Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2015.

[6] Y. Jinyao, Z. Hailong, S. Qianjun, L. Bo und G. Xiao, "HiQoS An SDN-based multipath QoS solution ," China Communications, Bd. 12, Nr. 5, p. 123–133, 2015.

[7] J. L. Valenzuela, A. Monleon, I. San Esteban, M. Portoles und O. Sallent, "A hierarchical token bucket algorithm to enhance QoS," in IEEE 802.11, In Vehicular Technology Conference, VTC2004-Fall. 2004 IEEE 60th (Vol. 4, pp. 2659-2662), 2004.

[8] Project Floodlight, "Floodlight OpenFlow Controller," [Online] Available: http://www.projectfloodlight.org/floodlight/.

[9] P. Bosshart, D. Daly, G. Gibb, M. Izzard, N. McKeown, J. Rexford, C. Schlesinger, D. Talayco, A. Vahdat und G. Varghese, "P4 Programming protocol-independent packet processors," ACM SIGCOMM Computer Communication Review, Bd. 44, Nr. 3, p. 87–95, 2014.

[10] Hewlett Packard, "HP Switch Software OpenFlow Administrator's Guide K/KA/WB 15.1," [Online] Available: http://h20628.www2.hp.com/km-ext/kmcsdirect/emr_na-c03991489-1.pdf.

[11] Brocade, "Brocade NetIron Software Defined Networking (SDN) Configuration Guide," [Online] Available: http://www.brocade.com/cotent/html/en/configuration-guide/NI_05800a_SDN/GUID-4C86703D-AF09-43B0-8DCA-8402D65624B0.html.

[12] Juniper Networks, "OpenFlow Support on Juniper Networks Devices," [Online] Available: https://www.juniper.net/documentation/en_US/release-independent/junos/topics/reference/general/junos-sdn-openflow-supported-platforms.html.

[13] Pica8, "PicOS Open vSwitch Configuration Guide," [Online] Available: http://www.pica8.com/wp-content/uploads/2015/09/v2.9/html/ovs-configuration-guide/.

[14] D. Wermser und O. Gebauer, "NaaS as Business Concept and SDN as Technology – How do They Interrelate," 20. ITG-Fachtagung, 2015.

The External Mechanical Effects on the Value of Additional Losses in the Telecommunicationsfiber Optic Cables Under Operating Conditions

Ali Mekhtiyev[1,2], Felix Bulatbayev[1], Yelena Neshina[2], Eduard Siemens[3],
Aliya Alkina[2] and Tatyana Shaigarayeva[1]
[1]*Faculty of Power Engineering, Automation and Telecommunications,*
Karaganda State Technical University, Mira, Karaganda, Kazakhstan
[2]*Engineering School of Nondestructive Testing and Safety, Lenina, Tomsk. Russia*
[3]*Future Internet Lab Anhalt, Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Köthen, Germany*
barton.kz@mail.ru, felix4965@mail.ru, 1_neg@mail.ru, eduard.siemens@hs-anhalt.de,
alika_1308@mail.ru, shaigaraeva@mail.ru

Keywords: Fiber Optic Cable, Reflectometer (OTDR), Data Loss, Signal Strength.

Abstract: This article deals with the impact of optical fiber bends on the level of optical losses. The main task of this research was to carry out a series of experiments using the optical time domain reflectometer (OTDR VISA) under field operating conditions for obtaining of optical loss parameter’s dependencies from different bending radii of optical fibers and from the fibers number. We have studied the processes of the occurrence of optical losses in the investigated object and elaborated deterministic models of the physics of the phenomena along with the description undelraying physical processes during the bending. As the outcome of the experiment, a number of dependencies of the optical losses intensity from the number of bends in the fiber, the radii and the wavelength is elaborated. A methodology of the calculation of the signal strength from the number of bends is given. During the work, the excessive losses were noted in case of non-compliance with the installation technology and operating rules. The results of cable research are adopted by Kazakhtelecom JSC.

1 INTRODUCTION

During the operation of the fiber optic cables mechanical loads on the cables occur, which lead to bends of the optical waveguide which in turn causes the signal attenuation and optical losses reducing the performance of the telecommunications network. Hereby the quality of the services provided to customers of Kazakhtelecom JSC is the most important criteria for the company and a special attention is payed to determination of that..

When laying fiber-optic communication cables different situations occur, which affect the technical characteristics of the fiber optic cable and the transport capacity of the telecommunications channels. The most important impact on the transport capacity has been proved to be the influence of bending parameters of optical fibers and the occurrence of losses of radiated power of the signal source in the fiber optic cable. As pointed out above, bends of a fiber-optic conductor lead inevitably to energy losses which in practice exceed the values given in the manufacturer's specifications. The loss data at such a bend radius is especially large. The signal attenuation values in optic fibers reach critical values, when the fiber optic conductor is on the verge of mechanical damage. The critical radius can be as small as a few millimeters for fibers with high numerical aperture, whereas the minimum allowable bend radius is much larger - often tens of centimetres - for single-mode fibers with large transverse mode area.

