

The topology of the Russian regions by the expected life expectancy of the population with the use of the Kohonen neural network

Petr Bondarenko

Volgograd State University,
Institute of Management and Regional Economics
Volgograd, Russia
bondarenko.volsu@gmail.com

Anna Trukhlayeva

Volgograd State University,
Institute of Management and Regional Economics
Volgograd, Russia
ann.tru@mail.ru

Elena Fokina

Volgograd State University,
Institute of Management and Regional Economics
Volgograd, Russia
fokina-ea@mail.ru

Abstract - The authors proposed a method to analyze the life expectancy of the population of regions of Russia, which is carried out using a Kohonen neural network. By the method of adjusting the input weights of the neural network, the self-organizing maps of Kohonen were chosen. As the determining socio-economic factors affecting the expected life expectancy of the population, statistical indicators are selected taking into account the correlation analysis, which is combined into the following blocks: public health, ecology, social participation and safety, employment and welfare of the population. With the help of Kohonen self-organizing maps, data analysis was performed, clusters were recognized and hidden interdependencies were established between the indicators that affect the expected life expectancy of the population of the regions of the country. Regions of Russia, depending on the level of significance of life expectancy at birth of the population and determining factors of influence on it, were classified into four clusters. Regional economic systems of the South of Russia are classified as a cluster with a high level of significance of life expectancy, due to a low incidence of the population, a balanced level of the employed population and the unemployment rate, as well as a positive ecological situation in the regions.

Keywords - *cluster analysis, neural network, Kohonen self-organizing maps, life expectancy, social and economic development of regions.*

I. INTRODUCTION

The assessment of the activities of public authorities is based on overall social effectiveness, expressed in the level and quality of life of the population of the regions of the country [1].

The life expectancy at birth is an integral characteristic reflecting the quality of life. This indicator is one of the four components of the human capacity index that the UN uses to compare and assess the socio-economic development of the

population in different countries, is taken into account when calculating the happiness index, the Gender Related Development Index, and the human poverty index [2, 3].

In recent years, scientists are actively engaged in research to identify factors that affect life expectancy. Among them are the following main factors: natural conditions, ecological environment, nutrition, socio-economic development, health, education, and urbanization, etc. [4, 5, 6]. The value of this indicator gives a general assessment of the viability of the population as a whole.

The nowadays, the use of such statistical methods of analysis as correlation-regression analysis, the method of factor analysis, canonical analysis, the method of comparison of means, frequency analysis, etc. is widespread [6, 7]. However, these statistical methods have applicability limits and do not work with incomplete or "noisy" data. Neural network technologies and intelligent information systems developed on their basis can be used as an alternative approach for solving weakly structured and unformalized tasks.

The purpose of the study is to examine the relationship between life expectancy at birth of the population and socio-economic indicators in Russia with the use of neural network technology.

II. MATERIALS AND METHODS (MODEL)

The empirical base of the research is Rosstat data for 83 regions of Russia for 2016. Socio-economic factors affecting the expected life expectancy were selected using correlation analysis using the parametric method of calculating the Bravais-Pearson coefficients ($r > 0.5$). The system of indicators is grouped in blocks, which have the following form:

1. The health of the population:
 - x1- provision of hospital beds for 10 thousand people;
 - x2 - morbidity per 1000 population;
 - x3- provision of the population with doctors for 10 thousand people.
2. Social participation and public safety:
 - x4- output of qualified workers and employees, 10 thousand people employed;
 - x5- the ratio of marriages and divorces (there are divorces per 1,000 marriages);
 - x6- the number of road accidents in them per 100 thousand people;
 - x7- the number of registered crimes per 100,000 people;
 - x8- the total marriage rate per 1000 population.
3. Employment of the population:
 - x9- the unemployment rate, in percentages;
 - x10- the share of the employed population with higher education, in percentages.
4. Welfare of the population:
 - x11- average per capita income of the population, rubles;
 - x12- average monthly nominal accrued wages of employees of organizations, rubles;
 - x13- consumer spending on average per capita, per month. rub.;
 - x14- average size of assigned pensions, rub.;
 - x15- the amount of the subsistence minimum (on average per capita), rubles per month.
5. Ecology:
 - x16 - emissions of pollutants into the atmospheric air, thousand tons;
 - x17 - use of fresh water, million cubic meters;
 - x18 - discharge of contaminated sewage into surface water bodies, million cubic meters.

This study uses a clustered data analysis method based on a neural network. By the method of setting the weights of the neural network was chosen Kohonen self-organizing maps. The algorithm for constructing Kohonen self-organizing maps is one of the variants of clustering multidimensional vectors - the projection algorithm with preservation of topological similarity. An important distinguishing feature is that all neurons, including class centers, are ordered by a two-dimensional structure. In the process of training the neural network of Kohonen without a teacher, the weight coefficients of the neuron-winners and the connections of nearby neurons are modified.

