

A Review of Methods for Processing Unstructured Data in the Assessment of Mining Personnel

Anna Barbara^{1,*}, *Alexander Pimonov*², and *Lyubov Sluder*³

¹ T.F. Gorbachev Kuzbass State Technical University, Mezhdurechensk Branch, 652881, Mezhdurechensk, Russia

² T.F. Gorbachev Kuzbass State Technical University, 650000, Vesennaya str., 28, Kemerovo, Russia

³ East Carolina University, 27858, E 5th St., Greenville, NC, USA

Abstract. Skilled staff, as well as conditions for their development and motivation, are key conditions for the successful operation of the company. Creation of a high-quality Human Resources (HR) personnel management system would allow to solve this problem. For the achievement of the ultimate goal - the most effective formation and development of the personnel potential of the enterprise requires the creation of conditions for all employees allowing the maximum use and increase of labour potential, as well use of creative abilities and creative thinking. Labour competences and competence evaluation represent real challenges for companies. When modelling a high-quality HR management system, it is important to take into account features such as presence of uncertainty and a large number of unstructured data. When evaluating personnel, the cognitive abilities of the decision-maker are involved and the use of fuzzy cognitive modeling (FCM) seems to be the most promising. In addition, cognitive models allow us to present complex relationships between investigated parameters revealing influence on each other. This paper considers an expert performance evaluation system based on competency model and a fuzzy logic model. The FCM based management personnel system's is proposed. There are many performance evaluation methods; however, none is universal and common to all companies. This work brings contributions to HR management solutions, finding new ways to apply artificial intelligence (AI) techniques to processes that typically were performed by humans.

1 Introduction

One of the main issues of human capital, which the mining and metallurgical industries are now encountering, is the social enterprise. This trend predicted to intensify in the future. Deloitte's annual reports on human capital in the energy, resource and industrial sectors show that digitalization and automation of technological processes create these new

* Corresponding author: barbara_ad@mail.ru

challenges for businesses. Findings from the report “Beyond HR: Rethinking work and operations in the mining industry” [1] clearly show that companies will need to adjust their operational and workforce strategies to provide the development of “Factories of Future”.

During the fourth industrial revolution (Industry 4.0), digital transformation aftermath effecting the workflow, the workforce, and the workplace. The nature of the work is changing and some mining companies are adapting their operations to take advantage of automation, remote operations centers, analytics, mobility and other digital developments. On the other hand, there are a number of problems, for example, young people are mostly not ready to work in the coal mining industry, which makes it difficult to attract enough specialists with the right sets of competences [2].

The sustainable development goals, which entered into force on 2016, drive the formation of a national strategy to achieve sustained per capita economic growth and achieve higher levels of economic productivity by diversification, technological renewal and innovation.

Socio-economic, increasing complexity of business processes, technologic and information transformations have led the need for expansion of employee’ qualifications and the labour productivity increase. Human resources have being one of the key subjects of socio-economic systems, and talent management is a complex demanding challenge in such conditions. The employer’s managerial decisions in the selection and placement of personnel affect the development, competitive growth and socio-economic stability of the organization. In such conditions, the development and implementation of new methods and algorithms for assessing the labour potential of workers is an urgent scientific and practical task.

The task of assessing engineers for mining regions is very important, since they are responsible for creating the sustainable and agile mining which will adjust to new technological challenges and digital progress. Sustainable development and transformation of the socio-economic sphere are not possible without highly skilled workers able to effectively solve production tasks. In modern market conditions, the selection of personnel plays, if not a key, then one of the most important roles in an innovative economy.

2 Materials and methods

In this study, the authors consider the task of assessing the engineering personnel of coal mining enterprises, focusing on the unstructured source data, a high amount of uncertainty. The object of the study is the processes of assessing the personnel of coal mining companies. The subject of the research is the application of evaluation results to the building of the company's personnel management system. The methods of system analysis, mathematical statistics, the theory of expert estimates, as well as the method of fuzzy logic and cognitive analysis are used. The authors have considered various methods for evaluating engineering personel. Based on those methods, using the fuzzy cognitive maps approach to the building of an engineering personnel management system is proposed. The right personnel policy would allow to build an effective socio-economic sustainable development strategy of enterprises. The following goals could be achieved: economic growth in the region and social stability.

