USE OF BIG DATA TECHNOLOGY IN PUBLIC AND MUNICIPAL MANAGEMENT

V.M. Ramzaev[1], I.N. Khaimovich[1,2], V.G. Chumak[1]

[1] International Market Institute, Samara, Russia
[2]Samara National Research University, Samara, Russia

Abstract. The development of a model for forecasting the competitiveness of territories requires using large amounts of stream data in real-time. The aim of this study is to develop the models and methods of management decision-making based on forecasting of competitiveness of territories. The objectives of this study include the identification of competitiveness factors, the development of a method of SMB management, the development of a model of competitiveness of territories using expert estimates, the presentation of information from the experts using the BIG DATA technology. The results of the study are models for making management decisions on competitiveness of territories using expert estimates and applying the BIG DATA technology. Practical outcomes include improving the quality and timeliness of making decisions on management of territories based on the model of forecasting of the region development.

Keywords: competitiveness, territory management, intensive data, mathematical models, clusters.

Citation: Ramzaev VM, Khaimovich IN, Chumak VG. Use of Big Data technology in public and municipal management. CEUR Workshop Proceedings, 2016; 1638: 864-872. DOI: 10.18287/1613-0073-2016-1638-864-872

Introduction

Information technologies are widely used in state and municipal management. The study covers the issues related to application of BIG DATA technology to territory management and SMB management. Both tasks are associated with the use of the “competitiveness” criterion. Competitiveness is an essential characteristic feature of the development of social and economic systems, including the territories. This area is one of the priorities for the R&D Centre of the International Market Institute. For a number of years we have been studying the competitiveness of territories: the region, the cities, including towns and single-industry towns, municipal districts, rural settlements.

Our approach is based on understanding the competitiveness as an ability to compete for limited resources [1,2].

Evaluation Model territory competitiveness

Our method is based on an economic and mathematical model of additive type for the evaluation of a territory competitiveness:

$$\begin{cases} KS = (\xi_1 GF + \xi_2 PRF + \xi_3 EF + \xi_4 PPF + \xi_5 APF + \\ +\xi_6 SF + \xi_7 FEF + \xi_8 IfF + \xi_9 UVF + \xi_{10} IF + \\ +\xi_{11} InF + \xi_{12} DF) \to \max \\ 0 \le \xi_i \le 1, i = \overline{1,12} \\ \sum_{i=1}^{12} \xi_i = 1 \\ 0 \le GF \le 1; 0 \le PRF \le 3; -2 \le EF \le 1; -3 \le PPF \le 12; \\ 0 \le APF \le 6; -3 \le SF \le 29; 0 \le FEF \le 11; \\ -2 \le IfF \le 13; 0 \le UVF \le 1; 0 \le IF \le 2; \\ 0 \le InF \le 3; 0 \le DF \le 5. \end{cases}$$

where *KS* is competitiveness; *GF* – geographic factor; *PRF* – natural-resource factor; *EF* – ecological factor; *PPF* – industrial production factor; *APF* – agricultural business factor; *SF* – social factor; *FEF* - financial and economic factor; *IfF* – infrastructural factor; *UVF* – the factor of the level of interaction with the superior authorities; *IF* – innovation factor; *InF* – investment factor; *DF* – mental factor; ξ – factor significance coefficient (defined based on the opinion of experts).

During the study, we have determined 12 factors of competitiveness typical for the modern level of social and economic development of the territories. Each of the factors has its significance, which determines its importance and contribution to the final value of competitiveness. The importance of factors vary for different types of territories, which reflects differentiation in the current state of the development process.

Since cogitability and visual expression are important for making management decisions, we offer multi-level visualization of analysis results and competitiveness evaluation. By choosing the dimensionality of the space we can illustrate the level and contribution of particular competitiveness factors for the purposes of management [3,4].

However, it is obvious that sustainable competitive development of economy cannot last forever. For example, at one of the territories the development is limited by energy and raw material resources of the planet and we are approaching the maximum of their use. There are other theories, according to which the development curve is approaching the saturation area [5].

Despite that, the competition remains a crucial factor stimulating the development and qualitative growth of social and economic systems. Hence, the application of the above state models, actually based on extensive direct addition of component factors, has its limitations when used for the purposes of management. In the areas being close to saturation such models involve poor accuracy or are inadequate.