Usually the energy losses increase during the optical fiber deformation at longer wavelengths. The interference of light reflected from the cladding of the coating determines the dependency of the losses from the wavelength. The increase in fiber losses on fibre bends at long wavelengths limits the transmission range of single-mode fibers [1].

Our work was carried out of the request of Kazakhtelecom JSC and is of applied nature, whereby cables of various manufacturers are used for tests and investigations, while installation is performed by several contract organizations that are not connected by common management of operations. At the same time, there is a possibility that not all technological requirements are met during assembly work.

The main task was to carry out a series of experiments using optical time domain reflectometer (OTDR)VISAunder actual operating conditions for obtaining the multifactor dependencies with different bend radii of optical fibers and their number. The results will be used to adjust the monitoring system of Kazakhtelecom JSC.

2 RELATED WORK

A scientific analysis of similar research performed by foreign authors that work with optical fibers and develop fiber-optic sensors has been performed in [2] - [10]. Most of that relates to the use of fiber to build sensors for the mining industry.

The author's experience and research methods have been taken into account, which allowed us to choose the direction of work and achieve the originality of the research, since our task was to solve actual problems associated with the transmission of optical signals in telecommunications systems. In recent research, we have elaborated methods of investigation of optical power losses and result processing in [11].

3 THE PERFORMANCE OF THE EXPERIMENT

The accuracy of the measurements made by the OTDR meter is determined in the same way as by optical wattmeters and light-sensitive detectors of other kinds [12]. The accuracy of any optical meter depends on how close the output power of the electrical signal corresponds to the input power of the optical radiation. Most optical meters convert the incoming power of optical radiation evenly over the entire operating range into an electrical signal of the appropriate level, but the output power of the electrical signal is extremely low. During the research we used OTDR VISA (Figure 1) which is an universal measuring system for equipping the trunk stations and operating the new FTTx access technologies (PON, etc.) operating at two wavelengths of 1310 nm and 1550 nm with an additional option of the optical power meter. OTDR VISA is used in the production of fiber and fiber optic cable, the construction of fiber optic communication lines (FOCL), diagnostics and maintenance of the state of fibers, cables and fiber optic communication lines.

Figure 1: Type of VISA OTDR.

A series of experiments was conducted to determine the dependency of the additional losses on the number of bends and the bend radius of the fiber optic conductor which can arise during its operation under normal operation conditions. A single-mode G-652 fiber with a wavelength of 1310-1550 nm was investigated.

The following initial conditions were adopted for the investigations:

- air temperature is 20 C,
- number of bends is from 1 to 10 in steps of 1,
- bend radius is in the range from 3 to 18 mm,
- length of the fiber optic cable under test is 71 m.

The blind spot of the OTDRvisa is no more than 3 m (according to the passport), the measurement method used is the measurement of optical losses on the optic fiber with an increaseed number of bends with different bend radii during the cable installation into the ground. The OTDR characteristics are: the refraction index (n) -1.4683, back-scattering coefficient (BC) -78 dB, allowed interval (dL) - 0.43 m, and the pulse length (Tp) - 10 ns [12].

Analysis parameters are: attenuation in the connections - 0.2 dB, reflection coefficient - 40 dB, fiber end - 3 dB, inner losses within the device - 0.002. During the research more than 30 types of fiber optic cable which was in operation by Kazakhtelecom JSC were examined, however three of them has been used for presenting in this work.

We have studied the processes associated with the occurrence of optical losses in the investigated object and developed deterministic models physical phenomena and the underlaying physical elementary processes in the fibers. The performed tasks where: the following: the determination (refinement) of the model parameters; d determination of the model parameters, reduction of the number of iterations. The method is based on the development results of M. M. Protodyakonov [13]. It allows working with data of uncontrolled and controlled experiments. The method is universal both in the field of application and in the variants of the model construction: the models can have the form of sum, products of partial dependencies, their combinations, with sequential neutralization of the influence of priority arguments or without it. An important advantage of the method is the possibility to estimate with an appropriate level of reliability the influence of those arguments that will be excluded from analysis in traditional methods as not significantly affecting the objective function (for example, by Fisher criterion). The results of our measurement results, repeated 10 times each, and the calculated approximations with exponential functionsxxx are presented by the dependency graphs below, whereby the parameters of the exponential functions are given for each graph, along with the correlation coefficient R between the measurement results and the respective approximations.

The measured fiber optic cable isSNR-UT-12, which is a very robust one and designed for laying intra-district and inter-building networks of cable providers of communication or cable TV. The technical parameters of the cable are shown in Table 1. Figure 2 shows the measurements results and their average loss values which were determined by means of the optical reflectometer. According to the average loss value it was constructed a dependency graph on the number of bends from the bend radius. We plotted the dependency curves of the additional losses on the number of bends at different values of the bend radii of the optical fiber. The value of the radius r for the curve is given for all three graphs as follows: 1: 3 mm, 2: 5 mm, 3: 10 mm, 4: 12 mm, 5: 15 mm, 6: 18 mm. For each radius two lines are given – a solid line with measurement points and a dashed line with the approximation, described above.

