Intelligent decision support system for iterative analysis of dysfunctions of the institutional system of human capital development based on neural structures

Aleksandr Zhukov[1], *Sergey* Pronichkin[1,2,3], *Yuri* Mihaylov[4], and *Igor* Kartsan[5*]

[1]Expert and Analytical Center, 33, Talalikhina str., Moscow, 109316, Russia
[2]Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, 40, Vavilov Street, Moscow, 119333, Russia
[3]Central Economics and Mathematics Institute of Russian Academy of Sciences, 47, Nakhimovsky Prospekt, Moscow, 117418, Russia
[4]National University of Science & Technology MISiS, 4, Leninsky Prospect, Moscow, 119049, Russia
[5]Marine Hydrophysical Institute, Russian Academy of Sciences, 2, Kapitanskaya str., Sevastopol, 299011, Russia

Abstract. Intelligent analysis of dysfunctions of the institutional system of human capital development is an iterative process of extracting useful information and decision-making templates from organized or unorganized arrays of big data of different formats. However, existing algorithms and approaches to intelligent analysis of human capital development data cannot be applied to solving problems of effective implementation of innovation programs, since they have a specific goal aimed at creating the results of intellectual activity. As a methodological basis for the system analysis of human capital development in the digital economy, it is proposed to use ensembles of decision trees and models of the Markov decision-making process together. A software and analytical toolkit for forecasting time series based on neural structures has been developed. For the iterative analysis of dysfunctions of the national innovation system, a multi-step forecasting of the human capital development trend has been performed and forecasting horizons in the digital economy have been determined.

1 Introduction

The quality and continuous updating of innovative programs in the context of digitalization is the subject of continuous analysis and research by innovative enterprises. Given the need for highly qualified employees, as well as the constant increase in attention to creative professions, all innovative programs necessary for the development of human capital can be assessed by the following characteristics: duration, price, scope of knowledge. In short, the most valuable are training programs that are most accessible both in terms of price and tools

* Corresponding author: kartsan2003@mail.ru

and a short period of training for the personnel of innovative enterprises [1, 2]. Such training programs should provide all the necessary knowledge and tools to master the necessary knowledge and behavior that are critical for the further development of the human capital of innovative enterprises. In general, the combination of skills, knowledge and behavior is the most important criterion for assessing the progress of personnel during training, as well as later in the performance of regular work duties of employees of innovative enterprises in the context of digitalization [3].

The information technology industry plays an important role in the growth of the innovative economy. Given its popularity and great opportunities, the interest in the IT industry from the already employed population in other industries increases every year. And in order to obtain the appropriate qualification, you need to pay significant attention to education, and this is the best way to acquire the necessary skills. It is necessary to develop a scientifically based tool for predicting the entry level of employees of innovative enterprises, taking into account non-academic factors and excluding preliminary assessments of academic performance, average score, etc. Within the framework of intelligent decision support systems, it is important to correctly determine the initial level of knowledge of candidates, as well as the specific result of a certain candidate for advanced training. A tool is needed that will allow for a systematic analysis of the impact of non-academic factors on the initial level of knowledge of employees of knowledge-intensive enterprises by using scientifically based methods and means of leveling dysfunctions of the institutional system of human capital development. To achieve this goal, it is necessary to solve the following main tasks: create a set of non-academic factors that can affect the initial result of the assessment of a candidate for advanced training; analyze the sample using decision tree construction algorithms; analyze the obtained results for accuracy and speed; determine the factors that have the greatest impact on the initial result and that can be taken into account to solve the problems of iterative analysis of dysfunctions of the institutional system for human capital development.

2 Materials and methods

Institutional development of human capital within national innovation systems is carried out through various institutions, which can be divided into categories: universities and institutes, private universities, corporate universities. And all of them face the same problems during the educational process, if we take into account the IT direction.

The first problem is the process of creating a curriculum and training schedule for employees of innovative enterprises. Due to the rapid change in information technology and approaches to software development, training programs do not fully cover all the needs of the knowledge-intensive industry. The IT industry does not have the opportunity to wait a long time to train and retrain employees [4,5]. Therefore, the ability to quickly and autonomously develop programs and high adaptability are critical skills of any educational institutions. Another problem is the initial level of candidates who want to start training in educational programs and the correct determination of the level of competencies. Knowledge at the initial stage of training is critically important in order to make the right decisions on participation in training programs for certain candidates and to create more personalized human capital development programs [6, 7]. Such programs make it possible to reduce the training time, with a greater likelihood of completing the entire cycle and obtaining the necessary knowledge and skills in the context of digitalization. The next problem is the timing and schedule of training. Employees of innovative enterprises must learn a lot of material in a short time. Accordingly, good self-training skills of employees of innovative enterprises are very important [8].

