Artificial intelligence technologies to identify similar objects and formalization of inference rules

N I Yusupova¹, O N Smetanina¹, E Yu Sazonova¹, A I Agadullina¹
¹Ufa State Aviation Technical University, Ufa, K. str. K. Marx 12, 450008, Russia

E-mail yussupova@ugatu.ac.ru

Abstract. Analysis of data accumulated during the activities of the enterprise may be applied for creating novel business models and more efficient use of information resources. Specific requirements for analysis of data in socio-economic areas entail the need to develop or select the required tools and methods. This article considers identification of implicit knowledge on a complex of historical and operational data. The task statement and the methodology of solution are given. Data mining technologies are applied for identifying hidden common factors in data. Interpretation of the results, formalization of knowledge, formation of the knowledge base and use of the inference mechanism allow us to organize decision support. As an example, the task of personnel selection is considered. The choice of certain factors that are significant for each enterprise, which are the basis for the selection of personnel, will make it possible to spread the solution method for enterprises where strict requirements for information security are imposed. An analysis of decision-making support research in this area, is conducted with the use of artificial intelligence technology. Neural network technologies for data analysis are proposed to be used by the authors.

Keywords: Neural network technology; Data mining; Decision support; Human resource management

1. Introduction

Effectiveness of the enterprise depends on the quality of the decisions taken (including in the field of organization and information security) at the human resource management. Large amounts of information are accumulated during the company’s activities in this area. Therefore, we attempted to use this data to identify implicit knowledge and use it to support decision making.

Issues of decision making in the field of human resource management are considered by many Russian and foreign researchers: Azhmukhamelev A. I. [1], A. Buzuev [2], Dergacheva M. Yu. [4], Evstratova E. I. [3], Kuts Yu.I. [6], M.A. Navrotsky [4], Prokushin Ya.E [9], Protalinskiy O.M., Fedyaev O.P [10], Shagieva Yu.R. [11], Farrés Rocabert A. [12], Hamidah Juntan [13], Hossein Hakimpoor [14], M. Steven Walczak [16], Loo-See Beh [17], Hao G. [18], Raphael Olufemi [19], Stefan Strohmeier [20], Ernst Biesalski [22]. Analysis of research has revealed that the decision support for various staff management tasks using neural network technology.

Here we propose and implement a methodology for organizing decision support for the selection of staff based on an analysis of collected historical data (for example, when filling out a questionnaire) and operational data (personnel assessment) and its implementation. The proposed generalized methodology can be in the complex developed applications for filling out questionnaire questions and collecting data.
in the form of completed questionnaires and forming a knowledge base can be used for other tasks in the field of human management resource, and many other areas. Part of the steps of the methodology is carried out using the analytical platform Deductor.

2. State of the problem

Authors [1,2,6,9] address issues of mathematical models development of the staff selection problem and methods for solution in order to create software.

There are many researches on decision support using artificial intelligence techniques. So, the authors [4] presented the concept of organizing intelligent decision support in the field of staff management, based case reasoning and ontology. It allows increasing the effectiveness of decisions in this subject area. Fedaykina I.P. [10] addresses decision support issues with using fuzzy knowledge base.

Evstratova E.I. [3] considers the possibility of using artificial neural networks in order to simplify and improve the efficiency of staff management processes. It should be noted that neural network technologies are widely used for organizing decision support in the field of staff management.

Farrés Rocabert A. [12] describes a Matlab toolbox for career-oriented social network that allows recommending a candidate at the request of an employer for a specific task. Authors [13]. explore the possibility of using machine learning to solve staff problems, and Daniel Faggella consider machine learning to select new team members. Hossein Hakimpoor [14] describes the use of neural networks for business intelligence in general. Steven Walczak [16] proposes to ensure an increase in the efficiency of the manager’s work and, in some cases, management actions using the neural network apparatus.

Loo-See Beh et al. [17] propose the use of hybrid intelligent methods. Example methods of this kind are machine learning and knowledge-based approach for extracting new knowledge. Hao G.[18] describes neural network model for solving typical task of staff planning — the task of drawing up a list of airport ground staff.