Table 1: Descriptions of SNR-UT-12fiber optic cable.

Number of optical fibers, pcs.	Attenuation coefficient, dB/km, acc. ITU-T:		Allowabl e tension strain, kN	Tem perat ure rang e, °C	Oute r diam eter, mm	Weig ht of 1 km of cable , kg
	G.652	G.651				
2-154	0,339	0,206	1,3...2,55	-40 ... +60	6	217

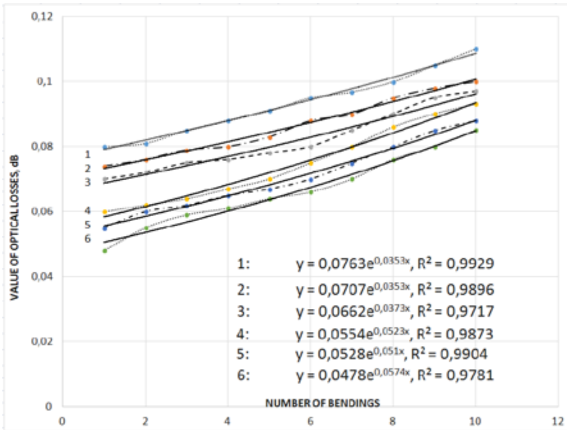


Figure 2: The value of the optical lossess of SNR-UT-12 fiber optic cable. The value of the radius r for the curves: 1: 3 mm, 2:5 mm, 3: 10 mm, 4: 12 mm, 5: 15 mm, 6-; 8 mm.

Futher the fiber optic cable M6A24, which is designed for data transmission over long distances and a high number of channels, has been analyzed. For this, the cable has to have low attenuation and dispersion and high data throughput. A single-mode fiber is used with core and shell dimensions of 8/125 μm. The wavelength is 1.3 ... 1.55 μm. The technical characteristics of the cable are presented in Table 2. Figure 3 shows the measurements results and their average loss value which was determined by means of the optical reflectometer. According to the average loss value a dependency graph has been compiled on the number of bends in dependence from the bend radius.

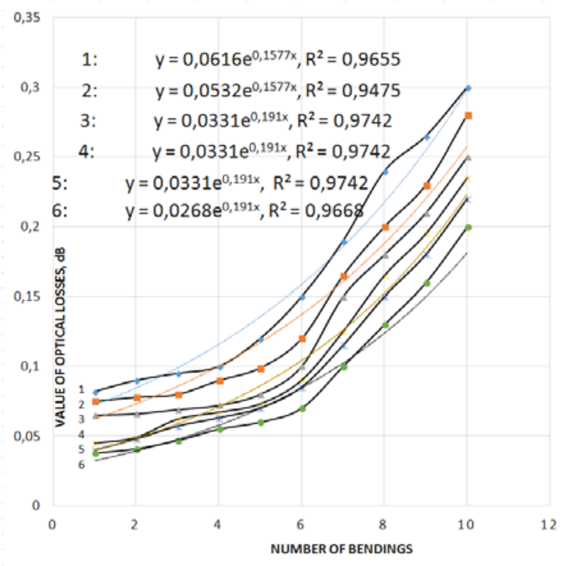


Figure 3: Value of additional losses of M6A24 fiber optic cable. The value of the radius r for the curve: 1-3 mm, 2: 5 mm, 3: 10 mm, 4: 12 mm, 5: 15 mm, 6: 18 mm.

Table 2: Descriptions of M6A24 fiber optic cable.

Number of optical fibers, pcs.	The attenuation coefficient, dB/km		Allowab letensio nstrain, kN	Temp eratur e range , °C	Exter nal diam eter, mm	Weig ht of 1 km of cable , kg
	G.652	G.651				
2-156	0,20	0,67	1,3 ... 2,55	-40 ... +60	14,9	215

In a next series, the OKLSM-12-FF fiber optic cable, designed for laying in cable channels, pipes, blocks, collectors has been investigated. The technical characteristics of the cable are shown in Table 3. Figure 4 shows the measurement results and the average loss values. According to the average loss value a dependency graph on the number of bends from the bend radius has been conducted.

Table 3: Descriptions of OKLSM-12-FF fiber optic cable.

Number of optical fibers, pcs.	The attenuation coefficient, dB/km		Allowab letensio nstrain, kN	Temp eratur e range , °C	Exter nal diam eter, mm	Weig ht of 1 km of cable , kg
	G.652	G.651				
2-144	0,22	0,7	1,5...3,5	-40 ... +60	14,4	194

