Application of decision tables transformations for prototyping knowledge bases in the case of forest fire risk forecasting

Aleksandr Yu. Yurin, Olga A. Nikolaychuk and Nikita O. Dorodnykh

Matrosov Institute for System Dynamics and Control Theory, Siberian Branch of Russian Academy of Sciences (ISDCT SB RAS), Lermontov St. 134, Irkutsk, Russia

Abstract

In this paper, we consider the application of the PESoT technology and a tool (namely, Personal Knowledge Base Designer) for prototyping rule-based knowledge bases by using the automated analysis and transformation of decision tables presented in the CSV format. Created prototypes of knowledge bases designed for intelligent decision-making support when analyzing and forecasting the risk (probability) of forest fires based on information about the forest fire hazard class, weather conditions, and other factors. A description of the main stages of the approach and an illustrative example are presented.

Keywords

Knowledge bases, transformation, decision table, fire risk, forest squares, PESoT, PKBD

1. Introduction

The complexity of creating intelligent systems and their knowledge bases can be reduced with the use of methods and tools based on the paradigm known as End-User Development (EUD), including End-User Programming (EUP) and End-User Software Engineering (EUSE) [1-3]. Visual programming and Model-Driven Development (MDD) are examples of EUD methods that are implemented within the PESoT technology (Prototyping Expert Systems Based on Transformations) [4-7]. The main benefit of these EUD methods and technology applications is reducing the risk of manual coding errors, and reusing conceptual models developed earlier.

One of the tasks that require the use of these methods is the development of an intelligent system in the form of a thematic WPS service to support forecasting the risk of forest fires. This task is a part of the grant No. 075-15-2020-787 of the Ministry of Science and Higher Education of the Russian Federation "Fundamentals, methods and technologies for digital monitoring and forecasting of the environmental situation on the Baikal natural territory" [8].

Two techniques for forming fire risk evaluations in forest quarters (squares) were considered when solving this task:

- The first technique is based on a statistical analysis of information about forest fires for the previous period, taking into account a certain forest quarter and a time interval. So, this technique involves statistical processing of large amounts of data, and the resulting evaluations do not depend on the query conditions;
- The second technique is based on artificial intelligence methods, in particular, rule-based expert systems. This technique involves not only statistical processing of data but also conceptual modeling, data mining to find patterns and their further formalization in the form of logical rules. In this case, the obtained evaluations take into account the query conditions.

In this paper, we consider the application of the PESoT technology for the automated creation of knowledge bases including formalization and code generation tasks in the context of the second technique. So, the main contribution of the paper is a description of the application of this technology when solving the task of decision-making support for forest fire risk forecasting.

The paper is organized as follows. Section 2 presents a background, including the main principles of the PESoT technology, our motivation, and the brief state of the art in the field of forest fire forecasting. Section 3 contains the application including a detailed description of separate steps, and an illustrative example, while Section 4 presents some concluding remarks and future works.

2. Background

Next, let's consider the main principles of the used PESoT technology, as well as the subject domain (forecasting the risk of forest fires).

2.1. PESoT: Prototyping Expert Systems Based on Transformations

The PESoT technology [4-7] implements the model-driven EUD principle and includes methods and tools for prototyping rule-based expert systems and decision-making software components for intelligent systems.

Formally, this technology can be described by the following [4]:

$$MDE^{ES} = \langle MOF, L^{ES}, CIM^{ES}, PIM^{ES}, PSM^{ES}, PDM^{ES}, F^{ES}_{CIM-to-PIM}, F^{ES}_{PIM-to-PSM}, F^{ES}_{PSM-to-CODE} \rangle$$

where L^{ES} is a set of languages and formalisms used for modeling expert systems and components; in our case $L^{ES} = \{UML, CM, DT, CT, RVML\}$ where UML is a Unified Modelling Language; CM is a concept or mind maps formalism; DT is a formalism for the representation of decision tables; CT is a formalism for the representation of canonicalized tables; $RVML$ is a Rule Visual Modeling Language [5];

CIM^{ES} is a computation-independent model for PESoT, in our case, it is a domain model represented with the aid of L^{ES};

PIM^{ES} is a platform-independent model for PESoT, in our case this model represent logical rules in our notation RVML;

PSM^{ES} is a platform-specific model for PESoT, in our case this model takes into account the features of the programming language, we use RVML;

PDM^{ES} is a set of platform description models for PESoT, in our case $PDM^{ES} = \{CLIPS, DROOLS, PHP, PKBD\}$.