3 Results and discussion

According to the study by VIVES [3] about 21.8 million jobs in EU, were linked to high-tech in 2011. Some of them relate to fields of engineering, technology and ICT. Such employees are in high demand, because they generate considerable economic value. Each

one of them represents talent that needs to be managed [4]. Back in 1997 McKinsey & Co began a study which said that the war on talent was the main area where many companies were fighting to obtain or maintain a competitive advantage [5]. Engineering activity is very difficult to formalize, since it is described mainly by qualitative features, it is constantly changes due to the transformations and innovations. There is uncertainty due to the complexity of taking into account all factors. Complex activities as development and design, situation analysis, decision-making, or work planning are difficult to assess, there is a high probability of error due to the human factor.

A number of methods are used in solving such problems, among which the following can be distinguished:

1. Statistical methods include correlation and regression, factor and discriminant analysis.
2. Neural networks are suitable for difficult-to-formalize tasks such as classification, pattern recognition, and approximation tasks.
3. Fuzzy logic is successfully used to analyze conflicting, noisy data that do not have a clear range of values.
4. Decision trees and hierarchical methods are sequences of judgements and conditions tested in the process of solving problems of classification, finding possible alternatives. Baseline data are often collected through expert assessments.

3.1 Evaluation of engineering personnel labour potential, based on competency model

The process of the estimation of an engineering employee can be split into the following main stages:

1. Determination of estimation dimensions and selection of scale.
2. Using selected values to evaluate the candidates.
3. Formalization of the estimates obtained, finding consistency of indicators.
4. Decision-making based on the results of the assessment.

Next, we study the task of assessing the engineering employees' labour potential by using expert methods.

First, it is necessary to collect information about candidates by performance indicators. The proposed performance indicators may include: production volume, penetration rate, overburden volume, frequency of targets achievements, the number of incentives (penalties) during the reporting period, the frequency of emergency situations, downtime, etc. As a result, employees with best and average indicators are selected. In combination with expert assessment of level of competences the system issues the recommendation to professional development, skills training or retraining, transfer in a talent pool or drop-out of the candidate. It should be noted that it is almost impossible to analyze the efficiency of a candidate having internship or probationary period, in which case the assessment is not performed. Each competency is evaluated on a 100-point scale rating and has five levels: "very weak", "weak", "medium", "high" and "very high". Since competence is a qualitative variable, to formalize it into quantitative variables is possible by expert assessments. The method of snowball sampling [6] is applied in order to form an expert group from among people most knowledgeable about professional skills of employees.

Suppose that m experts are evaluating n candidates. Let $X = \{x_1, x_2, x_3, \dots, x_m\}$ is a set of experts. A set of candidates is $Y = \{y_1, y_2, y_3, \dots, y_n\}$. Each one is characterized by a set of indicators united in four groups of competencies (professional, social and communicative, personal and cultural ones), determining the level of coherence to the given position. Let's define them as $\bar{K} = \{k_1, k_2, k_3, \dots, k_{\gamma_4}\}$. Based on the analysis of the survey of experts and

guided by research of the International Engineering Alliance, which presented graduate attributes and professional competency profiles for three professional tracks: engineer, engineering technologist and engineering technician [7], we have identified four groups of competences. $\bar{K}_1 = \{k_1, k_2, k_3, \dots, k_{\gamma_1}\}$ - the first group of competencies (professional); $\bar{K}_2 = \{k_{\gamma_1+1}, k_{\gamma_1+2}, k_{\gamma_1+3}, \dots, k_{\gamma_2}\}$ - the second group of competencies (social and communicative ones); $\bar{K}_3 = \{k_{\gamma_2+1}, k_{\gamma_2+2}, k_{\gamma_2+3}, \dots, k_{\gamma_3}\}$ - the third group of competencies (personal); $\bar{K}_4 = \{k_{\gamma_3+1}, k_{\gamma_3+2}, k_{\gamma_3+3}, \dots, k_{\gamma_4}\}$ - the fourth group of competencies (cultural).