Taking into account the above, we offer a more precise method of competitive development management based on correlations of competitiveness factors, which allows to define the vectors of management according to the key target parameters. To a

certain degree, this approach is similar to the method of evaluation of synergistic effect. But the latter is rather complicated in terms of numerical calculation of values.
Six groups of correlated factors were determined during the study.
From a practical perspective, the factor-based grouping allows to manage competitiveness more efficiently, as the maximum growth of the competitive development level can be achieved only by joint adjustment of factors in the groups.
After analyzing the correlations, we defined that the only factor that correlates with all the other factors is the investment factor. For this reason, the target function of management based on increasing the competitiveness was complemented with the model of controlling actions, the latter being represented by the limited investment resource:

$$\begin{cases} \Delta KS = 0,058(GF+\Delta GF(L))\times 0,072(PRF+\Delta PRF(L))\times \\ \times 0,064(EF+\Delta EF(L))\times 0,011(PPF+\Delta PPF(L))\times \\ \times 0,075(APF+\Delta APF(L))\times 0,115(SF+\Delta SF(L))\times \\ \times 0,113(FEF+\Delta FEF(L))\times 0,076(IfF+\Delta IfF(L))\times \\ \times 0,057(UVF+\Delta UVF(L))\times 0,101(IF+\Delta IF(L))\times \\ \times 0,104(InF+\Delta InF(L))\times 0,055(DF+\Delta DF(L)) \rightarrow \max \\ \Delta ...F = ...F(L_m(\bar{\phi}) - (...F) \\ L_m(\bar{\phi}) = \sum_{n=1}^{N}\sum_{m=1}^{M} \frac{PV_n^{\inf}}{(1+r^{\inf})^n}\cdot\frac{1}{IR}\cdot\phi_{im}, m=\overline{1,12}; i=\overline{1,12}; \\ 0 \le GF \le 1; 0 \le PRF \le 3; -2 \le EF \le 1; -3 \le PPF \le 12; \\ 0 \le APF \le 6; -3 \le SF \le 29; 0 \le FEF \le 11; \\ -2 \le IfF \le 13; 0 \le UVF \le 1; 0 \le IF \le 2; \\ 0 \le InF \le 3; 0 \le DF \le 5. \end{cases}$$

where ΔF is the change of discounted effect of each competitiveness factor of a municipal unit.
The models, which take into account the correlation, are effective when applied for the choice of investment projects at the development of territories, including the competition-based distribution, as they are oriented at taking into account not only the direct financial results, but the correlated indirect effects of increasing the competitiveness and its particular factors. For example, ecological, social and other.

Method of SMB development management for investment factor

The method of SMB development management can be applied to develop the investment factor in the Samara region.
However, the development of modern economy is unstable – kick-starts give way to deceleration and vice versa.
For example, the decrease in GDP of European economy amounted to 6% in 2008-2009, 1.5 % in 2011, 0.2% in 2012. In 2013, the GDP of European Union is forecasted to grow 0.6%, in 2014 – 1.2%. The Eurozone budget gap decreased from 4.2% of

GDP in 2011 to 3.7% in 2012. Further decrease of deficit to 2.8% of GDP is expected in 2013. That said, the government debt grown up to 91% of GDP in 2012 from 87% in 2011. The unemployment rate is estimated at 4 to 27% in various countries of European Union.

The Russian Economy, being less stable, has shown changes that are even more evident. GDP growth for the period from 2001 to 2008 amounted to 6.6%. In 2009-2011 there was a drop to 0.2% and an increase again up to 3.4% in 2012. The government debt amounted to minus 9.5% of GDP in 2011 and 3% in 2012. The 4% budget deficit of 2010 turned into a 0.8% profit in 2011. But the positive dynamics was not maintained and the year 2012 ended with a deficit of 0.02%.

Notably, these processes take place in a highly saturated competitive environment. So, it is evident that such situation is favorable for the territories, clusters and companies, which are highly sensitive systems – they react to the changes quicker and thus improve their competitiveness. In such conditions the traditional linear models cannot ensure fulfilment of management objectives for the following reasons:

- they are situation-related and can be used in short time intervals, which does not allow to perform long-term and strategical management;
- they do not take into account the speed of response to the control action;
- modern complicated multi-aspect relations and processes are nonlinear.

The most efficient competitiveness management in such environments can be achieved on the basis of dynamic models which are only coming into use in the contemporary economics.

As can be seen from the above, we understand the notion of competitiveness as a dynamic characteristic defined by the speed of system response to any changes of the external social and economic environment.

Since the size of this article is limited, we will provide the results of dynamic modelling illustrated by social and economic systems of industrial clusters.

Within the new approach that we offer, the determined competitiveness factors are divided in 3 dominants taking into account their correlations: production, labor resources and investments. During the process of modelling, cluster units shall be separated from the general economic system of the region, i.e. their borders shall be defined.

When analyzing the territorial cluster units from the manageability point of view, we shall define 2 key types:

1) Functional or manageable cluster, which:

- appears as a result of deliberate external influence in the areas of strategical importance for the country and in the course of implementation of strategic plans;
- has financial, economic and political support of the state;
- as a rule, has nuclear structure.

2) Self-organized or business cluster, which:

- is not a result of actions of state authorities;
- appears spontaneously, at the initiative of business and on the basis of economic relations;
- is not managed and has no institutional partners which guarantee its survival;
- as a rule, it has matrix structure.