$F^{ES}_{CIM-to-PIM}, F^{ES}_{PIM-to-PSM}, F^{ES}_{PSM-to-CODE}$ are the rules for model transformations.

The process of creating prototypes of knowledge bases and expert systems is represented by the sequence (a chain) of the following stages: building domain models, building platform-independent models, building platform-specific models, generating source codes and specifications, testing, and integration.

A detailed description of the technology is given in [4-6]. Below, a detailed description of the main stages of its application is considered.

2.2. Motivation

This project includes solving a set of tasks, one of which is the development of thematic WPS services for digital monitoring, analysis, modeling, and forecasting of the environmental situation, as well as the risk of natural and technogenic fires.

Researches are being conducted to solve this problem that is linked to:

- Collecting, cleaning, and analyzing forest fire data and identifying specific heuristics. These heuristics allow one to identify the territorial risk of forest fire hazard that will be considered as fire risk for forest quarter (aka dacha or forestry);
- Developing a knowledge base for analyzing and forecasting the risk (hazard) of fires based on information about the forest fire hazard class, weather conditions, and other factors.

2.3. State of art: Forecasting the risk of forest fires

The topic of the automated formation of knowledge bases for intelligence decision-making support in forecasting the risk (danger) of fires in forest quarters is poorly represented in the scientific literature. There are some works related to a model transformation and the application of a model-based approach (e.g., MDD), as well as some papers aimed at forecasting the risk (hazard) of forest fires.

The first group of works is considered in more detail in [4-6].

Several following works that illustrate the main areas of research that can be selected for the second group:

- The identification of factors affecting the fire hazard of forests [9-10], including a height, a slope, a topographic humidity index, a distance from urban areas, an average annual temperature, land use, a distance from roads, average annual precipitation, a distance to rivers, air temperature (average daily and maximum), dates of the transition of average daily temperatures through threshold limits, dates of onset and descent of a stable snow cover, a relative humidity (average daily and minimum), an air humidity deficit, a count of days with relative humidity < 30% in one of the observation periods for a certain period, an annual precipitation regime, a count of days with rain, a dryness index, a wind regime, a count of days with thunderstorms, etc.;
- The improvement of the scale for assessing forest fire hazard classes depending on weather conditions [11] to account new factors, in particular, humidity indicators [12], or taking into account regional features [13];
- The use of existing methods for assessing fire risk in different regions [14].

In the context of this study, the above works were used to analyze the subject domain and identify factors affecting the assessment of forest fire risk.

3. Decision tables transformations for prototyping knowledge bases

The PESoT technology provides the automated formation of knowledge bases by using various information sources, including conceptual models and tables of different types. One of the tabular forms supported by this technology is a specialized form of decision tables described below.

3.1. A specialized form of decision tables

The specialized form of decision tables is an extension of the standard one [15] and consists of columns and rows. Columns represent names of independent and dependent properties (components or parts of rules), and rows represent specific rules. At the same time, table cells contain values of their properties. The tabular form used in this work has some features. In particular, these features were determined by further automated processing of tables in the context of knowledge engineering.

The main features of our tabular form:

- A table may contain a column with rule names; it must be the first and have the "Rule Name" name;
- Headers of dependent columns are marked with the "#" symbol;
- A column header name can be compound, indicating an entity name (or a class name) and its property name separated by "::" string.
- There are no restrictions on values in cells, i.e. they can not only consist of a set of values {yes, no} as, for example, in [16]. So, values in cells can contain specific arbitrary values and not

only values that indicate the presence or absence of certain property (component) in a rule structure.

A specialized decision table fragment is presented in Figure 1.

	1	2	3	4	5	6	7	8	9
1	RuleName	Risk::grade	Risk::kind	...	Flood-haz.	...	#Conclusion::grade	#Conclusion::text	#Conclusion::cf
2	Risk+Flood hazard->	low	natural		low		low	safe situation	0.9

Figure 1: An example of a specialized decision table fragment.

This form of decision table provides the generation of logical rules of the "IF-THEN" type. In particular, an example of a decision table fragment presented in Figure 1 is interpreted as follows: **IF** there is a "Risk" of a certain "grade" and "kind", and a «Flood hazard» of a certain "level" and "probability", **THEN** some "Conclusion" with a certain "text" and "cf" (a certainty factor) is made. Accordingly, values for properties "grade", "kind" etc. are taken from the cells of a certain row.