The result of employee's activity can be evaluated by l key performance indicators and the candidates can be classified due to the results of performance. Denote $E = \{e_1, e_2, e_3, \dots, e_l\}$ - a set of key performance indicators. Each indicator has one of the levels $L = \{L_1, L_2, L_3\}$ "high", "mid", "low" respectively. Based on the analysis of performances indicators and the initial competences vector the competences making decisive impact on degree of compliance of the candidate of an estimated position are defined. The result is a vector $\bar{K}' = \{K'_1, K'_2, K'_3, K'_4\}$ - the level of conformity of the candidate of the engineering position based only on the most informative indicators. Depending on the force of influence each group of competences can be assigned its own weighting factor.

An expert assesses the rate of each competency. The result can be represented as a matrix R_X where the rows represent the opinion of an individual expert, and columns represent an assessment of each competency:

$$R_X = \begin{pmatrix} r_{11} & \dots & r_{1\gamma_1} & r_{1\gamma_1+1} & \dots & r_{1\gamma_2} & r_{1\gamma_2+1} & \dots & r_{1\gamma_3} & r_{1\gamma_3+1} & \dots & r_{1\gamma_4} \\ r_{21} & \dots & r_{2\gamma_1} & r_{2\gamma_1+1} & \dots & r_{2\gamma_2} & r_{2\gamma_2+1} & \dots & r_{2\gamma_3} & r_{2\gamma_3+1} & \dots & r_{2\gamma_4} \\ \dots & \dots \\ r_{m1} & \dots & r_{m\gamma_1} & r_{m\gamma_1+1} & \dots & r_{m\gamma_2} & r_{m\gamma_2+1} & \dots & r_{m\gamma_3} & r_{m\gamma_3+1} & \dots & r_{m\gamma_4} \end{pmatrix}.$$

After evaluation of the employee by a group of experts, a difference of opinion is analyzed, which identifies biased experts who are formally suitable to the procedure or are poorly aware of the candidate's job. For this purpose coefficient of variation CV (1) and Cronbach's Alpha (2) are calculated by the following formulas:

$$V_j = \frac{\sigma_j}{\bar{r}_j}, \tag{1}$$

$$\alpha = \frac{m}{m-1} \left[1 - \frac{\sum \sigma_j^2}{\sigma^2} \right], \tag{2}$$

where σ^2 is the variance of the sum of the estimates made by each expert; \bar{r}_j is the average values of the level of each competency; σ_j^2 is the variance of the estimates of each competency, m is the number of experts. CV and Cronbach's Alpha [8, 9] are used to determine the consistency of expert opinions, their values belong to the interval [0, 1].

Than the closer the CV is to 1, that the variation is greater and the studied objects are less uniformity. Data with a coefficient of variation (CV) greater than 33% should not considered due to extreme sampling variability. Than the closer the Cronbach's Alpha is to 1, that the expert opinions are more unanimous. Where expert opinions are poorly agreed,

them assessments are analyzed to identify the most divergent from the group's average opinion. After analysis, the opinion of each expert is given a weighting factor, and the opinions of experts found to be biased or formalistic may not be taken into account. The result is a vector of the average opinion of experts on each competence $R'_X = (r'_1 \dots r'_{\gamma_4})$. The coordinates are calculated by the formula:

$$r'_j = \sum_{i=1}^{m'} \alpha_i r_{ij}; \quad j = \overline{1, \gamma_4}, \tag{3}$$

where m' – the number of experts, α_i – the weight factor of opinion of the experts is calculated by the formula:

$$\alpha_i = \left[\frac{1}{\sigma_i^2} \right]; \left[\sum_{i=1}^m \frac{1}{\sigma_i^2} \right], \tag{4}$$

$$\sigma_i^2 = \frac{1}{n-1} \sum_{j=1}^n (r_{ij} - \bar{r}_j)^2 \quad \text{при } i = 1, 2, \dots, m, \tag{5}$$

where σ_i^2 – the variance of competency estimates given by each expert.