Author [19] describes the system that collects data from candidates via web-interface and compares them with the relevant posts. The basis is neural network web model of human resource management system.

Stefan Strohmeier [20] offers the potential of artificial intelligence in human resource management for using in six selected scenarios (prediction of turnover with artificial neural networks, search of candidates with intelligent search engines, staff rostering with genetic algorithms, analysis of personnel tonality with text mining, resume data collection with information extraction and self-service of employees with interactive voice response). Robert Charlier [21] uses artificial intelligence technology to speed up the search for talent.

Ernst Biesalski [22] notes possibility of using knowledge-based technologies to support decision-making in staff management.

The issues of decision support in staff management are discussed in this article. The proposed method allows implementing a number of tasks (functions), in particular, staff selection functions.

The authors of this article have experience in conducting research using artificial intelligence technologies [2, 12-13], in particular, using neural network technologies and technologies based on knowledge.

The general formulation of the task of processing large-scale data and identifying unobvious, but useful knowledge to use in decision-making will be formulated below. The range of tasks for which such an approach is characteristic, and this method of solution are much wider than it presented in the article

3. Identifying non-obvious knowledge: task statement and methodology

The general task of identifying non-obvious knowledge based on data is as follows. There are historical and current data. It is necessary to reveal knowledge based on these data that can be interpreted and used to support decisions in staff management. A feature of the statement is the separation of data into
historical and current (operational). The mathematical formulation of the problem can be stated as follows:

Given $X = \{x_1, x_2, \ldots, x_l, x_{l+1}, \ldots, x_n\}$ - historical data with $i = 1, \ldots, l$ and current data with $i = l + 1, \ldots, n$.

Find: $F: X \rightarrow Z$, function that maps data to knowledge.

The method for solving the problem (Fig. 1) includes the following stages: data collection, accumulation and preliminary processing (for example, transformation of non-numeric data into nominal data); checking the correctness of data sets and their cleaning (for example, based on filtering (allows working with data that meets specified conditions), evaluating data quality (filling in missing data, editing anomalous values, etc.); conducting data mining; interpretation of results (for example, clustering carried out at the previous stage, analysis of objects of one cluster and interpretation of results taking into account subject semantics), formalization of knowledge and formation of a knowledge base (choice of knowledge representation model and formation of a knowledge base), organization of support for management decisions based on the identified knowledge (application of inference engine, an adequate knowledge representation model).

Regarding data mining, among the methods for identifying and analyzing knowledge, we can distinguish predictive methods (classification, regression, time sequence prediction (series)) and descriptive (to describe patterns in data) clustering, association, sequence.

Here we will rely on descriptive methods. Neural networks can be used as tools and methods for analysis.

4. Solution for staff selection

To solve the task of staff selection, the following steps are implemented.

Stage 1. Description of the collection, accumulation of data and preliminary processing. It is assumed that for this task historical data (collected earlier in the selection of applicants), and operational data (as an assessment of the “active” staff), are used.

The stage is implemented using an application developed in framework of the research.

During the selection process information about the applicant was collected: age, education (secondary, secondary vocational, higher, humanitarian, technical, name of the university, specialized or non-core education), additional education, retraining, obtaining additional skills (advanced computer use, knowledge of a foreign language, driving skills and driving license), work experience (work experience in a similar position, work experience in a managerial position, length of service, terms of work one place, reasons for dismissal), level of wages, marital status and children, bad habits, reasons for choosing a company, etc.

Professional testing has revealed some information about the level of communication, ability to work in a team, ability to work on a computer at the level of an advanced user. All information is stored in the database.
For staff assessment the “360 degrees” method was used. The assessment is conducted according to the list of competencies presented to the position of applicant. The method can be adapted for other company positions.

E.g. the requirements for the seller-consultant are the following: customer-oriented (ability in its activity to proceed from the needs of the client; in the process of concluding a transaction shows what this proposal is beneficial for the client), knowledge of the whole range of equipment sold and its technical characteristics, ability to demonstrate the operation of equipment; ability to work on a computer (software product 1C: Enterprise 8), teamwork (ability to listen; ability to persuade; influence colleagues; ability to transmit information without loss; ability cooperate; cooperate; constructively overcome differences; use the potential of the group and achieve collective results; focus on team achievement of a common goal); communication skills (conflict-free, communication skills, diplomacy, ability to express thoughts verbally and in writing), etc.