The advantages of this form of representation of source data are the following:

- It is the most popular way to represent logical rules for non-programming end-users;
- It is used to represent the results of data mining, for example, when using the Deductor Studio or Loginom system (Base Group company);
- It provides the ability to use publicly available and widespread software, such as Microsoft Excel for generating data, and then saving them in the CSV format;
- There are examples of using this form when solving various practical tasks [17-19].

3.2. Main stages

The development process with the aid of PESoT and the PKBD (Personal Knowledge Base Designer) tool is similar to other PESoT cases [17] and can be presented in the form of the following scheme (Figure 2).
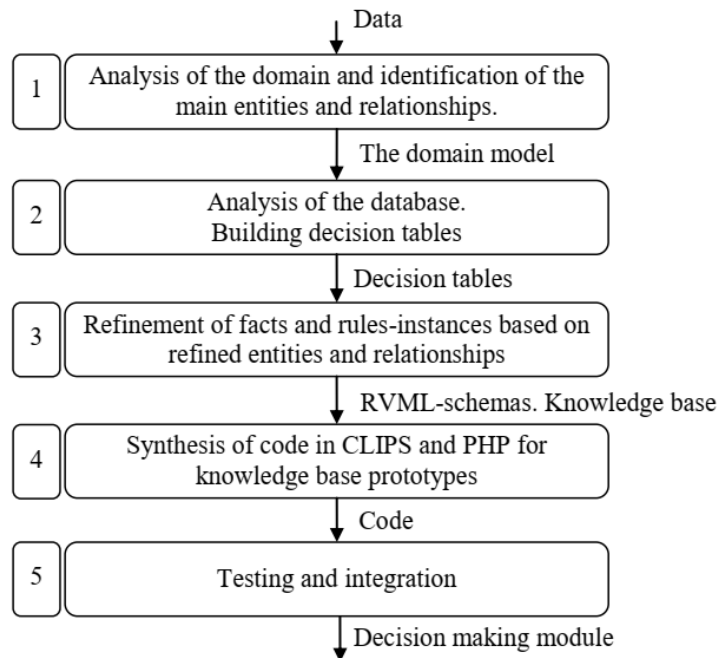

Figure 2: Knowledge base development using PKBD.

In the current case, stage 2 has the greatest computational complexity. This stage is associated with data analysis and the formation of decision tables. Let's consider the stages of this approach using an illustrative example in more detail.

3.3. An illustrative example

The task of prototyping knowledge bases for determining the risk of forest fires is considered an illustrative example. Information on forest fires in the Baikal natural territory for 2017-2020 years, weather data, as well as information on infrastructure (roads, settlements, etc.) and the type of vegetation were used as initial data. The database on fires includes more than 45 000 records describing information about heat points identified as a result of the analysis of satellite images.

Important and computationally complex tasks associated with the preparation of this data for the development of knowledge bases are the following:

- Grouping (aggregating) information about fires with the definition of duration, minimum and maximum area of a certain fire;
- Determining fires located within the boundaries of industrial zones, settlements, and mining zones that are not natural fires;
- Determining fire statistics for certain classes of fire hazard based on forest plans of districts of the Irkutsk region (taking into account the structure of the forestry, aka plots (dachas) or quarters);
- Determining a set of independent factors influencing the risk of fire hazard in a forest district;
- Calculating factor values affecting the risk of the fire hazard of a forest district and their transformation to interval or fuzzy form;
- Determining the risk (hazard) of a forest fire based on the current values of a complex of factors through a certain class of fire hazard and its statistics.

These tasks will be considered in more detail in other works. In this paper, the process of building knowledge bases based on the analysis of decision tables is considered, and we assume that the data preprocessing has already been completed.

In this case, the knowledge base consists of two segments and solving the following subtasks:

1. Forming a conclusion on the forest fire risk of a certain forest area based on the average monthly weather data, current weather conditions, information about the time of year, the proximity of rivers, lakes, roads, settlements, terrain, and vegetation type;

