LABOUR MARKET MONITORING SYSTEM

Sergey Belov [1,2,a], Irina Filozova [1,2], Ivan Kadochnikov [1,2], Vladimir Korenkov [1,2], Roman Semenov [1,2], Pavel Smelov [1], Petr Zrelov [1,2]

[1] *Plekhanov Russian University of Economics, 36 Stremyanny per., Moscow, 117997, Russia*

[2] *Laboratory of Information Technologies, Joint Institute for Nuclear Research, 6 Joliot-Curie, Dubna, Moscow region, 141980, Russia*

E-mail: [a] sergey.belov@jinr.ru

Last years, the prospects for digital transformation of economic processes were actively discussed. It is quite a complex problem having no solution with traditional methods. Opportunities of the qualitative development of the transformation are illustrated by the example of use of Big Data analytics, particularly text analysis, for the assessment of the needs of regional labour markets in the man-power. The problem is solved using the developed by the authors the automated information system of monitoring of matching the staffing needs of employers with the training level. The system presented use the information gathering from open data sources and provides additional opportunities to identify qualitative and quantitative interrelation between the education and the labour market. The system is targeted at a wide range of users: authorities and management of regions and municipalities; the management of universities, companies, recruitment agencies; graduates and prospective students.

Keywords: labour market, unemployment, regional economics, Big Data analytics, machine learning

1. Introduction

Many institutions and entities (state, educational institutions, employers, households, citizens, etc.) are involved in the process of interaction of the labor market and the system of vocational education. Ideally, changes in the labour market should be accompanied by a coherent and balanced transformation of the professional training system to meet the real needs of a changing economy.

However, scientific abstraction differs from the reality, and therefore in all countries there are lasting researches aimed to determine the expected level of unemployment, its forms and age and gender structure. The changes in the latter seem particularly important to us because the imbalance in the labour market primarily affects the employment of young people. The search for mechanisms to protect young people from the threat of unemployment (which is higher than that of older people) [1] is a crucial task for the governments of most countries. After all, the high level of unemployment among young people does not "just" increase social tension: it could result in a favorable environment for recruitment of some young people into extremist organizations.

There are different approaches to the problem, for example, state programs, special employment conditions, etc. However, despite their development and implementation, the horizontal and vertical mismatch of now available qualification and skill requirements of the market, continues to be widespread in developing and developed countries [2, 3]. This is especially important today, at the time of rapid changes in the needs of employers due to the accelerated technology shift and even the disappearance of the professions which were recently of a high demand, on the one hand, and the emergence of completely new ones - on the other. These may prevent young people from successfully entering the labour market. As a result, their expectations of education are falling. Voluntary exclusion of young people from employment and education is also possible. But the more important result of this widespread phenomenon is the "loss" of involvement of this group in the socio-economic processes, the group is most ready to accept the transformation which is characterized in recent years as "the transition to the digital economy".

In this article we discuss approaches providing qualitatively new opportunities to study the state and needs of the labour market. We believe that traditional sociological research has certain limitations, which are discussed in more detail below. To overcome them, as well as to obtain fundamentally new research tools, it is proposed to create an automated system for monitoring the labor market, based on the technologies of Big Data and text mining.

2. Automated estimation of the labour market’s state

As we believe, effective forecasting of the labor market needs in personnel is possible only basing on an objective assessment of its condition. And this seems to be impossible without the use of information and analytical systems designed to automate the collection of data from popular services for job search, and their subsequent analysis. The purpose of this work is to identify the most popular specialties and professions [4], the calculation and results according to query key status indicators of the labour market areas and the region as a whole [5]. Thus, it seems to be reasonable the creation and development of an automated information system for monitoring the compliance of personnel needs of the market and the level of training.

An additional argument for the development of such a system is that most publications on the identification of the needs of regional labour markets are based on the results of sociological surveys of employers and employees. There is no doubt that such surveys are appropriate. However, they have certain limitations, three of which reduce the possibility of assessing the real situation in the regional labor markets:

- generalized conclusions are made based on responses of a small group of respondents;
- respondents when answering questions sometimes are not frank enough, or base their conclusions on the incorrect interpretation of the situation;
- the collection and processing of the information takes a very long time.