Self-organizing maps of Kohonen allow you to project a multidimensional space into space with a lower dimension, to represent the projection of multidimensional data on a plane. Thus, the Kohonen mapping algorithm allows determining the location of clusters in a multidimensional space.

With the help of Kohonen self-organizing maps, data analysis was performed, clusters were recognized and hidden interdependencies were established between the indicators that

affect the expected life expectancy of the population of the regions of the country.

To conduct a cluster analysis of the life expectancy of the population of the Russian regions based on the Kohonen neural network, an information and analytical system was developed using the analytic platform Deductor. On this platform developed data warehouse designed to load data from various sources and build self-organizing Kohonen maps based on a neural network.

III. RESULTS AND DISCUSSION

In Fig. 1 presents the results of cluster analysis, where the regions of Russia are clustered, depending on the level of significance of the overall life expectancy at birth (Table 1). The level of significance was determined using the Student's t-test. Regions with a high level are assigned to the second cluster (5 regions), with a low level to the zero cluster (38 regions).

The profile of the second cluster is characterized by a high life expectancy at birth. The diagram shown in Figure 1 illustrates the average values of indicators in the context of clusters, which include: provision of hospital beds for the population, incidence per 1000 population, and the provision of the population with doctors for 10,000 people. The cluster is determined by the smallest values in comparison with other clusters, by the incidence of the population and by high values according to the indicator of the availability of the population by physicians.

The second cluster of the block "Social participation and security of the population" contains the smallest average indicators, where the largest differentiation is observed in terms of the output of skilled workers and employees per 10,000 employed population (Figure 2)

On the other hand, this cluster is characterized by the maximum values of the share of the employed population with

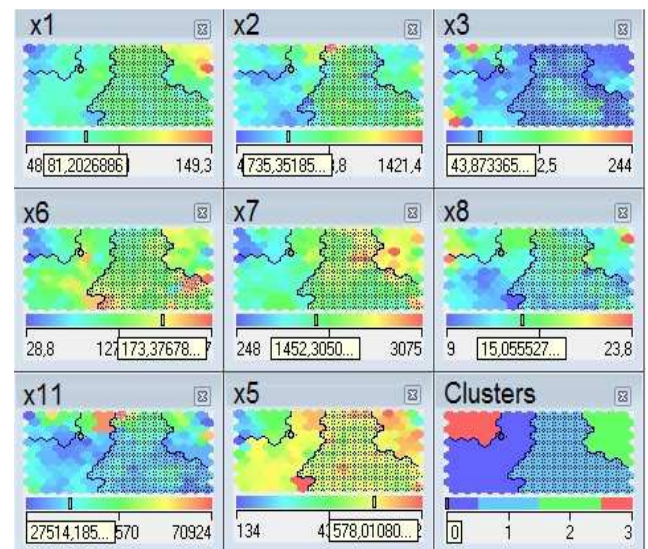


Fig. 1. Self-organizing maps of Kohonen (fragment)

higher education (Figure 3), which indicates a high employment of skilled workers and employees in the labor market in comparison with the regions belonging to other clusters.

One of the significant indicators is the unemployment rate for all activities, the maximum average values of the indicator

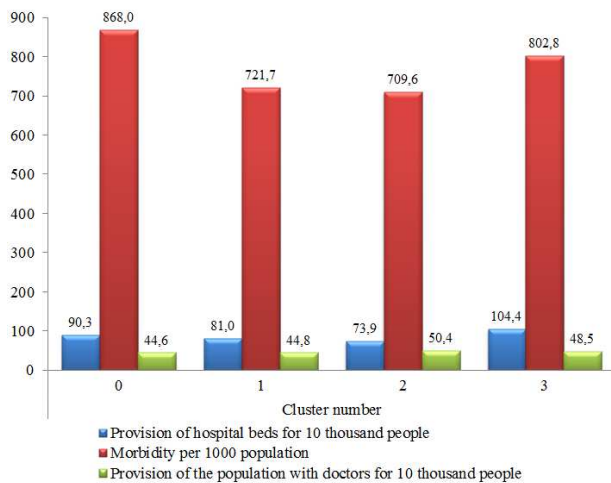


Fig. 2. The diagram of clusters in a cut of average values of indicators on the block "Health of the population"