2. Forming a conclusion on the risk of a forest fire according to the fire hazard class of a certain forest area using fire statistics and information about the time of year (season).
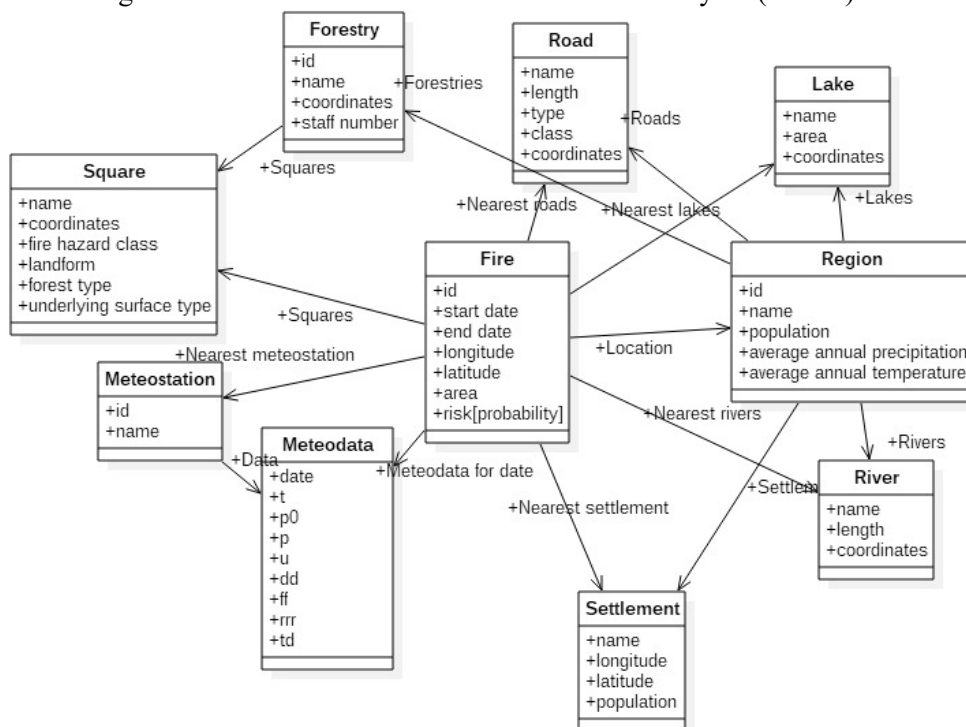

Figure 3: A fragment of the domain conceptual model.

Next, we will consider the stages in more detail.

Stage 1. As a model of the domain, a conceptual model was created that describes the factors affecting the class of the fire hazard of a forest area and the risk (probability) of a fire. A fragment of this model is shown in Figure 3.

Stage 2. Next, decision tables were developed that describe the structural aspect of the domain. These tables contain information about combinations of features describing the fire hazard class of a forest area and the risk (hazard) of a forest fire.

In particular, the following table structure (headers) is used to define the hazard class: `Road::distance_to_car_road,` `Road::distance_to_railway,` `River::distance_to_river,` `Lake::distance_to_lake, Meteodata::rrr, Meteodata::ff, Meteodata::u, Meteodata::t,` `Settlement::distance_to_settlement,` `Settlement::population,` `Region::population,` `Region::average_annual_temperature,` `Season::name,` `Forestry::staff_number,` `Square::landform,` `Square::forest_type,` `Square::underlying_surface_type,` `#Square::name, #Square::fire_hazard_class.`

To determine the risk (hazard) of a forest fire, the following table structure (headers) is used: `Square::name, Square::fire_hazard_class, Season::name, #Fire::risk[probability].`

Further, the obtained rules are analyzed to define the indicators of the frequency of their appearance in the analyzed data. In particular, the support and confidence of the rules are determined. The confidence is used as a certainty factor of a rule.

The decision table with intermediate data is shown in Figure 4.

	C	D	E	F	G	H	J	L	N	P	X
	Fire::dt	Fire::season	Fire::lat	Fire::lon	Fire::poly	Fire::area	Fire::distance_to_car_road	Fire::distance_to_railway	Fire::distance_to_river	Fire::distance_to_l	#forest_hazard_classes
	01.04.2017 08:01	april	51.6256	102.23	010600002C	2.421	0.0 - 0.100	0.400 - 0.500	0.000 - 0.050	0.100 - 0.200	['3']
	01.04.2017 08:02	april	57.9885	106.806	010600002C	2.065	0.200 - 0.300	0.200 - 0.300	0.000 - 0.050	0.300 - 0.400	['3']
	21.08.2017 00:59	august	58.8324	106.5	010600002C	10.887	0.400 - 0.500	0.400 - 0.500	0.000 - 0.050	0.300 - 0.400	['4', '3']
	01.09.2017 12:08	september	58.8261	106.499	010600002C	17.987	0.400 - 0.500	0.400 - 0.500	0.000 - 0.050	0.300 - 0.400	['4', '3']

Figure 4: A fragment of the decision table with intermediate data.