Data mining is an iterative process of extracting useful information and patterns from organized or unorganized data arrays of different formats. This approach is often used to predict the academic performance of subjects of innovative programs in the context of digitalization. Algorithms for the intellectual analysis of big data are the main core of any intelligent decision-making system [9, 10]. In addition to various approaches to predicting the academic performance of participants in innovative programs, the context and subject area, a specific educational institution and the characteristics of the intellectual activity results obtained in the future that depend on them are also important. All this determines the constant need to improve existing algorithms and their defining features of innovative enterprises [11]. Clustering and classification are most often used in solving problems in the intellectual analysis of data of innovative programs.

Most of the research focuses on solving scientific problems within the following perspectives:

- determining factors influencing the success of innovative programs in the educational process;
- accuracy of the selected methods and approaches to analyzing the academic performance of learners in innovative programs in the context of digitalization.

The problem of selecting data and attributes of innovative programs is relevant in all studies [12, 13]. Basically, this is a small amount of data and uniform cause-and-effect relationships. A large amount of factual data can be obtained by analyzing the information of online courses, where a large number of learners of innovative programs study simultaneously. In the context of digitalization, taking into account many factors of online courses, it is impossible to take into account the assessment of the academic performance of an individual learner of an innovative program. This factor can be easily compared through its exact value with other participants, but learners of online courses of innovative programs differ in age, education, experience, therefore, forecasting academic performance, obtaining the results of intellectual activity and choosing accurate methods is an urgent scientific task. The current average grade is usually taken into account when solving forecasting problems, but in the case of online courses of innovative programs, the current learning outcome is irrelevant, since the knowledge slice mainly reflects the voluntary desire of a particular learner.

Personal characteristics are also often considered in many studies [14]. For example, the age and gender of the learner in the innovative program. The reason why gender is considered is the fact that females have their own, different learning style. They are more disciplined and focused. They have a higher level of self-motivation and they pay more attention to self-discipline. Age is also often used in many studies [15-18]. There are studies that directly indicate that young employees of innovative enterprises cope with the material and assimilation of knowledge better than their older colleagues. According to the study [19], age prevails over other factors, such as: teaching method and learning environment. It is important to note that personal habits have a rather small effect. Factors that relate to personal qualities, such as: interests, hobbies, family support, are not often considered in forecasting algorithms, since it is difficult to obtain measurable results.

In the context of the analysis of the institutional system of human capital development in the digital economy, existing works consider only some socio-economic factors, such as: basic education, intangible assets of innovative enterprises, and the location of business. For example, it has been proven that basic education and high profits from the results of intellectual activity are highly correlated with the learning outcomes in innovative programs [20-21]. It has also been proven that age is a fairly important factor; young employees can influence and motivate their colleagues more easily and quickly than the older generation.

Another interesting area for research is the influence of psychosocial factors on the success of learners in innovative programs. Some studies indicate that very important factors of influence are the general attitude and interaction during training in innovative programs. Moreover, self-assessment has no relation to overall academic performance. Other factors that influence academic performance are stress, proper time management, involvement in various corporate activities and general emotional pleasure from training in the innovative program. Communication in a group of employees during a lesson does not directly affect the average assessment of academic performance in innovative programs aimed at creating results of intellectual activity.

Any knowledge-intensive enterprise implementing an innovation program has its own development features that were formed within the framework of the national innovation system, under the influence of the state, political processes and socio-economic relations. Insufficient state funding for scientific research, failure to fulfill the plan of scientific and technical programs have a significant negative impact. Typically, the key competencies of innovative enterprises are significantly dispersed between different structural divisions, which often do not pay due attention to them due to certain objective and subjective factors. In addition, current regulations and provisions regarding the use of the results of intellectual activity are often inconsistent, imperfect and ineffective.

The poor state of regulatory framework for the commercialization of the results of intellectual activity of innovative enterprises has led to the fact that in the context of digitalization, a value-added tax is levied on software licensing agreements. The result of such payments is that actors in the national innovation system have no incentive to actively use the results of intellectual activity and develop human capital. There is no perfect methodology for supporting management decisions to mitigate the dysfunctions of the national innovation system in order to effectively use human capital as a tool for implementing innovation programs aimed at creating the results of intellectual activity.