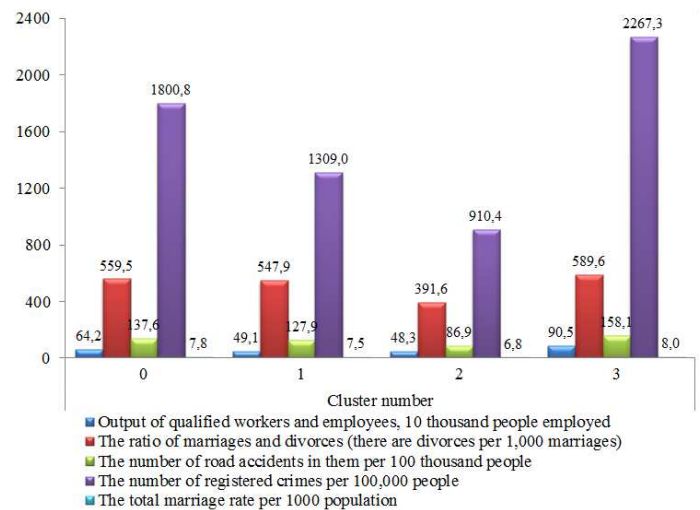


Fig. 3. The diagram of clusters in the context of the average values of indicators for the block "Social participation and public safety"

TABLE I. CLUSTER ANALYSIS OF EXPECTED DURATION OF LIVING OF THE POPULATION OF THE REGIONS OF RUSSIA

The level of significance of life expectancy at birth in Russia	
Cluster number	Region
<i>Low level of significance (38 regions)</i>	
0	Altai Territory, Arkhangelsk Region, Nenets Autonomous District, Bryansk Region, Vladimir Region, Vologda Region, Ivanovo Region, Kaliningrad Region, Kaluga Region, Kirov Region, Kostroma Region, Krasnoyarsk Territory, Kurgan Region, Kursk Region, Lipetsk Region, Murmansk Region, Nizhny Novgorod, Novosibirsk Region, Omsk Region, Orenburg Region, Oryol Region, Perm Territory, Primorsky Krai, Republic of Bashkortostan, Republic of Buryatia, Republic of Karelia, Komi Republic, The Republic of Mari El, the Republic of Sakha (Yakutia), the Samara Region, the Sverdlovsk Region, the Smolensk Region, the Tver Region.
<i>The level of significance below the average (23 regions)</i>	
1	Astrakhan Region, Astrakhan Region, Volgograd Region, Voronezh Region, Krasnodar Region, Leningrad Region, Moscow Region, Penza Region, Republic of Adygea, Republic of Kalmykia, Republic of Mordovia, Republic of Tatarstan, Rostov Region, Ryazan Region, Saratov Region, Stavropol Territory, Tambov Region, Tomsk Region, Tyumen Region, Khanty-Mansi Autonomous Area - Yugra, Chechen Republic, Chuvash Republic, Yamalo-Nenets Autonomous District.
<i>The level of significance is above the average (17 regions)</i>	
3	Amur Region, Jewish Autonomous Region, Transbaikal Region, Irkutsk Region, Kamchatka Region, Kemerovo Region, Magadan Region, Novgorod Region, Pskov Region, Republic of Altai, Republic of Tyva, Republic of Khakassia, Sakhalin Region, Khabarovsk Territory, Moscow, Saint Petersburg.
<i>High level of significance (5 regions)</i>	
2	Kabardino-Balkaria Republic, Karachay-Cherkess Republic, Republic of Dagestan, Republic of Ingushetia, Republic of North Ossetia-Alania.

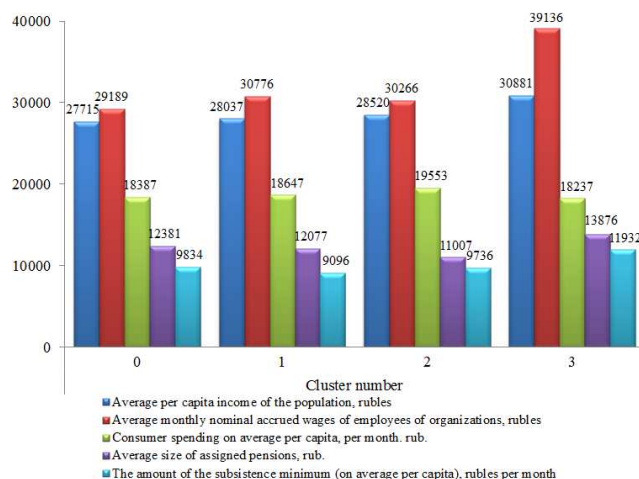


Fig. 4. The diagram of clusters in the context of average values for the block "Welfare of the population"

