Intellectual training algorithm of expert system's knowledge base at weakly structured problem area

G A Blagodatsky¹, S V Vologdin¹, M M Gorohov¹,², D A Perevedencev¹

¹Izhevsk State Technical University, Studencheskaja, 7, Izhevsk, Udmurt Republic, Russia, 426069
²Federal State Institution Research Institute of the Federal Penitentiary Service of the Russian Federation, 125130, Moscow, st. Narva, 15A building 1.

Abstract. The development of a mathematical algorithm for the self-learning of the intellectual system (IS) for the management of science-intensive projects for the creation of complex technical objects is considered. IS training occurring by optimizing the knowledge base rules using machine learning methods. This will increase the effectiveness of decision support in the implementation of science-intensive projects and the adaptation of the recommendations of the IS on the request of the decision-maker. Based on the analysis and markup of the available data, as well as the formulated requirements for the algorithm, the algorithm of regression and classification (CART) trees is chosen because of its high speed and the possibilities of formalizing the obtained dependencies in the format of production rules. In accordance with the requirements set, the training parameters were determined and the efficiency of the algorithm for classifying scientific projects for a given number of classes and for formulating new dependencies of the knowledge base was checked. A general scheme for processing the user's request with the described functionality of the intelligent module is presented.

1. Introduction

High technology projects implementation process of the complex technical objects, large volumes of heterogeneous and distributed data are often created and used; this information load makes it difficult not only to effectively use the accumulated data and knowledge, but also to understand the characteristics of the development of the project itself.

To solve this problem, scientists offer various methodological approaches to reducing the level of uncertainty of knowledge-intensive projects, for example: analysis of knowledge of experts and stakeholders [1], analysis of organizational structure, staff and information resources of the project [8], critical chain planning and principles of the theory of constraints according to the subject area of the project [2], analysis of the social network and project links [9].

At different times, models for the analysis of projects of complex technical systems were proposed. Takenberg S., Dukwitz S., and K. Schlick proposed a design engineering approach based on process and actor-oriented models [4]. Bukhin E. and Rozens H. [5] developed a multi-purpose optimization model for choosing a solution for implementing a project from a set of alternative options. In [6] examples of solving problems of project optimization in various situations are given on the basis of the structure of the decision tree constructed by the authors in time, called the "project tree". The authors
of [7] proposed graphical structures for representing knowledge about the project on the basis of a thematic and situational analysis of the life cycle. In [10] an optimization model of the planning task for project management processes with the constraints of several executive modes, multi-profile staff and priority of tasks is presented.

It should be noted that the modeling and management of knowledge-intensive knowledge-based projects is used in many fields of activity: the development of transport infrastructure [13], construction [12], the introduction of lean manufacturing methods [11], electricity generation [14], the development of strategies humanitarian logistics [15].

The urgency of developing special systems for intellectual processing of such data is confirmed by all the new research in this field. Thus, in [3] the classification of knowledge-based decision-support (KBDSS) is classified into three categories: modeling and knowledge representation technologies, reasoning and reasoning technologies and web technologies, and recommendations for their application as service systems. In particular, the paper [17] proposed an automatic model and analytical approaches to learning and forecasting the characteristics of projects, in [18] a model method for classifying new facts of various engineering systems is presented based on the intellectual analysis of the data estimates for the parameters of these systems, the authors of the study [19] describes a new approach to acquiring knowledge for the rapid development of expert systems, based on the use of linguistic rules, compatible with heuristic expert knowledge.

These circumstances significantly influenced the decision-making processes, which are increasingly based on advisory systems, develop approaches to the formation of expert solutions with collaborative filtering solutions, which are modeling confidence in the context of incomplete data [21].

In (Sakama C., Inoue K. (2003)) [20], the importance of the adductive approach to the formation of new knowledge is shown, as it relies on the search for explanatory hypotheses that are confirmed and substantiated by new relevant facts registered by the system. Within the framework of this approach, the effectiveness of the decision tree algorithm is shown for the training of intelligent system [16].