3 Results

The number of studies in the field of big data mining of innovative human capital development programs in the digital economy is growing rapidly, and the variety of methods used is also growing. Based on data mining methods, predictive modeling is used to forecast the development of human capital of innovative enterprises. Most often, research focuses on solving classification problems using support vector machines. However, existing data mining algorithms and approaches cannot be applied to solving problems in the educational process of innovative programs, since they may have a specific purpose and function of creating intellectual property. This means that you first need to apply a pre-processing algorithm and only then can you apply some specific big data analysis methods. One of these pre-processing algorithms is clustering.

Decision trees remain one of the most popular tools for solving problems of human capital forecasting. Decision tree methods employ a diverse set of algorithms and mostly small sample sizes. Given the mixed accuracy results of existing decision tree algorithms, it is important to note that the structure of the attributes to be studied plays an important role in the final results.

Such an attribute as initial knowledge was chosen as a key factor for the success of passing the test of the innovation program. The results of such a test are considered as the initial level of the candidate. Since it is impossible to reliably determine the academic success of learners during their studies, only non-academic factors are taken into account. The test result is successful if the learner answered more than half of the questions asked. The experience of employees of innovative enterprises was classified into three gradations, and only specialized certification obtained within the framework of transnational information and communication companies is taken into account. The algorithms for constructing the model classify learners into two categories depending on the results of the initial test.

Several different algorithms were used in the study. Special software was developed for analysis and visualization. Decision trees and Markov decision process models were used together to predict the development of human capital in the digital economy.

The success of developing the competencies of employees of innovative enterprises was assessed as follows:

$$g(a_l^m) = \sum_{k=1}^{N_l^m} s_k(a_l^m) g_k(a_l^m) \tag{1}$$

where a_l^m is the assessment of the success of mastering the innovative program at the m-th node of the tree at the l-th iteration of decision making; g_k is the assessment of the usefulness of mastering the innovative program depending on the k-th gradation of the attribute $k = 1, \dots, N_l^m$; s_k is the probability of successful mastering of the innovative program aimed at creating the results of intellectual activity.

A Markov decision problem with a finite number of decision tree levels has the following form:

$$f_i(q_t^m) = \max_l \left[f_{i-1}(q_t^m) + \sum_{k=1}^{N_l^m} s_k(a_l^m) g_k(a_l^m) \right] \tag{2}$$

where $f_{i-1}(q_t^m)$ is the utility of the decision q_t^m taken at the t-th step to level out the i-th dysfunction of the institutional system for the development of human capital in the context of the digital economy.

In the first step, the usefulness of the decision being made within the institutional system of human capital development is determined through royalties for the special use of licenses for the results of intellectual activity:

$$f_0(q_t^m) = \sum_{j=1}^{V} \left(R_j^{min} \cdot \frac{a_j \cdot (100 - d_j)}{1000} \cdot P_j(q_t^m) \cdot H_j(q_t^m) \cdot G_j(q_t^m) + R_j^{max} \right) \cdot C_j(q_t^m) \tag{3}$$

where R_j^{min} is the minimum price of a license for the j-th type of intellectual property; a_j is the share of innovators of the j-th type of intellectual property among the actors of the national innovation system; d_j is the natural exit of actors from the market for supply and demand of licenses for the j-th type of intellectual property; $P_j(q_t^m)$ is the possible increase in licenses from one innovator within the j-th type of intellectual property; $H_j(q_t^m)$ is the volume of the market for supply and demand for the j-th type of intellectual property; $G_j(q_t^m)$ is the norm for public procurement of licenses for the j-th type of intellectual property; R_j^{max} is the maximum price of a license for the j-th type of intellectual property; $C_j(q_t^m)$ is the optimal number of consumers of the j-th type of intellectual property; V is the number of types of intellectual property.

The main essence of the proposed methodology is to calculate the economic assessment of the dysfunctions of the institutional system of human capital development through the establishment of minimum and maximum levels of payment for the special use of the results of intellectual activity. The economic component of the proposed methodology consists in the economic stimulation of the intensive use of the results of intellectual activity through improving their quality, and consequently increasing the economic assessment.

The conditions for the effectiveness of the proposed methodology for analyzing the dysfunctions of the institutional system for the development of human capital are:

- mandatory planning and costing of the results of intellectual activity;
- fees for the special use of the results of intellectual activity are established taking into account availability for subsequent commercialization.

The cross-validation method was used to evaluate the classification accuracy. The cross-validation process was repeated ten times for each execution of the classification algorithm. The following metrics were used to compare the performance, usefulness, and accuracy of the algorithm: accuracy, F-measurement, correctly classified instances, incorrectly classified instances. The identification and analysis of redundant attributes were used to check the attributes. The evaluation was performed on all training data sets. The attribute ranking was evaluated.