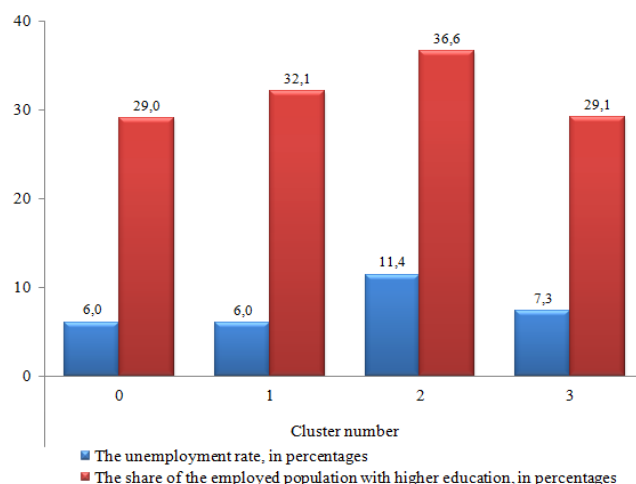


Fig. 6. The diagram of clusters in the context of average values for the block "Employment of the population"

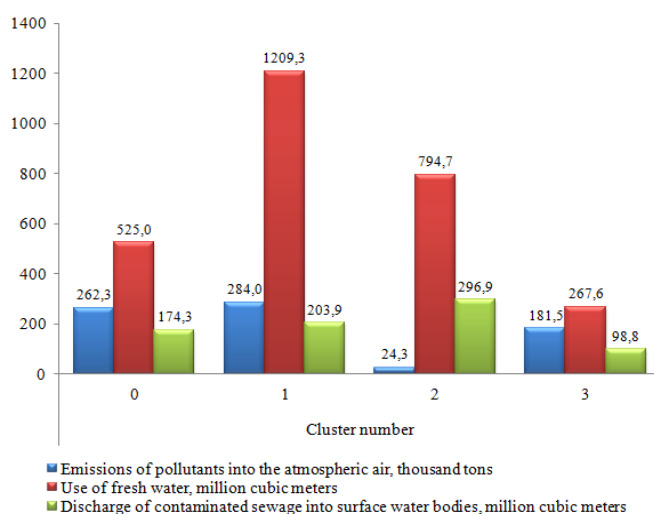


Fig. 5. The diagram of clusters in the context of the average values of the indicators for the block "Ecology"

The diagram of clusters in the context of average values of the level of incomes and expenditures of the population is presented in Figure 4. The regions included in the third cluster are average for all indicators of the "Welfare of the population" block. All other clusters have the same differentiation of the average values of incomes and expenditures of the population.

Analyzing the influence of the environmental factor on the expected life expectancy, it was revealed that the regions of the second cluster (Figure 5) are determined by low average values for the number of pollutant emissions into the surrounding atmospheric air and sufficiently high values for the use of fresh water, the leader in this indicator are the regions of the first cluster.

IV. CONCLUSION

Studies have shown that the regions included in the third cluster are characterized by the largest average values for the "Well-being of the population" block in comparison with other clusters, but they are inferior to the regions belonging to the second cluster in terms of life expectancy at birth. This indicates about hidden or not explicit dependence of the incomes and expenditures of the population in relation to the expected life expectancy.

Thus, there is a strong enough link between life expectancy at birth and social and economic indicators, where factors such as the incidence of the population, the level of the employed population, the level of unemployment, and also the factors characterizing the ecological situation in the regions have the greatest impact.

References

- [1] Presidential Decree of 11.14.2017, № 548 [Ukaz prezidenta RF ot 14.11.2017 g. № 548] [Electronic resource]. URL: publication.pravo.gov.ru/File/GetFile/0001201711150027?type=pdf (circulation date 11/26/2017).
- [2] United Nations Development Program: Human Development Index in the World in 2015 [Programma razvitiya OON: indeks chelovecheskogo razvitiya v stranakh mira v 2015 godu] [Electronic resource]. URL: <http://gtmarket.ru/news/2015/12/16/7285> (circulation date 26.11.2017).
- [3] Pandey M.D., Nathwani J. S. Measurement of Socio - Economic Inequality using the life-quality Index // Social Indicators Research. - 1997. - Vol. 39. - pp. 187 - 202.
- [4] Gupta S., Verhoeven M., Tiongson E. Does Higher Government Spending Buy Better Results in Education and Health Care? // IMF Working Paper. - 1999. - № 99/21. - 25 p.
- [5] Jayasuriya R., Wodon Q. Efficiency in improving health and education outcomes - provincial and state-level estimates for Argentina and Mexico // Estudios Económicos. - 2007. - Vol. 22. - № 1. - pp. 57 - 97.
- [6] Chukmaitova A. Determinants of life expectancy and mortality: comparative analysis of different regions in Kazakhstan / Working Paper. BSP/2003/072 E. 2003.
- [7] Grigoriev P., Shkolnikov V., Andreev E., Jasilionis D., Jdanov D., Meslī F., Vallin J. Mortality in Belarus, Lithuania and Russia: Divergence in Recent Trends and Possible Explanations // European Journal of Population. - 2010. - Vol. 26. - № 3. - pp. 245 - 274.