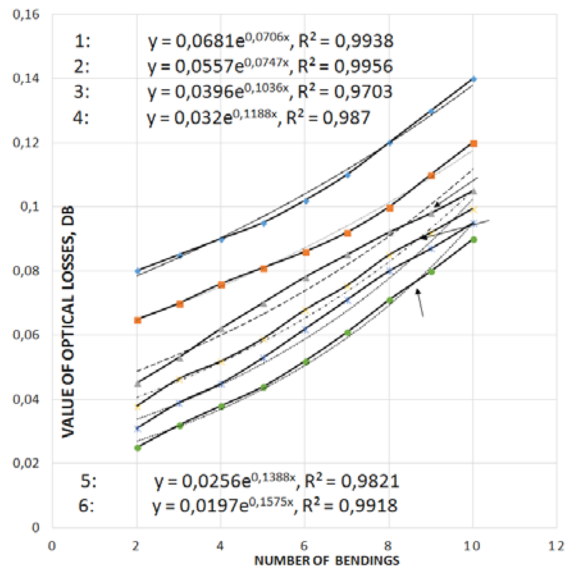


Figure 4: Value of additional losses of OKLSM-12-FF fiber optic cable. The value of the radius r for the curve: 1: 3 mm, 2: 5 mm, 3: 10 mm, 4: 12 mm, 5: 15 mm, 6: 18 mm

The performed tests and measurements have shown that under normal operating conditions not all technical characteristics of the manufacturers correspond to the claimed ones. Additional optical losses arising from mechanical effect on them may be higher than the normative ones, even if the manufacturer indicates that the cable is not susceptible to mechanical bends and deformations.

Based on the research, the recommendations on the selection of fiber optic cables for use at Kazakhtelecom JSC have been made.

4 CONCLUSIONS

During the experiments, a number of dependencies of the increase in additional losses are established with an increase in the number of fiber bends and a decrease in the bend radius, and also with an increase in the wavelength of the optical range of the light source. This analysis is new and is not yet known to be performed by other research groups. The results allow to make ajustments for more accurate results of reports of telecommunication devices sent to the central server in the monitoring system to assess the overall transmission quality situation. In the event of an increase in optical losses, the information is sent to the operator desk for making appropriate decisions.

During the work, excessive losses were noted in case of non-compliance with the installation technology and operating rules. Recommendations for elimination of points of massive optical losses were made which were sent to the management of this telecommunications company to eliminate the issues. The database has been gathered and put on the server in the monitoring system to assess the quality of the telecommunications network. This research was conducted with the participation of teachers, undergraduates and representatives of the technical service of Kazakhtelecom JSC.

REFERENCES

[1] V.T. Vorob'ev, R.S. Avramenko, V.N. Kozlov, «Research Of Of Signal Attenuation In Fiber-Optic Cable At Cable Line Nonlinear Lining», Vestnik Kostromskogo Gosudarstvennogo universiteta im. N.A. Nekrasova, vol. 3, pp. 8-10, 2011.

[2] V.K. Kryzhanovskii, V.V. Burlov, A.D. Panimatchenko, Iu.V. Kryzhanovskaia, Technical properties of polymeric materials, Training manual, Sp-B, 2005.

[3] Kumar V. Fiber optic methane and strain sensors for mines. – Photonics (ICP) International Conference on. 5-7 July 2010, pp.79–84. doi: 10.1109/ICP.2010.5604439

[4] Kim S.T., Park Y.-H., Park S.Y., Cho K., Cho J.-R. A sensor-type PC strand with an embedded FBG sensor for monitoring prestress forces. – Sensors (Switzerland), 2015, v.15,No 1, pp.1060–1070. doi: 10.3390/s150101060.

[5] Liu T., Wei Y., Song G., Li Y., Wang J., Ning Y., Lu Y. Advances of optical fiber sensors for coal mine safety monitoring applications. – Proceedings of the 2013 International Conference on Microwave and Photonics, 2013, pp.102–111.

[6] Yiming Zhao, Nong Zhang,1 and Guangyao Si. A fiber bragg grating-based monitoring system for roof safety control in underground coal mining. – Sensors (Switzerland), 2016, v. 16, No10, pp. 112-117. doi: 10.3390/s16101759.

[7] Chunde Piao, Jun Yuan, Bin Shi, Haijun Lu, Guangqing Wei, Chunsheng Gu Application of distributed optical fiber sensing technology in the anomaly detection of shaft lining in grouting. – Journal of Sensors, 2015, v.2015, pp. 163–169.

[8] Liu X., Wang C., Liu T., Wei Y., Lv J., Fiber Grating Water pressure sensor and system for mine. – ACTA Photonica Sinica, 2009, v. 38, pp. 112–114.

[9] Kumar Atul, Kumar Dheeraj, Singh U.K., Gupta P.S., Shankar Gauri Optimizing fibre optics for coal mine automation. – International Journal of Control and Automation, 2011, v.3, pp. 63–70.

[10] Naruse H., Uehara H., Deguchi T., Fujihashi K., Onishi M., Espinoza R., Pinto M. Application of a distributed fibre optic strain sensing system to monitoring changes in the state of an underground mine. – Measurement Science and Technology, 2007, v.18, No 10, pp. 3202–3210. doi: 10.1088/0957-0233/18/10/S23.

[11] A.V. Yurchenko, N.I. Gorlov, A.D. Alkina, A.D. Mekhtiev, A.A. Kovtun, «Research of the additional losses occurring in optical fiber at its multiple bends in the range waves 1310nm, 1550nm and 1625nm long», Journal of Physics: Conference Series, vol. 671, 2016.