As a result, persons limited by this kind of information may make incorrect management decisions that lead not to a reduction, but to an increase in the unemployment rate. To avoid such an outcome, we offer a qualitatively different study having the purpose to create an intelligent system for

monitoring the real situation in the regional labor markets. Achieving this goal includes the following tasks:

- Gathering the most complete information about the real needs of employers in the graduates;
- Analysis of compliance of these needs with existing professional and educational standards.

Open sources of information served as the basis for the collection of information. As initial data on vacancies in development resources of the Internet portals "Work in Russia" [6] (the information site of the Rostrud agency), HeadHunter [7], SuperJob [8] are used. Based on this data set, we are able, if necessary, to assess changes in the needs of employers in the labour markets of the constituent entities of the Russian Federation on a daily basis. The registry of the approved professional standards is used as normative documents base [9]. The object of a separate study is to assess the completeness of the reflection of the needs of the labor market vacancies presented on the Internet.

As we believe, the practical use of this intelligent monitoring system will allow, first, to optimize the budget expenditures (of the Federation and the regions) for the training of the specialists required by the regional economies. Secondly, to develop the recommendations for making (if necessary) changes and additions to the educational programs of Universities. Third, to allow undergraduate applicants and graduates to be better guided in the demands of the labour market.

3. Matching labour market with the professional standards

As it is known, modeling of word semantics (meaning) is one of the key problems related to natural language processing. The results of semantic analysis are used in search engines, automatic translation systems and other areas related to natural language text processing.

Currently, the so-called "predictive models" based on the use of neural networks occupy a leading place in the approaches of vector representation of words (word embedding) [10]. One of the main tools for vector representation of words is "word2vec" [11].

The basic principle of "word2vec" is to find connections between word contexts according to the assumption that words in similar contexts tend to denote similar things, that is, to be semantically close. The problem solved by "word2vec" could be formalized as follows: to minimize the distance between the vectors of words that appear next to each other and to maximize the distance between the vectors of words that do not appear side by side. "Near " in this case means "in close contexts". For example, the words "analysis" and "research" are often found in similar contexts, "word2vec" analyzes such contexts and concludes that these words are similar in meaning. Context analysis is performed on large text corpora. In our task we used the Russian Wikipedia corpus and the national corpus of the Russian language, as well as models of distributive semantics of RusVectōrēs [12].

There are attempts to create a predictive model for translation of a document to vector space [13]. However, the task of comparing short sentences to semantic similarity has a certain specificity, and the use of existing models for translating words or documents into vector space without modifications gives an unsatisfactory result.

Since the texts of the formulations of educational competencies, as well as the wording of the requirements in the vacancy announcements, the analytical part of the system is based on the calculation of the semantic proximity between two short sentences. The authors developed an algorithm for translating sentences into vector space based on "word2vec".

Thus, each word corresponds to a vector of dimension *n*, which affects the accuracy of the model. Metric space of word mappings is called semantic. Projections of the vectors close in meaning to words are close and form some semantic clusters. Vector representation allows to calculate the "similarity" of words based on the calculation of the cosine distance. By analogy with the calculation of word similarity, the semantic similarity of competencies and requirements is being calculated (they are short sentences contain 10 words in average). The calculation of the vector of the described sentences is defined as the average weighted of the word's vectors. It is worth noting that words that do not have a meaning (conjunctions, particles, prepositions, pronouns, etc.) are excluded from forming of the sentence's vector.

Data collection and processing is carried out based on modern methods and technologies for obtaining information from web-based sources. In the next step, machine learning algorithms are used to translate words into a vector representation. Then, the vectors of the sentences are being calculated, which allows to identify the semantic similarity of the labor market requirements and professional competencies of higher education, which are nothing more than short text sentences. The obtained results are used to identify the relationships between two sides, labour market and educational system.

3. Labour market monitoring system

Every day several million job offers are to be actualized, analyzed and stored. To track the dynamics of the indicators and make the base for forecasting the state and needs of the labor market, it is necessary to effectively store, analyze and visualize the data on the job offers for the maximum available time (we consider data from 2015, more than 3 years by now). Therefore, the basis of the created system was created on the Big Data technologies. First of all, the following free software products were used: Spark [14], Hadoop, Kafka, Flume, Marathon, Chronos, Docker.