2. Description of the problem environment and the formulation of the research problem

The article discusses the online platform of the regional business incubator [25], this is an information and analytical system for the evaluation and selection of high technology projects, which is a set of databases, accounting, analytical and intelligent modules, as well as an interface with the user, the technological scheme of this system consists of: ARM1 is an automated place for a specialist, ARM2 is an automated place for an employee, ARM3 is an automated expert site, DB1 is a database of users, DB2 is a database of projects, DB3 is a database of contests, DB4 is a database of bid requests for contests, fuzzy logic, BZh - knowledge base for expert recommendations, HCD - document storage, CD - data storage, AM - analysis module, MU - project accounting and change module, EM - expert advice generation module, MIA - validation module and filter and data MPO - project appraisal calculation module MRPr - projects ranging module MAPr - Project Analysis Module, MMOb - unit machine learning system.

The intellectual core of the system is the module of expert recommendations, on the basis of which recommendations are formed in the system about the possibility of financing the project based on the analysis of its prospects and current implementation risks.

Thus, the architecture of the software complex unites the subsystems of operational work with projects, their multidimensional analysis and expert module, which realize a wide range of decisions and procedures necessary for improving the quality of organization, promotion and implementation of high technology projects into a single information space.

The quality of this system, like any ES, is determined by the size and relevance of the knowledge base used to make recommendations, therefore support for decision-making under dynamic conditions in which scientific and technical systems operate, places high demands on the adaptability and functional development of the decision support system. However, attracting experts to expand and adjust the knowledge base of ES in accordance with new data in this case is inexpedient, today this problem is effectively solved using the algorithms of machine learning.
Thus, the subject of this article is the research and solution of the scientific and technical problem of actualizing the basis of the rules of the expert system with the use of modern technologies of mathematical modeling and computational experiment.

3. Description of the method for solving the problem

Within this article, we will present an approach to modeling expert knowledge in the process of analyzing and selecting high-tech business incubator projects. On the basis of an expert survey and a description of the ontology of the subject area, according to the method developed earlier [22, 23], criteria for evaluating projects are formulated and their mathematical models are constructed. Thus, the following list of project criteria was obtained: manager's score (K1), project recognition score (K2), knowledge-intensiveness score (K3), project resource estimate (K4), commercial viability score (K5), project current performance score (K6), novelty (K7), assessment of knowledge of the problem area (K8), profitability (K9), evaluation of analogues (K10), Strategy assessment (K11), risk assessment (K12) and overall project evaluation (Y). All values of the criteria are normalized and are in the range from 0 to 10.

Thus, the concept of a common project assessment and an assessment of the structure of its criteria for identifying the project development potential is formalized. The developed algorithm for evaluating the project based on the totality of these calculated characteristics makes it possible to most accurately determine its state.

Within the framework of the described online platform, the task of selecting projects comes down to choosing the most promising project for investment at a certain point in time. In this case, the developed methodology for the formalization of projects allows to combine the algorithm for the selection and analysis of data.

Next, we describe the methodology for developing an algorithm for optimizing the rules for selecting projects based on the constructed models for evaluating it. However, before proceeding to the description of the rules of the knowledge base, it is necessary to classify projects and decide on a set of features, on the basis of which the project can be attributed to a particular class. If we formalize the conditions on the basis of the 12 criteria presented and the overall assessment (Y) of the project, we will get a satisfied complex structure of the rules, so we will analyze the significance of the signs in order to reduce.