When using the scoring method, it should be noted that the staff assessment is carried out by those who indirectly or directly interact with applicants. The “grouping system” is based on the fact that in each applicant gains one or another number of points. Depending on this number each employee can be assigned to one of the groups. Comprehensive assessment is used as an output for clustering. The higher the assessment, the more the employee meets the criteria.

After the data for analysis has been collected, they are pre-processed, in particular, non-numeric data is replaced with nominal data.

Stage 2. Checking the correctness of data sets and their cleaning. Verification is carried out using the analytical platform Deductor. The presence of special nodes allows for cleaning. In particular, filtering was carried out, which allows selecting the most important data and improving the quality of analysis. The data quality assessment allowed filling in the missing data, editing anomalous values. The factor analysis allows reducing the data dimension.

One of the methods of factor analysis is the principal component method, which is based on determining the minimum number of factors that make the greatest contribution to data variance.

The general model of factor analysis is: \( X_j = \sum_{i=1}^{m} b_{ij} * F_i + \sum_{i=k+1}^{m} b_{ij} * U_i + e_j, i = 1..m, k < m; \)

where \( X_j \) are original variables in space of \( m \) measurements; \( F_i \) are common factors in \( k \) measurements; \( U_i \) are specific factors in the space of \( m - k \) measurements; \( b_{ij} \) are factor loads; \( e_j \) are random errors.

The principal component model is: \( X_j = \sum_{i=1}^{m} a_{ij} * Z_i; i = 1..m; j = 1..m, \)

where \( X_j \) are original variables; \( Z_i \) are principal components; \( a_{ij} \) are transition factors from a system of variables \( X \) to the system of components \( Z \), constituting the matrix of eigenvectors \( A \), which is a solution of the matrix equation: \( R * A = \lambda * A \), where \( R \) – a correlation matrix between the original variables; \( \lambda \) – matrix eigenvalue vector \( R \).

Moreover, since the matrix \( A \) is orthogonal, its inverse matrix is equal to the transpose: \( A^{-1} = A^t \).

Therefore, the matrix equation can be rewritten as: \( R = \lambda * A * A^t \) or \( R = B * B^t \), where \( B \) is the factor load matrix.

Estimates of factor loads are the values \( l_{ij} = a_{ij} * \sqrt{\sum_{i} Z_i} \), and estimates of specific factors are given by equalities: \( e_i = \sum_{i=k+1}^{m} a_{ij} * Z_i. \)

Thus, an estimate of the factor model is obtained: \( X_j = \sum_{i=1}^{k} l_{ij} * Z_i + e_j, j = 1..m. \)

Stage 3. Conducting data mining. To identify implicit knowledge, clustering, with Kohonen self-organizing maps is used. To solve the clustering task, it is necessary to divide into clusters applicants with similar characteristics.

The set of applicants \( X \) is given, from which the training sample \( X^m = \{x_1, x_2, ..., x_m\} \subset X \) is selected, as the initial information for the considered task. The output is represented by a set of clusters \( Y = \{y_1, y_2, ..., y_l\} \), including objects with similar characteristics and estimates. The function of the distance between objects \( \rho(x, x') \) is given. It is required to split the learning sample into clusters based on the similarity of features of objects for one cluster and the difference of features of objects of
different clusters. Moreover, each object \( x_i \in X^m \) is assigned a cluster number \( y_j \). Further, each object under consideration belongs to one of the classes.

To solve such tasks, a number of algorithms and methods have been developed, in particular, on the basis of artificial intelligence technologies: K-means fuzzy clustering method, Kohonen's genetic algorithm and neural network.

The solution of the problem with the help of Kohonen's self-learning network is based on the fact that the neural network, focusing on the structure of the input vectors supplied, relates the object to a specific class.