The attribute with the highest rank is education, and the lowest rank is age among all attributes (education, experience, profile, gender, age). The best average accuracy was obtained using the ensemble of decision trees.

For the iterative analysis of the dysfunctions of the institutional system, a multi-step forecasting of the human capital development trend in the digital economy was performed and forecasting horizons were determined. For this purpose, a software and analytical tool for forecasting time series was developed based on the linear neural structure of the model of successive projection transformations.

The assessment of the accuracy of forecasting the development of human capital in the digital economy has certain specifics and features. To assess the accuracy of forecasting time series, commonly used indicators in regression methods that are not related to time dependence are usually used. After the generally recommended comparison of the developed methods with the naive forecast method, having received sufficiently small error values, it can be argued that high accuracy of forecasts for the development of human capital in the digital economy has been achieved. Therefore, a study was carried out on a new forecasting model based on easy-to-configure and use linear neural structures of the successive projection transformation model.

The algorithm for the iterative analysis of dysfunctions of the institutional system for the development of human capital in the context of the digital economy contains the following steps:

1. Creation of separate matrices $\|f_i(q_t^m)\|_{N_l^m \times N_l^m}$, $i = 1 \dots I$ values for each counting of human capital quality indicators using the sliding window method.
2. Isolation of one principal component and training of a linear neural structure of the model of successive projection transformations on the created samples with duplicated outputs as additional inputs.
3. From the outputs obtained during training, matrices are again formed for each parameter for a different number of time windows.
4. The generated matrices are divided into training and testing.
5. Application of the neural structure of the model of successive projection transformations on matrices with different numbers of time windows for one-step forecasting of dysfunctions of the institutional system of human capital development.
6. Determining the number of time windows for which forecasting results are most accurate by calculating forecasting errors.
7. Forecasting the trend of dysfunctions of the institutional system of human capital development in the context of the digital economy.
8. Comparison of the method based on the neural structure of the model of successive projection transformations with the naive forecast method.
9. Multiple short-term forecasting of dysfunctions of the institutional system of human capital development in the context of the digital economy is carried out.
10. The forecast horizon is determined by the last found indicator, the forecast error of which approaches ten percent.

4 Discussion

The importance of creating an extension of the analysis of human capital of innovative enterprises and new algorithms for predicting dysfunctions of the institutional system of human capital development to improve the quality of implementation of innovative programs aimed at creating the results of intellectual activity is substantiated. It is comparatively and experimentally determined that the developed method based on neural structures of the model of successive projection transformations shows more accurate results.

The best indicators of multi-step forecasting of dysfunctions of the institutional system of human capital development with a fixed window satisfy forecasting using neural structures.

The trend of these time sequences, the weight of which in the sum of the components is decisive, as experiments have shown, is quite accurately predictable and gives an idea of the general nature of changes in the parameters of dysfunctions of the institutional system of human capital development in the short and medium term.

The forecasting errors of the developed scientific and methodological tools are less than the errors of a naive forecast. Since the developed method is performed with more accurate results, therefore, using this method, further multi-step forecasting of the parameters of dysfunctions of the institutional system of human capital development in the context of the digital economy was performed based on multiple short-term trend forecasting.

As the conducted research has shown, changes in the institutional system of human capital development should be positive in the direction of a flexible system of state financing of critical scientific research and development, creation of modern conditions for the functioning of socio-technical ecosystems, organization of the necessary institutional innovation programs, bringing fiscal measures to the optimal level, involvement of the scientific community in control and monitoring actions in the process of intellectual activity.

The application of the proposed scientific and methodological approaches will stimulate the leveling of dysfunctions in the institutional system of human capital development, and the catalyst for such change will be the high demand for the results of intellectual activity.

5 Conclusion

The basic factors that can influence the results of the initial test of candidates for training in innovative programs aimed at creating the results of intellectual activity were analyzed. Only those factors that can be easily collected from the resumes of candidates were taken into account. To determine the factors that most affect the result of intellectual activity, algorithms for creating a decision tree and Markov decision processes were used together. It was found that algorithms of an ensemble of decision trees are most applicable for making management decisions for the iterative analysis of dysfunctions of the institutional system of human capital development in the context of the digital economy. Such algorithms are user-friendly and achieve maximum accuracy.

The results of the study confirm that candidates who do not have the appropriate level of knowledge and experience have no chance of successfully passing the initial selection test for training in innovative programs aimed at creating the results of intellectual activity.