Key performance indicators (KPI) are set for each position.

In each group of candidates classified by key performance indicators, it is necessary to conduct an analysis by competencies. Competencies that really do not affect the work of engineers are excluded from the database. As a result, a new database of competencies for assessment with the most valid parameters is formed. The results of assessment of candidates of each KPI-class can be presented in the form of matrix R_E^p , where p - number of KPI-class, rows of matrix represent assessment of candidate, columns indicate estimates of each competence, n_p - number of candidates included in class p :

$$R_E^p = \begin{pmatrix} r_{11} & \dots & r_{1\gamma_1} & r_{1\gamma_1+1} & \dots & r_{1\gamma_2} & r_{1\gamma_2+1} & \dots & r_{1\gamma_3} & r_{1\gamma_3+1} & \dots & r_{1\gamma_4} \\ r_{21} & \dots & r_{2\gamma_1} & r_{2\gamma_1+1} & \dots & r_{2\gamma_2} & r_{2\gamma_2+1} & \dots & r_{2\gamma_3} & r_{2\gamma_3+1} & \dots & r_{2\gamma_4} \\ \dots & \dots \\ r_{n_p 1} & \dots & r_{n_p \gamma_1} & r_{n_p \gamma_1+1} & \dots & r_{n_p \gamma_2} & r_{n_p \gamma_2+1} & \dots & r_{n_p \gamma_3} & r_{n_p \gamma_3+1} & \dots & r_{n_p \gamma_4} \end{pmatrix}.$$

For each column of the matrix, the coefficient of variation is calculated. It characterizes the relative measure of deviation of individual values from the average. After excluding competences for which the coefficient of variation exceeds the permissible value, we obtain a vector $K' = \{K'_1, K'_2, K'_3, K'_4\}$, here K'_1, K'_2, K'_3, K'_4 are additive convolutions by significant parameters of competences of each group. The resulting vector $K' = \{K'_1, K'_2, K'_3, K'_4\}$ associated with the value φ - employee's work potential level:

$$\varphi = \sqrt[A_1 + A_2 + A_3 + A_4]{K_1^{A_1} K_2^{A_2} K_3^{A_3} K_4^{A_4}},$$

where A_1, A_2, A_3, A_4 are the weights of the competence groups.

The advantage of this approach is that there is no simple data averaging of different qualitative characteristic groups of competences, which eliminates the compensation of weakly expressed competencies by stronger ones. Moreover, for close values, the result of multiplicative aggregation is close to additive convolution.

3.2 Applying fuzzy logic to personnel assessment by key performance indicators

Fuzzy logic is one of methods used to personnel assessment. Since qualities describing engineering personnel production tasks are of qualitative nature, it is difficult to assign to them numerical values. For that reason, the fuzzy logic is applied in the matching process. Fuzzy logic is one of artificial intelligence technique. It's used to develop models and programs of the intellectual behavior. One of the approaches this technique is Logic. The main purpose of which is formalization of natural reasoning [10]. Researches of scientists in the field of the application of fuzzy logic in socio-economic systems make it clear that it's possible to apply these methods for personnel assessment [11, 12, 13].

Let us consider the task to determine engineering personnel performance based on fuzzy data. Employee's activity can be assessed by k performance indicators X_1, X_2, \dots, X_k . Each indicator corresponds to one of three terms $L = \{L_1, L_2, L_3\}$ - "high level", "medium level" and "low level", respectively. Engineering personnel should be classified by performance based on the analysis of these indicators. At the output, we get the value of variable Y - "labour efficiency". The set $W = \{W_1, W_2, W_3\}$ is used as a term-set of the linguistic variable, respectively {"high", "acceptable", "insufficient"}.