[12] I.A. Ovchinnikova, P.A. Semenov, «Investigations of the influence of external factors on the elements of optical cable designs // Cables and wires, vol. 3, pp. 8-9, 2009.

[13] M.M. Protod'iakonov, The method of rational planning of experiments, Moscow: Publishing house "Science", 1970.

Efficiency of a PID-based Congestion Control for High-Speed IP-Networks

Nikolai Mareev[1], Dmitry Kachan[1], Kirill Karpov[1], Dmytro Syzov[1],
Eduard Siemens[1] and Yurii Babich[2]
[1]Department of Electrical, Mechanical and Industrial Engineering,
Anhalt University of Applied Sciences, Bernburger Str. 55, 06366 Köthen, Germany,
[2]Department of Telecommunication Networks,
Odessa National O.S. Popov Academy of Telecommunications, Odessa, Ukraine
nikolai.mareev@student.hs-anhalt.de, {dmitry.kachan, kirill.karpov, dmytro.syzov, eduard.siemens}@hs-anhalt.de,
y.babich@onat.edu.ua

Keywords: Congestion Control, Transport Protocol, Ip Networks, Pid Control, Rmdt, High-Speed Networks

Abstract: The current situation in IP networks shows the need for new congestion control algorithms that can be flexible, scalable, and capable of avoiding additional queue delays caused by loading the bottleneck buffers. Most common internet flows use loss-based congestion controls, which can achieve high bottleneck bandwidth utilization and fair resource sharing but cause overload bottleneck buffers. In this paper we present an investigation of the performance of a PID-based congestion control solution for high-speed IP networks. It uses measurements of a round trip time and receiver delivery rate to reach and keep maximum available bottleneck performance and constant node buffer load caused by bottleneck queue on some low level. This algorithm can be effective in high-speed IP networks and delay-sensitive applications. It is designed to be flexible and scalable for different connectivity cases. This algorithm then is investigated on the example of RMDT.

1 INTRODUCTION

Network congestion occurs, when a receiving node is receiving more data than it can handle or forward to an output interface. It leads to a significant performance degradation: additional delays and massive packet losses. Congestion control algorithms are aimed to solve such problems. This is an automatic control of a sender's parameters, which describe performance of data send process, adaptability for different connection cases and the ability to share link resources fairly with other connections.

The aim of this paper is to present an investigation on PID-based congestion control in terms of CloudBDT and BitBooster projects. These projects use the Reliable Multi-Destination Transport protocol RMDT [1][2], where the results of the present work may be used.

The idea behind the usage of PID (Proportional – Integral – Derivative) control in congestion control algorithm lies in the fact that this type of control can be very flexible, scalable and adaptive. It can be easily extended by additional modules like auto tune loop or artificial neural network.

The main challenges for modern congestion control are: high bottleneck bandwidth utilization, low bottleneck queue delays, automatic scalability to different channel conditions (different bandwidth and different delays) [3], adaptation for sudden changes over connection like rerouting, applicability in wireless networks and resource sharing.

The rest of this paper is organized as follows: In section 2 a short overview of modern congestion control solutions and their main disadvantages is presented. Section 3 describes main states of PID-based congestion control solution and its principles. The experimental setup is presented in Section 4. Test results and evaluations are given in Section 5. Section

6 includes conclusions based on the evaluation results, further work and describes benefits of such solutions.

2 RELATED WORK

In [4] various approaches to TCP host-to-host congestion control algorithms and its evolution due to modern network sharing issues have been reviewed.

Loss based congestion control algorithms (Reno, Cubic) interprets packet losses as an indicator of a congestion. TCP Cubic [5] (which is the default congestion control in Linux kernels 2.6.19 and above.) can be very effective - with high bottleneck capacity utilization and fair resource sharing. However, they cause significant bottleneck queue delays and performance degradation in cases with tiny bottleneck queue buffers. Moreover, packet losses can be caused not only by a congestion in a network, but also by a link itself as well (e.g. wireless connections).

Another important solution is a delay-based congestion control, like TCP Vegas described in [6] and its future improvements. It is a proactive algorithm that uses bottleneck queue delay and packet losses as congestion indicator. Such strategy allows to predict a congestion before losses occur, caused by bottleneck buffer overload happen and also to keep queue delays on the levels, lower than the loss-based algorithms. Anyway, those algorithms have no aim to keep bottleneck buffer load at a low level, they keep it at some constant level. The most significant disadvantage of such algorithms is an unfair network resource sharing – especially with loss-based congestion control algorithms [7]. In addition, use of packet losses as a secondary congestion indicator can lead to the same problem of non-congestion caused losses as with pure loss-based algorithms.

BBR [8] algorithm (Bottleneck Bandwidth and Round-trip propagation time) is a new solution in congestion control. It uses round trip time and bottleneck bandwidth probing cycle to keep bottleneck queue load on a low level along with queueing delays and tries to reach effective bottleneck capacity utilization. Such technique under some conditions leads to a higher performance in comparison to loss-based or "delay-loss-based" algorithms. However, the probing cycle leads to data rate decrease and in some cases to unfair resource sharing [9].