The number of neurons in input and output layers, the learning rate and the stopping criterion are determined in the process of Kohonen network development. The number of layers is two; the number of neurons in the second layer is determined by dimension of feature vector; the number of neurons in output layer is determined by the number of classes; the learning rate is 0.3; stop criterion is error less than 0.05.

The neurons in the output layer compete with each other in the process of inputting features of an object for the right to be a “winner” whose incoming weights \( \omega \) are closest to the input image. The relation \( |\omega_j - x| \leq |\omega - x| \) is for all \( j \) for the winning cell \( j' \). The winner has the right to regulate their weight.

The Kohonen network learning algorithm is defined by the following steps.

Step 1. Initiate the weight, set the parameters of neighborhood function and learning speed.

Step 2. If the stop condition is not satisfied, go to steps 3-9.

Step 3. Perform steps 4-6 for each input vector \( x \).

Step 4. Calculate \( D(j) = \sum_i (\omega_{ij} - x_i)^2 \) for each \( j \), where \( \omega_{ij} \) – synaptic weights between neurons \( i \) and \( j \).

Step 5. Find an index \( j' \) such that \( D(j') = \min \).

Step 6. Calculate for all cells \( j \) within a certain neighborhood from \( j' \), find for all \( i \) \( \omega_{ij}(new) = \omega_{ij}(old) + \eta \Lambda(j,j') [x_i - \omega_{ij}(old)] \), where \( \omega_{ij}(new) \) and \( \omega_{ij}(old) \) – new and old weights, \( \eta \) – learning speed, \( \Lambda(j,j') \) – neighborhood function.

Step 7. Change the learning speed.

Step 8. Reduce the radius of the neighborhood function.

Step 9. Check the stop condition (reduce the learning speed to zero).

Kohonen's self-organizing map, as one of the varieties of neural networks, is used for clustering. The analytical platform Deductor is used for implementation. The algorithm determines the location of clusters in a multidimensional feature space, and the construction of a map of Kohonen allows to display multidimensional space in more convenient two-dimensional space for visualization.

The clustering scenario is ensured by the following steps:

- importing source data from a file and specifying input columns;
- settings of dividing method for the original data set into train (95% of sample) and test (5% of sample);
- settings of the map parameters (the number of cells along axes \( X = 16 \) and \( Y = 12 \), the shape of cells - hexagonal);
- setting of the training parameters (speed at the beginning of training – 0.3 and at the end - 0.005; radius at the beginning of training - 4 and at the end - 0.1; neighborhood function is step; a fixed number of clusters — 4) and the learning stops (authors consider an example recognized if the error is less than 0.05);
- implementation of the learning process (the number of recognized examples and the current error values are determined);
- settings of mapping for Kohonen network.

Kohonen maps (Fig. 2) allow to demonstrate visually the obtained results.
Stage 4. Interpretation of results, formalization of knowledge and formation of a knowledge base.

Interpretation of the results is made with the participation of an expert and taking into account the subject area. It allows us to reveal new knowledge. Knowledge formalization as a way of representing the knowledge of subject area using sign systems with the participation of a knowledge engineer, allows deciding that the identified knowledge can be represented by production rules.

Let’s consider the cluster with the highest employee scores. This cluster is characterized by the presence of family and children; age from 25 to 43 (the absence of older staff in the position, more than 43); the presence of technical education (better higher education); the presence of additional education or additional knowledge (for example, increasing knowledge in the field of information security); length of experience at the previous job from 25 to 50 months (in positions different from the considered); the number of jobs over the past 2 years is no more than one; the reason for leaving a previous job is the lack of professional growth or low wages; high level of stress resistance; having the ability to learn; the true answer to the question about the presence of friends in the company at the time of arrival, etc.

All clusters are described in the same way. Results allow us to formulate the production rules. As part of the study, an application was developed to form a knowledge base and recommendations for decision-making.

The main operations of the developed software solution are discovery, creation, editing and saving of the knowledge base.

Inference engine allows us to receive the appropriate recommendations.

The quality evaluation of the applicant’s selection is made using the formula: \[ S = \frac{K}{L} \times 100\% \], where \( K \) is the number of correctly recommended, \( L \) is the total number of applicants.