Stage 3. Next, with the aid of PKBD [6] (it is a tool of PESoT), the decision tables were imported and presented in the form of logical rules. The imported decision tables were refined in the RVML (Rule Visual Modeling Language) form (Figure 5).
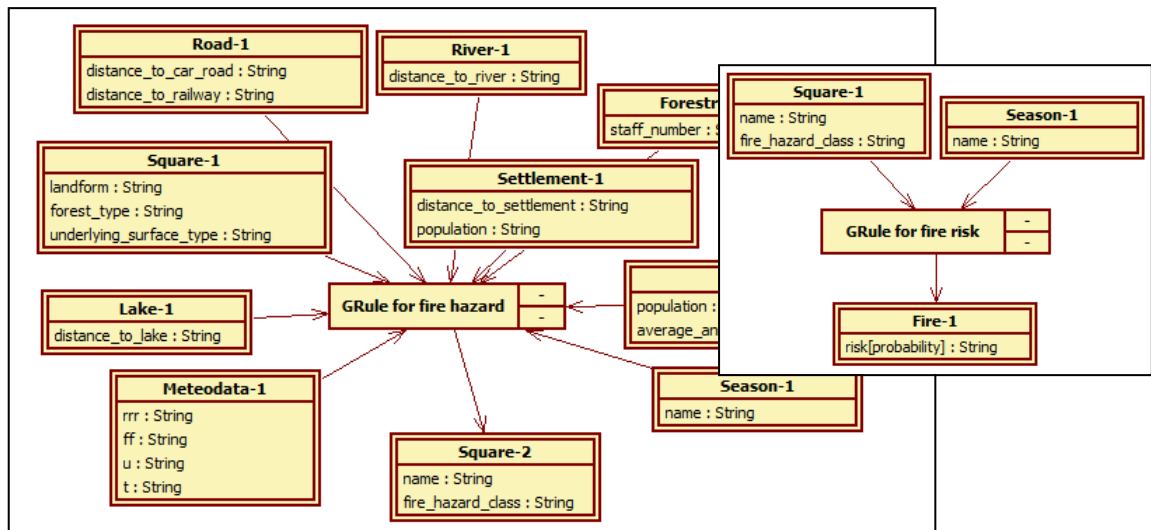

Figure 5: Rule templates (generalized rules) for the formation of specific knowledge base rules.

Stage 4. For two segments of the knowledge base, a code was generated on CLIPS, which was used to debug the obtained knowledge bases, later presented in the form of PHP codes (Figure 6).

Stage 5. Testing and integration will be done in the future.

4. Conclusion and Future Works

In this paper, we consider the use of the PESoT technology and tools for prototyping rule-based

knowledge bases by using automated analysis and transformation of decision tables. The formed knowledge bases can be used to create an intelligent decision-making support software module in the form of a WPS service for analyzing and forecasting the risk (hazard) of forest fires based on information about the forest fire hazard class, weather conditions, and other factors. An illustrative example demonstrating the fundamental applicability of this approach is presented.

Figure 6: Fragments of generated codes.

The technology is designed for end-users and reduces the time for creating prototypes of AI modules and expert systems by automating the codification stage and using existing domain models.

Our approach has a certain level of universality and after its improvement can be used in various domains, for example, for solving tasks in the field of industrial safety inspection [20].

In the future, we plan to make a quantitative evaluation of the proposed technology by conducting computational experiments.

5. Acknowledgements

6. References

[1]   E. Coronado, F. Mastrogiovanni, B. Indurkhya, G. Venture, Visual Programming Environments for End-User Development of Intelligent and Social Robots, a Systematic Review, Journal of Computer Languages 58 (2020) 100970. doi: 10.1016/j.cola.2020.100970.

[2]   B.R. Barricelli, F. Cassano, D. Fogli, A. Piccinno, End-user development, end-user programming and end-user software engineering: A systematic mapping study, Journal of Systems and Software 149 (2019) 101-137.

[3]   M. Santos, M.L.B. Villela, Characterizing end-user development solutions: A systematic literature review, Lecture Notes in Computer Science 11566 (2019). doi: 10.1007/978-3-030-22646-6_14.

[4]   A.Yu. Yurin, Technology for Prototyping Expert Systems Based on Transformations (PESoT): A Method, CEUR Workshop Proceedings 2677 (2020) 36-50.