The developed assessment algorithm involves the calculation of several complex features (“manager's assessment”, “novelty”, “profitability”, “risk assessment” and “overall assessment”), therefore, as a hypothesis, we make an assumption that these criteria are more important in classifying projects. With the help of a specially designed questionnaire, data were collected from 2570 projects submitted for registration to a business incubator. Experts assigned 4 classes to the submitted projects according to the following hierarchy (table 1).

| Class | Description                                      | Number of observations |
|-------|--------------------------------------------------|------------------------|
| A     | The project has a high potential for implementation | 310                    |
| B     | The project is attractive                         | 845                    |
| C     | The project has a commercial potential            | 759                    |
| D     | The project has a low implementation potential    | 656                    |
|       | Total:                                           | 2570                   |

The bagging method over the CART algorithm allows to evaluate the importance of signs by building a set of trees and then averaging their results. Another name for this approach is Random
forest. Figure 1 shows the results of 1000 independent rounds of cross-validation of signs.

![Figure 1. The level of significance of signs of classification projects](image)

In the Random forest algorithm, the metrics Mean Decrease Accuracy (RMSE) and Mean Decrease Gini are used to assess the importance of the variables.

Mean Decrease Accuracy measures how much the inclusion of this predictor in the model reduces the classification error. RMSE can be calculated using formula (1):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{f}(x))^2}, \quad (1)$$

where $y_i$ – $i$-th variable observation, $\hat{f}(x)$ – $i$-th prediction to $i$-th variable observation and $n_i$ – amount of samples group.

Mean Decrease Gini is a measure of the quality of separation for each tree variable using the Gini index and shows how each variable contributes to the uniformity of the nodes and leaves in the resulting random forest, calculated as (2):

$$\sum_{i \in \text{left}} (y_{\text{left}} - y_i)^2 + \sum_{i \in \text{right}} (y_{\text{right}} - y_i)^2, \quad (2)$$

where $y_{\text{left}}$ and $y_{\text{right}}$ – classification of observation errors while splitting respectively on the left and right branch of the tree.

Higher values of Mean Decrease Accuracy and Mean Decrease Gini thus indicate a higher predictive value of the variable [24].

From fig. 3 we see that the most significant 3 features can be distinguished: “manager's assessment”, “overall assessment” and “risk assessment”, which confirms the hypothesis put forward earlier.

Determining the optimal level of data description is relevant for all heuristic algorithms, in particular for the decision tree this is the size of the final tree. So, a little branched tree may not reflect some of the information important to the study about the sample space, on the other hand, too detailed a tree can load the knowledge base with too detailed and insignificant rules. In this regard, it is quite difficult to determine when the algorithm should stop building up internal nodes, because it is impossible to predict which node addition will most significantly reduce the uncertainty in the data structure. Therefore, we further evaluate the impact of a different set of features and the value of the learning stoppage criterion on the quality of project classification, for this we divide the data set into a training and test sample for 80% by 20%. To assess the quality of the model, we use $F$-measure – metric, which is the average harmonic precision and recall of the project classification algorithm (3):
Precision Recall F (1) = \beta \times \frac{\text{Precision} \times \text{Recall}}{\beta^2 \times (\text{Precision} + \text{Recall})}.

(3)

If priority is given to precision, \( \beta \) takes values in the range \( \{0;1\} \), and if recall – then \( \beta > 1 \). If \( \beta = 1 \) the formula gives the same weight of accuracy and completeness, the so-called balanced \( F \)-measure. In the framework of this work, we take \( \beta = 1 \).

The results of calculating the effectiveness of the training model with a different set of features and the level of the stopping criterion \( b \) shown in the graph (Figure 2). Here \( b \) is the minimum number of observations in a leaf of a tree, calculated as a percentage of the total number of observations.

**Figure 2.** Assessment of the quality of the model, depending on the level of the learning stoppage

From Figure 2 we see that the criteria “Overall Assessment”, “Risk Assessment” and “Manager’s Assessment” together make it possible to classify projects in terms of the subject area at the level of 90% efficiency, which is an acceptable level of model quality.

The choice of the stop criterion level equal to the presence of at least 2% of objects in each class with a slight decrease in the classification accuracy allows to obtain the optimal structure of the decision tree for its formalization, thus, you can take this value for the learning stoppage criterion. Thus, we select the specified parameters of the model for the formation of the rules of the knowledge base and its learning.