100 employees who have worked in the organization for more than a year, and information about which was not included in the initial sample for factor and cluster analysis, were selected to assess the correct functioning of the software solution developed using the proposed approach.

Thus, the following information were used: information from the questionnaires, additionally collected data in the selection process, data collected in the process of evaluating their activities. Also, production rules from the knowledge base. As a result, for 96 employees out of 100, the software solution gave the correct recommendation.

The rules of logical inference are based on the interpretation of Kohonen maps obtained during the clustering process.

5. Conclusion
Research analysis by foreign and Russian authors have shown that neural network technologies demonstrate good results in solving personnel management problems. A number of publications devoted, among other things, to the identification of implicit knowledge for decision support in this area. As a rule, operational knowledge is used in these cases.

We propose the common method for identifying implicit knowledge and organizing decision support on the example of personnel selection using both operational knowledge and historical knowledge. The neural network technologies are used, in particular, Kohonen maps of analytical platform Deductor to identify non-obvious knowledge through clustering.
The data collected about candidates in the process of their selection is used as initial data for clustering, including personal data, and evaluation data in real-time of those candidates who have been selected.

The results of clustering allow us to reveal implicit knowledge based on the similarity of characteristics (features) of those workers, who in the process of their professional activities can be assessed as effective or not.

Interpretation of the results of cluster analysis with the participation of an expert and a knowledge engineer allows us to formulate production rules.

The performed experiments using the developed software solutions for formation of the knowledge base and its use showed a good result in 96% of cases. Validation of the decision made it possible to confirm the possibility of using the proposed approach for new applicants.

6. Acknowledgements
The results of the research presented in this article were supported by Grants RFBR 19-07-00709 and 18-07-00193. The part of study which associated with use of methods for identifying similar objects and formalizing inference rules is conducted with financial support from the Ministry of Education and Science of the Russian Federation as part of the basic part of the state assignment to higher education educational institutions # FEUE-2020-0007.

References
[1] Azhmukhamedov A.I., Protalinskiy O.M. 2016 Mathematical model of motivational personnel management Vestnik of Astrakhan State Technical University. Series: Management, Computer science and Informatics vol 2 pp 50-59
[2] Buzuev A.I. 2014 Matematicheskoye i metodicheskoye obespechenie podderzhki prinятиya resheniy v sisteme upravleniya personalom krupnoy organizatsiyi Phd Diss [Mathematical and methodological support of decision support in human resource management of large organization. Phd Diss.]. Penza State University Penza p 146
[3] Evstratova E.I. 2018 The use of neural network analysis in personnel management processes [Primenenie nejrosetevogo analiza v processah upravleniya personalom] Young Scientist vol 35 pp 31-33 URL: https://moluch.ru/archive/221/52515/
[4] Navrotskiy M. A., Zhukova I. G., Dergacheva M. Y. 2014 Intelligent support of decision making in human resource management using case-based reasoning and ontology Izvestia VSTU vol 12(139) pp 91-98
[5] Danilenko A.N. 2015 Developing methods and algorithms for a decision-making intellectual support in personnel management systems Proceedings of International Conference Information Technology and Nanotechnology (ITNT-2015) pp 381-388 URL: http://ceur-ws.org/Vol-1490/paper45.pdf
[6] Kuts Yu.I. 2008 On development of automated support system for development and decision-making on personnel management [O razrabotke avtomatizirovannoy sistemy podderzhki vyrabotki i prinятиya reshenij po upravleniyu personalom]. Programmnye Produkty i Sistemy vol 1 URL: http://www.swsys.ru/index.php?page=article&id=105
[7] Guzairov M.B., Yusupova N.I., Kudryashev A.F., Elkhova O.I., Popov D.V., Rizvanov D.A., Enikeeva K.R., DolomatoY.M.Yu., Yartsev R.A., Smetanina O.N., Verkhoturov M.A., Verkhoturova G.N., Makeev G.A., Gayanova M.M., Boyko M.V., Bogdanova D.R., Klimgova A.V., Galyamov A.F., Orekhov Yu.V., Orekhov E.Yu., Verkhoturov O.M., Minigulova I.R., Malakhov N.E., Malakov E.N., Furman K.G., Vinogradova N.V., Sultanova L.B., Salakhova N.R., Varygin D.V. 2014 Methodical aspects of artificial intelligence (in Russian) (Moscow) Mechanical Engineering
[8] Yusupova N.I., Smetanina O.N., Rizvanov D. A., Enikeeva K.R. 2016 Knowledge representation models for decision support in the management of complex systems taking into account
uncertainty and resource constraints [Modeli predstavleniya znanii dlya podderzhki prinятия решений при управленiи сложными системами в условиях неопределенности и ресурсных ограничений] Proceedings of the 4th International Conference Information Technologies for Intelligent Decision Making Support (ITIDS'2016) (USATU, Ufa) vol 1 pp 24-27