3 PID-BASED CONGESTION CONTROL

A PID controller is a widely used control loop feedback mechanism, it continuously calculates an error value as a difference between a desired level of a controlled value (Setpoint, SP) and a measured process value (PV). It applies a correction based on proportional, integral and derivative terms. In case of this congestion control solution, the process value is round trip time. The correction can be done by changing the send data rate.

The first state of an algorithm is a "*Gain*" state (see figure 1), used to quickly estimate bottleneck bandwidth (BBW). PID congestion control requires presets of main parameters such as round trip time SP and factors for send data rate correction. To estimate that the algorithm has a second state named "*Manage*" state. The third, "*Control*" state is a PID-controller itself. Figure 1 illustrates main states of PID-based congestion control.

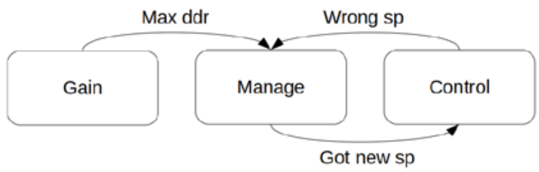


Figure 1: Main states of a PID-based congestion control.

3.1 Gain

Algorithm enters the *Gain* state at the very beginning of the transmission. It tracks the delivered data rate (DDR) and rapidly increases the send data rate. When last three reports show that there is no significant growth of delivered data rate, then congestion control switches to *Manage* state. *Gain* state enables the algorithm to quickly reach a bottleneck bandwidth limit and to make delay measurements for future processing.

3.2 Manage

Manage state tries to get the minimal round trip time (RTT) of a transmission by omitting the bottleneck queue buffer. It is achieved by decreasing data rate by half for 50 ms. It allows to set an acceptable round trip time setpoint (SP). For the current implementation, the acceptable setpoint is:

$$SP = max\ (minRTT + \alpha;\ 1.25\ minRTT), \qquad (1)$$

where α is the minimal growth of SP.

Minimal level value of SP (10 ms in this work) is caused by some instability of RTT measurements in the current solution. Otherwise, it can be even lower.

This state is also useful for resetting a setpoint if sudden rerouting is detected (RTT significantly dropped/raised). The third role of *Manage* state is fairndwidth sharing. For PID based congestion control, fair share is possible if chosen SPs of both links are almost equal.

3.3 Control

In the *Control* state, a modified PID digital controller tracks the delivery data rate and round trip time and tries to keep RTT near a setpoint by changing send data rate. If SP is not reachable for the last 10 packets or DDR has suddenly dropped down (more than 20% DDR drop), algorithm goes to *Manage* state to estimate a new RTT setpoint.

4 EXPERIMENTAL SETUP

Figure 2 shows testbed network topology. All tests have been performed in 40 GE Laboratory of Future Internet Lab Anhalt [10] (FILA).

The core element here is the WAN emulator Netropy 40G [11] that can be used to create an emulation of WAN links up to 40 Gbps throughput and up to 10^6 ms delay. Sender and receiver both run in Ubuntu 16.04 (kernel: GNU/Linux 4.13.0-17-generic x86_64) and are equipped with Intel(R) Xeon(R) CPU E5-2643 v4 3.40GHz, 64GB of RAM and 40000baseSR4/Full supported link modes on Emulex Corporation OneConnect NIC.

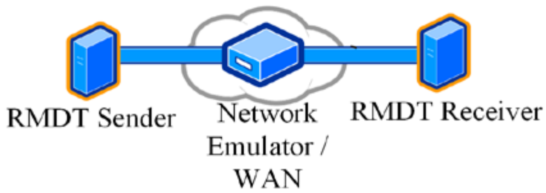


Figure 2: Testbed network topology.

5 EXPERIMENTAL RESULTS

Tests have been performed in next scenarios: single flow test with BBW 10 Gbps and {10, 50, 100, 150, 200}ms RTT; resource sharing test with two flows, bottleneck bandwidth 5 Gbps and 50 ms RTT. Queue management is set to drop-tail in all test. Figures from 3 to 7 demonstrate experimental results of congestion control: high bottleneck bandwidth utilization and queue load level control in high speed IP network with different base RTT (one way delay is one half of RTT). The fair resource sharing of proposed congestion control method is shown in Figure 8. All statistics are collected by WAN Emulator. It can collect only per-second mean statistics, which leads to unclear view of bottleneck buffer load level in the different states on some plots.

Results of first experiment with 10Gbps bottleneck bandwidth and 10 ms RTT are presented in figure 3. Such bottleneck buffer load level deviations (highlighted zones) are caused by the current RTT measurement solution in RMDT ver. 0.97 alpha, however this issue has no significant effect on bottleneck bandwidth utilization or performance of the control itself. 10 Gbps data rate was achieved with 6 Mbytes bottleneck buffer load in the *Control* state.