### 4. Development of an algorithm for optimizing the database of expert recommendations

Denote the original data set \( T \), whose observations are classified into \( n \) classes, then the criterion for splitting a tree, which in the CART algorithm is represented by the Gini index, is defined as (4):

\[ Gini(T) = 1 - \sum_{i=1}^{n} p_i^2, \]

(4)

where \( p_i \) – class probability (relative frequency) \( n_i \) in \( T \). If set \( T \) can be divided on \( T_1 \) and \( T_2 \) with \( N_1 \) and \( N_2 \) amount of samples, then split quality score will be equal to (5):

\[ Gini_{split}(T) = \frac{N_1}{N} Gini(T_1) + \frac{N_2}{N} Gini(T_2). \]

(5)

Thus, when building a decision tree using the CART method, a variant of branching is revealed, at which the value of \( Gini_{split}(T) \).

The presented mechanism for analyzing the database of high-tech projects can be described by the
following procedure:

1. Calculate the entropy of the original set $Gini(S)$:
   1.1. if $Gini(S) = 0$ then all objects of the initial set belong to the same class, save this class as a leaf of the tree;
   1.2. if $Gini(S) \neq 0$, then we iterate over all elements of the original set and select $k$-th tag $f_k$ with set of values $X(k)$, on the basis of each element, generate a predicate of the form $x_i \leq c < x_{i+1}$, where $c$ — a certain threshold that is most often chosen as the arithmetic mean of two adjacent ordered values of the learning sample variable and dividing the original set into two subsets.

2. We calculate the value $x_0(k) \in X(k)$ to all tags $f_k$, $k = 1, \ldots, m$, so that the measure of uncertainty of $Gini$ is minimal: $x_0^{(k)} = \arg \min_{f_k, x \in X(k)} Gini_{split}(T, X^{(k)})$.

The predicate found is part of the decision tree, it is stored, and the original set is divided into two subsets, according to the selected received condition.

3. This procedure is performed recursively for each resulting subset until the stop condition criterion.

With an acceptable stop level of the learning algorithm, we also choose the one at which it is possible to formalize the full set of rules with a slight deviation from the maximum level of accuracy. From Figure 4 we see that the most critical drop in the efficiency of the algorithm occurs at the level of the learning stopping criterion in 3% of the objects of each class.

Analyzing the structure of decision trees at the level of learning stopping criteria equal to 1%, 2% and 3% showed that at the level of 3% the model has less explanatory power, therefore the choice of the stopping criterion level equal to having at least 2% of objects in each class with a slight decrease in accuracy classification allows you to get the optimal structure of the decision tree for its formalization, so you can take this value for the criterion for stopping learning.

In this case, the following rules will be added to the knowledge base:
1) if $Y \geq 5.7$ and $K1 \geq 6.1$, then the project belongs to class A;
2) if $Y < 5.7$ and $K1 < 6.1$ and $K12 \geq 4.5$, then the project belongs to class B;
3) if $Y \geq 5.7$ and $K12 \geq 4.5$, then the project belongs to class C;
4) if $Y < 5.7$ and $K1 < 6.1$ and $K12 < 4.5$, then the project belongs to class D;
5) if $Y < 5.7$ and $K12 < 4.5$, then the project belongs to class D.

According to the results of the analysis, an algorithm for teaching an intelligent system is implemented, which allows you to expand the knowledge base based on new data on the dynamics of project development; each new rule is stored in the knowledge base until a rule excluding it is added.

Thus, the developed algorithm allows real-time generation of recommendations, current accumulated data on projects and investors' preferences.

5. Description of the format of the knowledge base rules

As was shown earlier, the final decision when choosing a project for financing depends on the overall assessment of the state of the project and the level of risk acceptance by the investor.