The membership functions $\mu(x)$ and $\mu(Y)$ assigns a grade of membership to each object based on a characteristic. To form membership functions generally use expert assessments or fuzzy clustering methods. We propose to use expert methods to solve the assessing task. To form the inclusion functions, use the formula

$$\mu^i(x) = \frac{n^i}{N}, \quad (6)$$

where x - the numerical value of characteristic, n^i - the number of experts who believe that the employee is characterized by the i -th term; N - the total number of experts.

After linear approximation of the obtained estimates, we obtain membership functions for all three terms of each of the input variables.

When calculating the output membership functions for Y - "labour efficiency", fuzzy rules base for output system is formed. It's based of empirical experience or expertise in the problem area like the fuzzy products of a standard form. Then output is translated to numeric format using the gravity center (centroid) method. This step is called the defuzzification.

$$Y = \frac{\sum a_i Y_i}{\sum a_i}, \quad (7)$$

where Y_i - the numerical value of i -th point coordinate the x -axis, a_i - the corresponding coefficient the y -axis.

It is possible to use expert methods to construct membership functions, by directly setting the membership function by a table, graph, or formula. This option can be implemented on the basis of information contained in regulatory documents, regulations, rules, requirements, etc. If information to form the membership functions is not sufficient, the construction of a fuzzy model should begin with piecewise linear membership functions, and then they can be adjusted.

3.3 Applying a fuzzy cognitive maps to personnel management

The results of the assessment of engineers personnel, described above, can be used to manage the workforce and to develop solutions to improve performance of enterprise. One of the most promising methods is cognitive modeling. The cognitive approach is good to apply in management as the factors affecting target concepts much to do with uncertainty. They are heterogeneous and difficult to measure values. It makes difficult to apply classical methods. The use of fuzzy cognitive maps (FCMs) allows the development of formal methods and models to intelligently support decision-making in management. Such tasks are not easy to solve: to begin with, it is necessary to define targets (concepts) that control concepts to identify causal relationships between them.

Cognitive maps are mathematical models used to describe complex weakly structured systems. This term first was proposed by E. Tolmen [14]. In turn, R. Axelrod presented a cognitive map in the form of a directed graph when considering the decision-making problem in managing economic and political processes [15]. Cognitive maps represent a simulated socio-economic system in the form of a directed graph with the set of nodes (concepts) and edges (relationships between concepts). Relationships reflect the impact of one concept on another and are basic elements of the cognitive map. However, the applying possibilities of classical cognitive maps for analysis are significantly limited. Fuzzy Cognitive Maps (FCMs), were proposed by B. Kosko, allow you to do a more exact description of the subject area model [16]. Concepts can take values from the segment $[0, 1]$. "Fuzzy" is due to the fact that the weights of edges (the degree of impact between concepts) can take not only discrete values 0 or 1, but from the real range $[0, 1]$. But Kosko's FCMs are not correctly considered "true" fuzzy. In fact, many modifications of the FCMs that currently exist allow to study the fuzzy parameters of the processes [17, 18].

Recently, many different approaches to modeling using FCMs have been developed, and methods are constantly being modified. Interest in cognitive analysis is not accidental, since the study of complex socio-economic processes quantifiably but also qualitatively with use of modern information technologies opens up wide opportunities for research.

When constructing cognitive maps, the following methods and their combinations can be used:

1. Modeling by the decision maker. This method does not take a long time, as no additional resources are attracted; but this is relatively subjective method that depends on competence of each decision maker.
2. Modeling by experts on the basis of available data, however, such a method is rather labour intensive and time-consuming.
3. Aggregation of the views of the expert group obtained from the survey. This method, like the second, is very labour-intensive, despite its advantage of a larger estimates range.
4. Modeling FCMs based on open sample surveys. Advantages include the possibility of modeling comparative cognitive maps, but like the previous two, the method is labour-intensive.

Cognitive modeling is carried out in stages, for example, like on Figure 1.