The implemented prototype of the automated information system is a web-based application with an intuitive user interface that provides reliable data storage. The system is built on a modular principle and includes, firstly, a module for collecting text data (functioning automatically using open sources, which are the Internet portals of recruitment agencies). Second, a data loading and storage module consisting of a distributed data warehouse (providing replication and archiving). Thirdly, the automatic processing performing the preparation of information for analysis, auto-linking requirements and competences, and machine learning. Fourth, the user interface for generating and displaying reports based on business data analysis technologies.

4. Conclusion

Speaking about the practical results of the study, it should be noted that today we have created a prototype of an automated information system for monitoring and analyzing the personnel needs of the economies of the Russian Federation. With its help, as a result of the analysis of constantly updated large data sets, it is possible to determine the compliance of higher education programs with the current expectations of employers.

This system is included in the program and technological solutions of the Situational Center for social and economic development of Russia in the Plekhanov Russian Economic University. It is also used in the activities of the Russian Institute of Labor to analyze the compliance of educational programs and standards to the needs of the labor market.

Since the system is based on a stack of Big Data technologies and machine learning methods, it could easily scale and be flexibly configured for different tasks. Thus, using the developed software and hardware platform and methods of intellectual analysis of text and media information, the problem of finding signs of illegal content on pages in social networks was solved.

The development and adaptation of the developed system can vary in accordance with the requirements of the customer, depending on the specifics of the problem – the characteristics of the region, University, etc. We believe that this system, as well as algorithms and principles of its construction, in the future it is advisable to use and solve a wider class of socio-economic problems by reconfiguring it, determined by the characteristics of the problem and the type of input data.

The special importance and timeliness of the research direction we are developing are even more obvious if we look at the prospect of predicting the expected changes in the labor market. And this applies not only to Russia, but also to global trends. The whole world is entering a new stage of development of the post-industrial, information society with an accelerating change of priorities in socio-economic development, and hence with a rapid change in the picture of the labor market. The task of scientists is to help both the generation, just entering the working life, and representatives of older generations, once mastered interesting and well-paid professionals, which suddenly ceased to be in demand, relatively painless to adapt to future changes; to offer them a kind of "compass", allowing one to consciously navigate in a changing world. And if a variety of sociological studies of the labor market can only point to the existing difficulties and dangers, then the proposed method, with its

consistent application, is quite capable of providing real, concrete assistance to a variety of actors. These are public administration bodies, entrepreneurs, senior workers thinking about retraining, and young people who are only choosing the labor way to find the best way to use their abilities, knowledge and professional competencies.

5. Acknowledgement

The work was supported by the Russian Foundation for Basic Research (RFBR), grant 18-07-01359 "Development of information-analytical system of monitoring and analysis of labour market's needs for graduates of Universities on the basis of Big Data analytics".

References

[1] J. Dolgado et al., No Country for Young People? Youth Labour Market Problems in Europe, 2015

[2] European Commission. Labour Market and Wage Developments in Europe. Annual Review, 2016

[3] A. Wolf, Review of Vocational Education – The Wolf Report, 2011

[4] Cheremisina E.N., Belaga V.V., Samoilenko Yu.I. Informational and educational environment for teaching information technologies on the basis of the Institute for System Analysis and Management of the University "Dubna" // "Open Education", 2/2014 - P. 59-65. (in Russian)

5] Petrunina O.E. Designing of information-analytical system of management of the regional labor market, Modern science-intensive technologies. - 2005. - No. 5 - P. 75-78.

[6] “Work in Russia”, Russian nationwide base of vacancies: https://trudvsem.ru/

[7] HeadHunter staffing agency: https://hh.ru/

[8] SuperJob staffing agency: https://www.superjob.ru/

[9] Russian national registry of professional standards: http://profstandart.rosmintrud.ru

[10] Barkan Oren (2015). «Bayesian Neural Word Embedding». arXiv:1603.06571.

[11] Mikolov Tomas et al. «Efficient Estimation of Word Representations in Vector Space». arXiv:1301.3781v3 [cs.CL] 7 Sep 2013.

[12] Kutuzov A., Kuzmenko E. (2017) WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. In: Ignatov D. et al. (eds) Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science, vol 661. Springer, Cham

[13] Le Quoc et al. «Distributed Representations of Sentences and Documents». arXiv:1405.4053.

[14] M. Armbrust et al., Spark SQL: Relational Data Processing in Spark. SIGMOD 2015. June 2015.