Let $x$ is project, $S(x)$ is project’s state. Based on the selected variables, we will describe the algorithms of the rule base for the formation of expert recommendations and the conditions for their triggering; we will present the obtained data in the form of a production model reflecting the current state of the project and specified by the function (6):

$$S(x) = Y \land K1 \land K12,$$

where: $S(x)$ — conclusion on the state of the project $x$ and its possible risks, $S_x = \{\text{Conclusion}\}$; $Y$ — overall project evaluation, $Y_x \in (0; 10)$; $K1$ — evaluation of project manager experience, $K1_x \in (0; 10)$; $K12$ — evaluation of project’s risks, $K12_x \in (0; 10)$.

So, the conclusion about the state of the project is based on a comparison of the overall assessment, the level of evaluation of the manager’s experience and the level of the project’s risk assessment. The
rule of the knowledge base for this block is as follows:

\[
\text{«Conclusion»} = \langle \text{«Project’s parameters», actualization block, } Y \land K_1 \land K_12 \rightarrow \text{solving} = \{\text{conclusion}\}, \text{else} = \text{«not enough data», decision on the status of the project and its prospects } R(x_i) >
\]

The formation of the initial rule base occurs by describing a set of rules based on the possible combinations of input and output parameters, we note that the use of this approach is advisable with a small number of variables. In addition, this approach will also allow this knowledge base to meet the criterion of redundancy and completeness, since in this case at least one inference rule exists for any input values, and the developed optimization algorithm will eliminate rules with the same inference functions.

Thus, a knowledge base, expressed in natural language in terms of the subject area, has been created; this database is the core of the decision-making module of the business incubator's intellectual system.

6. Experiment
As an example, we will select 10 projects participating in one competition of innovative projects at business incubator of Izhevsk State Technical University, and we will analyze them using the tools developed by IAS. The results of the comparison of methods for evaluating projects are shown in Table 2.

| № | Project name                                                                 | The decision of the expert committee | Information system decision |
|---|------------------------------------------------------------------------------|-------------------------------------|----------------------------|
| 1 | The eccentric skewing mechanism for oil and gas directional drilling columns | 3 degree                             | Class C                    |
| 2 | Development of an integrated technological communication system of an enterprise using RoIP technology | Did not win                          | Class D                    |
| 3 | Development of software and hardware complex to reduce the noise of ventilation systems | 2 degree                             | Class B                    |
| 4 | Automated diagnostic system for monitoring electric aircraft engines          | 3 degree                             | Class B                    |
| 5 | Automated system for the design of local area networks in residential and office premises | 1 degree                             | Class C                    |
| 6 | Internal combustion engine powered by a forced ignition gas generator        | Did not win                          | Class C                    |
| 7 | Information-measuring system for determining the parameters of the movement mechanisms of automatic weapons of small arms in a contactless way | 2 degree                             | Class B                    |
| 8 | Device for vibratory drilling                                                | Did not win                          | Class D                    |
| 9 | Development of the frequency converter for the main drive of the elevator   | 2 degree                             | Class B                    |
| 10| Broadband Shortwave SDR Modem                                                 | 1 degree                             | Class A                    |

Figure 3 shows a visual comparison of the presented decision data, along the Y axis, the options of the expert committee decisions and the automatic ES recommendations for each project are plotted in ascending order of their level.
From Figure 3 we see that the developed algorithm from a sample of projects in only two cases and no more than one rank deviated from the evaluation results of the expert committee. In this case, the Spearman's rank correlation coefficient for the given example is 91.4%, which indicates a high comparability of the results of the evaluation of projects by the competition commission and the automatic recommendations of the developed analytical module of the online platform.

On the business incubator website, the recommendations of the intelligent decision support module are displayed in the form of a visual highlight of the selected project in the list.

7. Comparison with analogues
High-tech projects include projects with a high degree of scientific novelty in the approaches used to solve a specific task, therefore, methods for evaluating and selecting such projects are developed and used in various organizational systems: scientific foundations (1), universities (2), industrial enterprises (3), special project support programs (4). The method presented in this article is denoted by (5).