The management personnel system's FCM applied is listed below. FCM is representing graph $G=(K,W)$, where K is a set of the concepts, and $W, w \in [-1,1]$ a set of relationships between concepts. A set of the concepts are characterized by linguistic variables. The relation W is represented as an adjacency matrix $W=\{w_{ij}\}; i, j=1, \dots, n$, where n is the number of concepts.

Target concepts represent key performance indicators (KPIs). They can be divided into three groups: work quality, plan execution, labour safety. Managed concepts include the competencies of engineering personnel, labour discipline and working conditions.

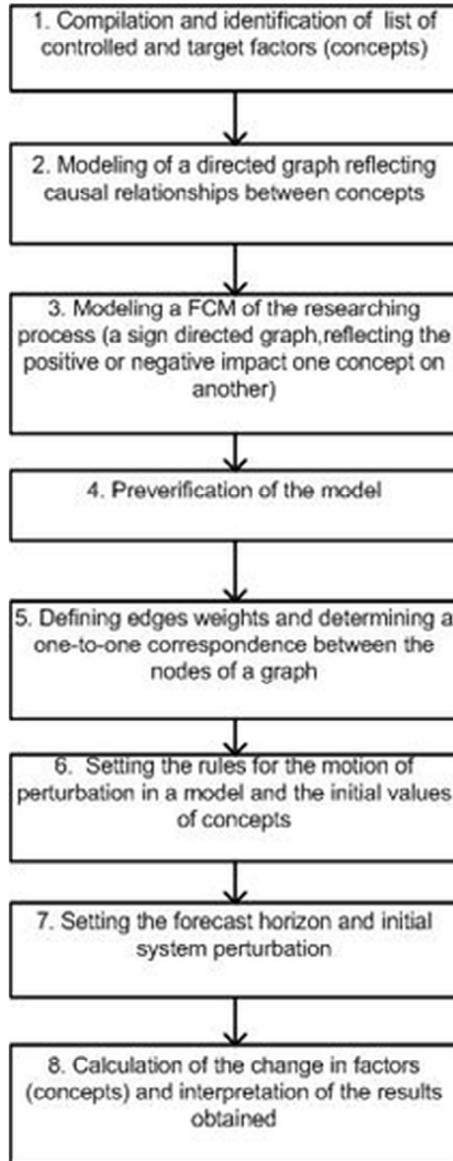


Fig. 1. Cognitive modeling stages.

The relationships between the concepts are presented the personnel management system's FCM (Figure 2). The relationships between nodes (concepts) marked with a dotted line reflect the negative influence of one concept on another. The relationships marked with solid lines determine the positive influence of concepts. Then, it is possible to analyze and predict the impact of the level of competence, working conditions and labour discipline on key performance indicators by created fuzzy rules for characterizing concepts and given a weight to the edges of the graph. Thus, it is possible to monitor the effects of managed concepts on targets. These operations are able to allow getting the desired results by adjusting the control effects.

11. G. Kłosowski, A. Gola, A. Świć, 13th International Conference. Advances in Intelligent Systems and Computing. Springer, Cham, **474**, 505-513 (2016)
12. R. Md Saad, M.Z. Ahmad, M.S. Abu, M.S. Jusoh, AIP Conference Proceedings, **1522**, 462-469 (2013)
13. O.W. Samuel, M.O. Omisore, E.J. Atajeromavwo, *Online fuzzy based decision support system for human resource performance appraisal*, Measurement, **55**, 452-461 (2014)
14. E.C. Tolmen, Psychological Review, **55(42)**, 189-208 (1948)
15. R. Axelrod, *Structure of Decision: the cognitive maps of political elites* (Princeton University Press, New York, 1976)
16. B. Kosko, Journal of Man-Machine Studies, **24**, 65-75 (1986)
17. C.D. Stylios, P.P. Groumpos, J. Intel. Rob. Sys., **26**, 389-403 (1999)
18. P.P. Groumpos, Studies in Fuzziness and Soft Computing. Ed. by M. Glykas, **247**, 1-23 (2010)