[9] Prokushev Y.E. 2015 Support of decision-making for the procedure of selection of the personal for the training and development (in Russian) Statistics and Economics vol 4 pp 144-148 URL: https://doi.org/10.21686/2500-3925-2015-4-144-148

[10] Fedyakina O.P. 2004 Интеллектуализация поддержки решений по управленческому персоналу организации на основе нечетких моделей знаний. Phd Diss [Intellectualization of support of organization's personnel management solutions based on fuzzy knowledge models. Phd Diss.] Voronezh State Technical University Voronezh p 194

[11] Shagieva Yu.R. 2013 Информационная поддержка принятия решений при управленческом процессе отбора технического персонала на основе комплексной диагностики. Phd Diss [Information support for decision-making in management of process of selection of technical personnel based on integrated diagnostics. Phd Diss.] Ufa State Aviation Technical University Ufa p 200

[12] Farrès Rocabert A. 2017 Artificial Neural Network System Applied to Human Resource Management ETSEIB Barcelona Spain

[13] Jantan H., Hamdan A. R., Othman Z. A. 2010 Intelligent Techniques for Decision Support System in Human Resource Management Decision Support Systems Advances Book edited by: Ger Devlin (INTECH, Croatia) ISBN 978-953-307-069-8 pp 261-276

[14] Hakimpoor H., Arshad Kh. A. B., Tat H. H., Naser Khani, Rahmandoust M. 2011 Artificial Neural Networks’ Applications in Management World Applied Sciences Journal 14 (7) pp 1008-1019

[15] Faggella D. Machine Learning in Human Resources – Applications and Trends https://www.techemergence.com/machine-learning-in-human-resources last accessed 2020/10/20

[16] Walczak S. 2016 Artificial Neural Networks and other AI Applications for Business Management Decision Support International Journal of Socio technology and Knowledge Development archive 8 (4) pp 1-20

[17] Masum A. K., Beh L.-S., Azad A.-K., Hoque K. 2018 Intelligent Human Resource Information System (iHRIS): A Holistic Decision Support Framework for HR Excellence The International Arab Journal of Information Technology vol 15(1) pp 121-130

[18] Hao G & Lai, Kin Keung & Tan, Manzhi. A. 2004 Neural Network Application Personnel Scheduling. In: Annals of Operations Research vol 128 pp 65-90 DOI: 10.1023/B:ANOR.0000019099.29005.17

[19] Akinyedede R. O., Daramola O. A. 2013 Neural Network Web-Based Human Resource Management System Model ( NNWBHRMSM ) International Journal of Computer Networks and Communications Security vol 1 No 3 pp 75-87

[20] Strohmeier S., Piazza F. 2015 Artificial Intelligence Techniques Human Resource Management— A Conceptual Exploration Kahraman C., Çevik Onar S. (eds) Intelligent Techniques in Engineering Management. Intelligent Systems Reference Library (Springer, Cham) vol 87 pp 149-172

[21] Artificial Intelligence in HR: a No-brainer, https://www.pwc.nl/nl/assets/documents/artificial-intelligence-in-hr-a-no-brainer.pdf, last accessed 2020/10/19

[22] Knowledge Management and e-Human Resource Management, https://km.aifb.kit.edu/ws/LLWA/fgwm/Resources/FGWM03_08_Ernst_Biesalski.pdf, last accessed 2020/10/22