In addition, it has been established that the age factor has a significant impact. Applicants from the last age category, compared to other factors, show worse results. Additional education and any work experience also have a great impact on passing the initial test.

The most significant node of the decision tree, which best corresponds to the forecasting factor, is the presence of basic education. Accordingly, it was found that the educational level remains the most important factor that affects the academic performance of learners in innovative programs aimed at creating the results of intellectual activity.

As an area for further research, it is proposed to expand the list of factors in order to find hidden factors influencing the academic performance of learners in innovative programs aimed at creating results of intellectual activity. It is necessary to conduct additional research among applicants in the first five years of work in innovative enterprises, since it is obvious that this is a separate segment that can be considered a different generation with different key factors in the conditions of the digital economy.

The work was carried out within the framework of the state task of the Ministry of Education and Science of the Russian Federation on the topic "Conceptual modelling of the information and educational environment for the reproduction of human capital in the digital economy" № 121102600069-2 (code FNRN – E).

References

1. J. Brodny, M. Tutak, Journal of Open Innovation: Technology, Market, and Complexity **10**, 100210 (2024). https://doi.org/10.1016/j.joitmc.2024.100210
2. G. Qu, L. Wang, J. Chen, Technology in Society **77**, 102531 (2024). https://doi.org/10.1016/j.techsoc.2024.102531
3. Y. Su, J. Wu, Finance Research Letters **60**, 104902 (2024). https://doi.org/10.1016/j.frl.2023.104902
4. S. Fareri, R. Apreda, V. Mulas, Technological Forecasting and Social Change **196**, 122844 (2023). https://doi.org/10.1016/j.techfore.2023.122844
5. Y. Yeo, S. Jung, Telecommunications Policy **18**, 102871 (2024). https://doi.org/10.1016/j.telpol.2024.102871
6. X. Huang, F. Yang, J. Zheng, Asia Pacific Management Review **28**, 598-610 (2023) https://doi.org/10.1016/j.apmrv.2023.04.004
7. D. Hooshyar, X. Weng, P. Sillat, Computers & Education **223**, 105169 (2024). https://doi.org/10.1016/j.compedu.2024.105169
8. Y. Dai, S. Qin, Y. Tang, Acta Psychologica **246**, 104278 (2024). https://doi.org/10.1016/j.actpsy.2024.104278
9. F. Poszler, B. Lange, Technological Forecasting and Social Change **204**, 123403 (2024). https://doi.org/10.1016/j.techfore.2024.123403
10. J. Girodon, D. Monticolo, E. Bonjour, Advanced Engineering Informatics **29**, 696-713 (2015). https://doi.org/10.1016/j.aei.2015.07.001
11. I. Lee, S. Kim, S. Green, Journal of Business Research **132**, 491-506 (2021). https://doi.org/10.1016/j.jbusres.2021.04.033
12. F. Lomello, L. Bard, M. Maglione, Computational and Structural Biotechnology Journal **25**, 186-193 (2024). https://doi.org/10.1016/j.csbj.2024.09.019
13. Y. Li, G. Su, H. Pan, Journal of Cleaner Production **447**, 141569 (2024). https://doi.org/10.1016/j.jclepro.2024.141569
14. A. Zhao, Y. Yu, Computers in Human Behavior Reports **9**, 100265 (2023). https://doi.org/10.1016/j.chbr.2022.100265
15. K. Haldorai, W. Kim, K. Mussina, International Journal of Hospitality Management **122**, 103864 (2024). https://doi.org/10.1016/j.ijhm.2024.103864
16. E. Melia-Marti, G Tormo-Carbo, J. Fernandez-Guadano, European Research on Management and Business Economics **30**, 100257 (2024). https://doi.org/10.1016/j.iedeen.2024.100257
17. V.S. Averyanov, I.N. Kartsan, Voprosy cybersecurity **2(54),** 65-72 (2023). https://www.doi.org/10.21681/2311-3456-2023-2-65-72
18. Q. Huang, Y. Wang, K. Yuan, Safety and Health at Work **13**, 207-212 (2022). https://doi.org/10.1016/j.shaw.2022.02.002
19. S. Lorenz, T. Zwick, M. Bruns, The Journal of the Economics of Ageing **21**, 100363 (2022). https://doi.org/10.1016/j.jeoa.2021.100363

20. X. Li, Heliyon **10**, e37978 (2024). https://doi.org/10.1016/j.heliyon.2024.e37978
21. M. Okon, Thinking Skills and Creativity **54**, 101678 (2024). https://doi.org/10.1016/j.tsc.2024.101678