Second experiment with 10 Gbps bottleneck bandwidth and 50 ms RTT is shown in figure 4. The highlighted zone shows here the bottleneck buffer load during the *Gain* state (rapid growth of send data rate and bottleneck bandwidth estimation). It allows to make measurements of maxRTT and congestion reporting. 10 Gbps data rate was achieved with 13 Mbytes bottleneck buffer load in the *Control* state.

As shown in figure 5, in terms of next experiment with 10 Gbps bottleneck bandwidth and 100 ms RTT.

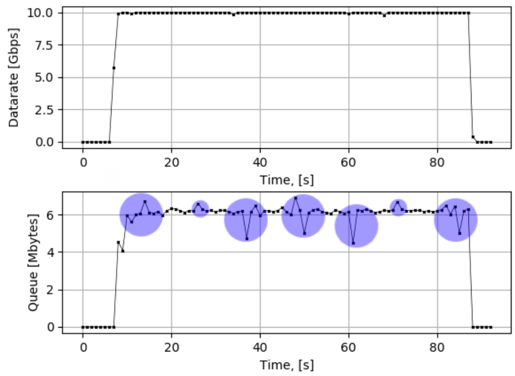


Figure 3: Test 1: BBW 10 Gbps, base RTT 10 ms, single flow.

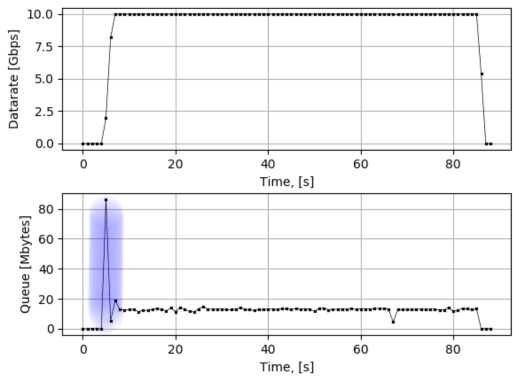


Figure 4: Test 2: BBW 10 Gbps, base RTT 50 ms, single flow.

10 Gbps data rate was achieved with 30 Mbytes bottleneck buffer load in the *Control* state.

Figure 6 demonstrate single flow experiment with 10 Gbps bottleneck bandwidth and 150 ms RTT. IP in high speed networks with large RTT. However, various RTT/ BBW cases require

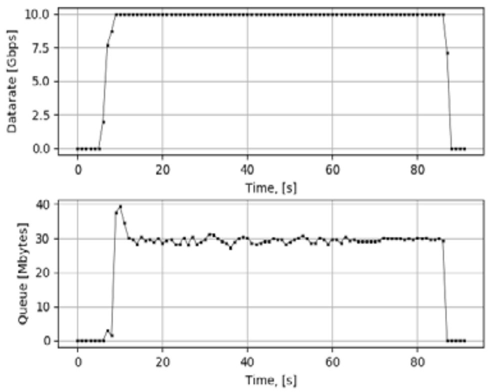


Figure 5: Test 3: BBW 10 Gbps, base RTT 100 ms, single flow.

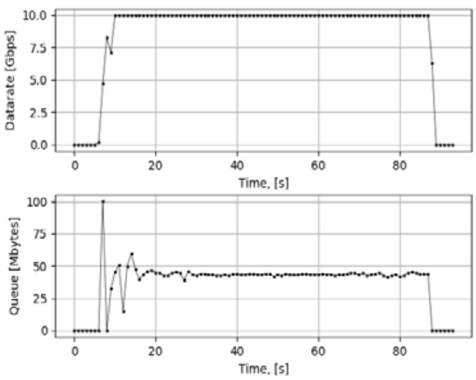


Figure 6: Test 4: BBW 10 Gbps, base RTT 150 ms, single flow.

different presets done by *Manage* state for PID-controller. Higher RTT require more bottleneck buffer memory in a relatively current realization of an algorithm (figures 3-7). 10 Gbps data rate was achieved with 43 Mbytes bottleneck buffer load in the *Control* state.

Last single flow experiment with 10 Gbps bottleneck bandwidth and 200 ms RTT is presented in figure 7. Highlighted zone shows the transition from *Manage* state (drop data rate and omitting bottleneck buffer) to *Control* state (growth of buffer load to SP). 10 Gbps data rate was achieved with 59 Mbytes bottleneck buffer load in the *Control* state.

Figure 8 shows fair network resource sharing test by two PID-based congestion control flows, with lution has following features: it is scalable, keeps bottleneck buffer load on some low level and achieves high throughput with minimal losses. Fair resource sharing is achieved by dynamically network is a system with high transport delays, however PID-based congestion control acts quite effective even

5 Gbps bottleneck bandwidth and 50 ms RTT. Highlighted zone shows overload of bottleneck buffers in the *Gain* state of flow 2. After *Gain* state cross-interference, both flows come to such SP‘s, that result in low bottleneck buffer load. 2.53 Gbps rate was achieved in sharing by flow 1 and 2.46 Gbps by flow 2 (5 Gbps in total). Bottleneck buffer load before interaction is near 14 Mbytes by only flow 1 and 17 Mbytes in sharing by two flows.