[5]   A.Yu. Yurin, N.O. Dorodnykh, O.A. Nikolaychuk, M.A. Grishenko, Prototyping Rule-Based Expert Systems with the Aid of Model Transformations, Journal of Computer Science 14 (5) (2018) 680-698. doi:10.3844/jcssp.2018.680.698.

[6]   A.Yu. Yurin, N.O. Dorodnykh, Personal knowledge base designer: Software for expert systems prototyping, SoftwareX 11 (2020) 100411. doi: 10.1016/j.softx.2020.100411.

[7]   I.V. Bychkov, A.Yu. Yurin, A method and tools for prototyping components of intelligent systems based on transformations, Journal of Physics: Conference Series. 13th Multiconference on Control Problems (MCCP 2020) 6-8 October 2020, Saint Petersburg, Russia. 1864 (2021) 012042. doi:10.1088/1742-6596/1864/1/012042.

[8]   I.V. Bychkov, G.M. Ruzhnikov, R.K. Fedorov, A.E. Khmelnov, A.K. Popova, Organization of digital monitoring of the Baikal natural territory, IOP Conference Series: Earth and Environmental Science 629(1) (2021) 012067. doi:10.1088/1755-1315/629/1/012067.

[9]   H.R. Pourghasemi, A. Gayen, R. Lasaponara, J.P. Tiefenbacher, Application of learning vector quantization and different machine learning techniques to assessing forest fire influence factors and spatial modelling, Environ Res 184 (2020) 109321. doi:10.1016/j.envres.2020.109321.

[10]   L.V. Golubeva, I.V. Latisheva, K.A. Loschenko, A.S. Schebilkin, Investigation of the influence of meteorological factors on the occurrence and spread of forest fires in the Irkutsk region, The bulletin of Irkutsk State University. Series «Eearth science» 22 (2017) 30–40. (in Russian)

[11]   S.V. Zalesov, G.A. Godovalov, E.Yu. Platonov, Updated scale of distribution of forest fund plots by natural fire hazard calsses, Agrarian Bulletin of the Urals 10(116) (2013) 45–49. (in Russian)

[12]   A.V. Rubcov, A.I. Suhinin, E.A. Vaganov, System analysis of weather fire hazard in forecasting large fires in the forests of Siberia, Earth exploration from space 3 (2010) 62–70. (in Russian)

[13]   Yu.Z. Shur, V.Yu. Neshataev, A.A. Stepchenko, N.V. Shapoval, Regional scales of assessment of natural fire hazard of forests, Proceedings of the St. Petersburg Scientific Research Institute of Forestry 2 (2020) 59–69. (in Russian)

[14]   A.V. Sofronova, A.V. Volokitina, Assesment of the natural fire hazard of forest areas on the territory of oil and gas comlexes using remote sensing data of the Earth, Siberian forest journal 5 (2017) 84–94. (in Russian)

[15]   Decision table. URL: https://en.wikipedia.org/wiki/Decision_table

[16]   O.A. Nikolaychuk, A.I. Pavlov, A.B. Stolbov, Rule Creation Based on Decision Tables in Knowledge-based Systems Development Platform, CEUR Workshop Proceedings, 2677 (2020) 102–112.

[17]   A.Yu. Yurin, N.O. Dorodnykh, Creating Web Decision-Making Modules on the Basis of Decision Tables Transformations, Communications in Computer and Information Science 1341 (2021) 167–184. doi:10.1007/978-3-030-68527-0_11.

[18]   J. Roger, B. Pelletier, J. Aucan, Update of the tsunami catalogue of New Caledonia using a decision table based on seismic data and marigraphic records, Natural Hazards and Earth System Science 19 (2019) 1471–1483. doi:10.5194/nhess-19-1471-2019

[19]   P.V. Senchenko, Y.P. Ekhlakov, Use of Decision Tables in Monitoring of Performance Discipline, IEEE 13th International Conference on Application of Information and Communication Technologies (AICT) (2019) 1-4. doi: 10.1109/AICT47866.2019.8981795.

[20]   A.F. Berman, O.A. Nikolaichuk, A.Y. Yurin, K.A. Kuznetsov, Support of Decision-Making Based on a Production Approach in the Performance of an Industrial Safety Review, Chemical and Petroleum Engineering 50(1-2) (2015) 730-738. doi:10.1007/s10556-015-9970-x.