Let us compare the various methods of evaluation and selection of high-tech projects in terms of aspects of the presented article, namely, according to the criteria of the functional capabilities of the methodology, as well as scalability and automation of assessment processes. We take the following scale of the degree of compliance of the considered analogue with the comparison criterion: 3 – high, 2 – medium, 1 – low. The results of the comparison of analogues are shown in table 3.

| № | Comparison criterion | Analogue | |
|---|----------------------|----------|---|
| 1 | Ability to scale the methodology on the subject areas beyond the main field of activity | (1) | 2 | (2) | 1 | (3) | 1 | (4) | 3 | (5) | 3 |
| 2 | Ability to simultaneously evaluate a large number of projects | | (1) | 1 | (2) | 2 | (3) | 2 | (4) | 3 | (5) | 3 |
| 3 | Automatic optimization of method conditions based on accumulated data | | (1) | 1 | (2) | 1 | (3) | 1 | (4) | 1 | (5) | 3 |
| 4 | Ease of introducing new conditions and parameters into the methodology | | (1) | 1 | (2) | 2 | (3) | 2 | (4) | 1 | (5) | 3 |
| 5 | Automate the final decision | | (1) | 1 | (2) | 1 | (3) | 1 | (4) | 2 | (5) | 3 |

*Sum of scores:* 6 7 7 10 15
Thus, the developed algorithms of the analytical module of the online business incubator platform provide greater efficiency in the evaluation and selection of high-tech projects in comparison with analogues.

8. Conclusion
In the presented work, the algorithm of self-learning and optimization of the knowledge base of the expert system for management of high technology projects is described, which contributes to the speed of analysis and information processing. Improving the effectiveness of processing information about the state of the project at a certain point in time in the conditions of a given situation and choosing the forms of its presentation to the decision-maker leads to an increase in the certainty of the choice of the management decision, as well as to expand the horizon of forecasting the scientific potential and the feasibility of the project.

As an example, the implementation of an algorithm for sampling projects participating in the same competition was presented, but this approach is applicable to teaching the expert system by project types, their scientific direction and other parameters, which, taking into account the speed of the algorithm, allows to formulate recommendations for each request, conditions and accumulated data.

9. References
[1] Caron F 2015 Data management in project planning and control Int J of Data Science Vol 1 no 1 pp 42–57 DOI: 10.1504/IJDS.2015.069050
[2] Yuan Q, Polychronakis Y E 2012 The development of a robust resource constrained project scheduling framework Int J of Project Organization and Management Vol 4 no 4 pp 339–367 DOI: 10.1504/IJPOOM.2012.050329
[3] Zarate P, Liu Sh – 2016 A new trend for knowledge based decision support systems design International Journal of Information and Decision Sciences Vol 8 no 3 pp 305–324 DOI: 10.1504/IJIDS.2016.078586
[4] Tackenberg S, Duckwitz S, Schlick Ch 2010 Activity and actor oriented simulation approach for the management of development projects Int J of Computer Aided Engineering and Technology Vol 2 no 4 pp 414–435 DOI: 10.1504/IJCAET.2010.035395
[5] Bukchin Y, Rozenes Sh 2011 A multi objective approach for decision making during the project life cycle Int J of Project Organisation and Management Vol 3 no 2 pp 184–203 DOI: 10.1504/IJPOOM.2011.039820
[6] Muhlemann A P, Lockett A G, Gear A E 1978 Portfolio modeling in multiple criteria situations under uncertainty Decision Sciences Vol 9 pp 612–626 DOI: 10.1111/j.1540-5915.1978.tb00749.x
[7] Daradkeh M 2017 Information visualization for decision support under risk Int J of Information and Decision Sciences Vol 9 no 3 pp 276–296 DOI: 10.1504/IJIDS.2017.1007807
[8] Burke R, Barron S 2015 Project Organization Structures In Project Management Leadership: Building Creative Teams (John Wiley & Sons) DOI: 10.1002/9781119207986.ch4
[9] Chinowsky P, Taylor J E 2012 Networks in engineering: an emerging approach to project organization studies Engineering Project Organization Journal Vol 2 no 1 pp 15–26 DOI: 10.1080/21573727.2011.635647
[10] Liu Q, Geng X, Dong M, Ye Ch 2017 Scheduling optimization of design stream line for production research and development projects Engineering Optimization Vol 49 no 5 pp 896–914 DOI: 10.1080/0305215X.2016.1216113
[11] Ballard G, Tommelein I 2012 Lean management methods for complex projects Engineering Project Organization Journal Vol 2 no 1 pp 85–96 DOI: 10.1080/21573727.2011.641117
[12] Fengjie T, Lahmer T 2018 Shape optimization based design of architype dams under uncertainties Engineering Optimization Vol 50 no 9 pp 1470–1482 DOI: 10.1080/0305215X.2017.1409348
[13] Jia J, Yang K, Yang L, Gao Y, Li Sh 2018 Designing train speed trajectory with energy
efficiency and service quality *Engineering Optimization* (Vol 50 no 5) pp 797 – 818 DOI: 10.1080/0305215X.2017.1358712