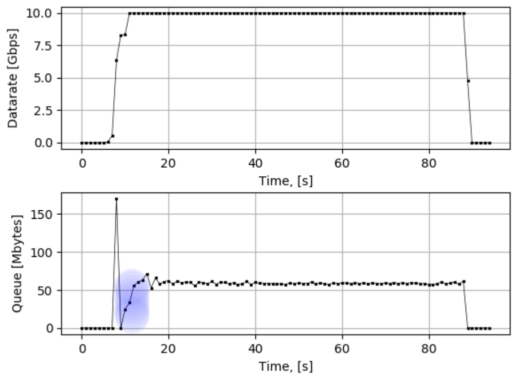


Figure 7: Test 5: BBW 10 Gbps, base RTT 200 ms, single flow.

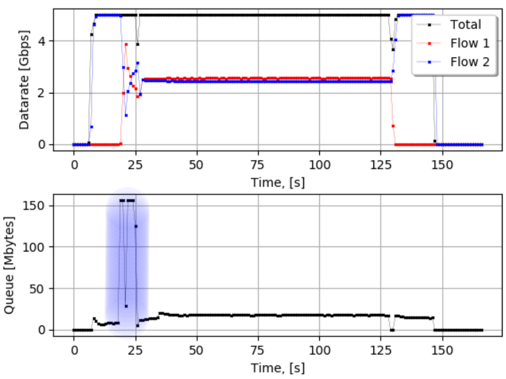


Figure 8: Test 6: BBW 5 Gbps, base RTT 50 ms, two flows in one link.

6 CONCLUSION & FURTHER WORK

This article presents the results of the investigation of a PID-based congestion control solution. Basic tests with UDP-based transport protocol show that proposed so

changing the RTT setpoint within the Manage state of the proposed algorithm. For long fat networks a PID-based congestion control is also can be used, but it requires additional RTT setpoint fitness algorithm, for keeping bottleneck buffers on lower level. It is under active development; next steps include additional auto tune loop for more precise scalability to any bandwidth and any delays; more intelligent setpoint management; fairness with TCP congestion control algorithms; boost performance in wireless networks.

ACKNOWLEDGMENT

This work has been funded by Volkswagen Foundation for trilateral partnership between scholars and scientists from Ukraine, Russia and Germany within the project CloudBDT: Algorithms and Methods for Big Data Transport in Cloud Environments.

REFERENCES

[1] E. Siemens, D. Syzov and D. Kachan, “High-speed UDP Data Transmission with Multithreading and Automatic Resource Allocation” Proc. of: the 4th International Conference on Applied Innovations in IT, (ICAIIT 2016), pp. 51-56, Koethen, 2016

[2] S. Maksymov, D. Kachan and E. Siemens, “Connection Establishment Algorithm for Multi-destination Protocol” Proc. of: the 4th International Conference on Applied Innovations in IT, (ICAIIT 2014), pp. 57-60, Koethen, 2016

[3] Eduard Siemens, Ralf Einhorn, Andreas Aust, Lars Fuerst, "Multi-Gigabit Challenges: Similarities between Scientific Environments and Media Production" Proc. of: The IASTED International Conference on Automation, Control, and Information Technology (ACIT-ICT 2010), Novosibirsk, Russia.

[4] A. Afanasyev, N. Tilley, P. Reiher, and L. Kleinrock, “Host-to-Host Congestion Control for TCP,” Communications Surveys Tutorials, IEEE, vol. 12, no. 3, pp. 304–342, Third 2010.

[5] I. Rhee and L. Xu, “CUBIC: A New TCP-Friendly High-Speed TCP Variant,” Proc. of PFLDnet 2005, February 2005, Lyon, France.

[6] L. S. Brakmo, S. W. O’Malley, and L. L. Peterson, “TCP Vegas: New Techniques for Congestion Detection and Avoidance,” in SIGCOMM ’94. Techniques for Congestion Detection and Avoidance,” in SIGCOMM ’94.

[7] M. Hock, R. Bless, and M. Zitterbart, “Toward Coexistence of Different Congestion Control Mechanisms,” in 2016 IEEE 41st Conference on Local Computer Networks, November 2016, pp. 567–570.

[8] N. Cardwell, Y. Cheng, C. S. Gunn, S. H. Yeganeh, and V. Jacobson, “BBR: Congestion-Based Congestion Control,” ACM Queue, vol. 14, no. 5, pp. 50:20–50:53, Oct. 2016.

[9] M. Hock, R. Bless and M. Zitterbart, “Experimental Evaluation of BBR Congestion Control,” in 2017 IEEE 25th International Conference on Network Protocols (ICNP), Oct 2017.

[10] FILA. Future Internet Lab Anhalt [Online]. Available from: https://fila-lab.de. 2017.11.12

[11] Apposite. Apposite Technologies: Linktropy and Netropy Comparison. [Online]. Available from: http://www.apposite-tech.com/products/index.html. 2017.11.12