[14] Malekpoor H, Chalvatzis K, Mishra N, Mehlawat M K, Zafirakis D Song M 2018 Integrated grey relational analysis and multi objective grey linear programming for sustainable electricity generation planning *Annals of Operations Research* Vol 269 no 1 pp 475–503 DOI: 10.1007/s1047901725664

[15] Roben S, Ottjes J A, Lodewijks G, Dool A 2010 A design approach for asset supply *logistics Int J of Computer Aided Engineering and Technology* Vol 2 no 4 pp 311–323 DOI: 10.1504/IJCAET.2010.035388

[16] Kumar A, Sarkar B K 2017 A case study on machine learning and classification *Int J of Information and Decision Sciences* Vol 9 no 2 pp 179–208 DOI: 10.1504/IJIDS.2017.10005873

[17] Shi L, Newnes L, Culley S, Allen B 2017 Learning to predict characteristics for engineering service projects / L Shi L Newnes S Culley B Allen *AI EDAM* Vol 31 no 3 pp 313–326 DOI: 10.1017/S0890060417000129

[18] Quaglio M, Fraga E S, Cao E, Gavriilidis A, Galvanin F 2018 A model based data mining approach for determining the domain of validity of approximated models *Chemometrics and Intelligent Laboratory Systems* Vol 172 pp 58–67 DOI: 10.1016/j.chemolab.2017.11.010

[19] Xing H, Huang S H, Shi J 2003 Rapid development of knowledge based systems via integrated knowledge acquisition *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* Vol 17 no 3 pp 221–234 DOI: 10.1017/S0890060403017502

[20] Sakama C, Inoue K 2003 An abductive framework for computing knowledge base updates *Theor y and Practice of Logic Programming* 2003 Vol 31 no 3 pp 671–715 DOI: 10.1017/S1471007003001716

[21] Gohari F S, Aliche F S, Haghhighi H 2018 A new confidence based recommendation approach: Combining trust and certainty *Information Sciences* Vol 422 pp 21–50 DOI: 10.1016/j.ins.2017.09.001

[22] Blagodatsky G A, Gorokhov M M, Perevedentsev D A 2017 Modelling the system of fuzzy logical inference for evaluating science intensive projects *Automation of control processes* Vol 2 no 48 pp 84–91

[23] Gorokhov M M, Perevedentsev D A 2017 Designing the knowledge base of the expert system for managing scientific projects *Bulletin of Tomsk State University Management Computer Science and Informatics* Vol 1 no 38 pp 47–51

[24] Han H, Guo X, Yu H 2016 Variable selection using mean decrease accuracy and mean decrease Gini based on random forest *Software Engineering and Service Science: 7th IEEE International Conference on IEEE* pp 219–224

**Acknowledgements**

This work was financially supported by the grants of ISTU № 05.1318/18GMM and 09.04.02/18GMM