As the cluster is an open dynamic system, its borders are unclear, which resulted in using a frame of fuzzy sets and fuzzy logic. This way the degree of membership of an element to the system is determined on the basis of necessary and sufficient conditions of cluster existence. Application of methods of fuzzy logic allows to find the areas where the clusters overlap. These are the zones of particular innovative capacity, which can provide qualitative breakthroughs in the development of cluster systems.

This objective shall be approached using modern information technologies. One of such technologies is BIG DATA directly associated with data mining [6,7]. At the same time, the use of modern BIG DATA technologies allows to highlight the areas of active consumption of goods and services, which can be rapidly developed and supplied to the market.

A special method based on BIG DATA was developed to manage the development of SMB in the region. The method includes the following stages:

1. determine the role and place of small business in the region;
2. define the main types of goods and services offered by small business in the region;
3. create the image of a customer using the services of small business in the region based on mathematical modelling in the form of models of correlation and regression analysis [8,9] or simulation modelling [10,11];
4. create an informational model of an SMB customer in the region;
5. form the zones of small business in the region;
6. develop recommendations on making management decisions.

The role and place of SMB in the region, major types and services offered by the entrepreneurs in the region were analyzed in the beginning of the article, as for creating the image and informational model of the customer, it requires using the BIG DATA technology. The method of using data mining includes the following:

1. form the area of BIG DATA in Hodoop [12,13,14] from twitter using the filter “Samara region” showing the hit count;
2. divide the obtained memory area according to various filters connected with the basic factors of small business;
3. perform monitoring of stream content analysis according to filters;
4. take operative measures in the cases of stable spikes in the number of hits;
5. develop a program on Scala to work with filtering in the area of BIG DATA;
6. perform debugging and testing of the program and gather practical data;
7. analyze the calculation results.

The “twitter” software is used for data analysis, as it is an open-source product, its use does not require additional investments, and 50% of internet users have profiles on twitter. With the BIG DATA technology it is possible to store and update the data in “Hodoop” area with the filter “Samara region” (filter1= {Samara region}). Then this area shall be filtered according to the base factors of SMB by setting the following filters e.g.:

Filter2 (meals) = {cafe, bar, restaurant, cooking, beer, meat, fish, pub};
Filter3 (clothes)= {jacket, blouse, dress, skirt, bra, stuff};
Filter4 (entertainment)= {nightclub, concert, session, hangout};
Filter5 (children) = {kindergarten, baby-club, sports club, study group}.

We get the diagrams of dependency of the number of hits by filters on the time of data collection (Fig. 1). The time of data collection from the internet is unlimited in the BIG DATA technology [15,16,17].

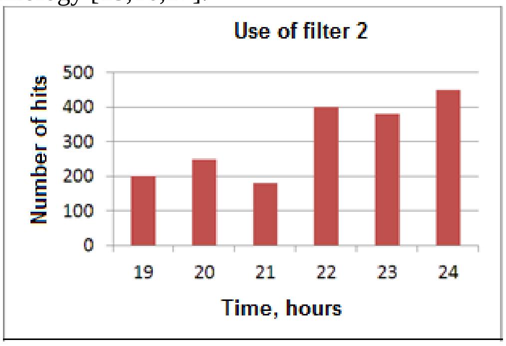

Fig. 1. Diagram of dependency of the number of hits by filter on the time of data collection

As a result, we get dynamic change of information from the internet in real time, which allows to monitor the stream analysis of unstructured information (In-Memory Data Processing and Stream Technology) by filters [18,19]. In order to implement this method a program on Scala [20,21] was developed:

```
val file = spark.textFile("hdfs://… ")
val errors=file.filter(line=>line.contains("Samara region"))
//count all the data
errors.count()
//count data mentioning Filter
errors.filter(line=>line. contains("meat")).count()
//Fetch the filter as an array of string
errors.filter(line=>line. contains("meals")).collect()
```

After the program operation we get the diagrams of dynamic change of parameters in the BIG DATA environment (Fig. 2), which allow to define the areas of SMB in the region according to the analysis of unstructured information.

If stable spikes are detected in the number of hits at the diagrams according to the forms of entrepreneurship, than investment support shall be provided for the development of SMB in this particular type of activities in the analyzed area.

If we proceed to look into competitiveness of territories, than the models of competitiveness dynamics shall be used to make management decisions. Dynamic models of competitiveness of functional and business clusters have a particular appearance.

Then we define the starting points of competitiveness management for the regional industrial clusters of the Samara region. For this purpose, the target function of the state model of slide 1 is supplemented with a set of CL parameters developed by us. The set of CL parameters includes:

1) Cluster type according to the manageability criterion;
2) Cluster type according to the development dynamics;
3) Type of cluster structure;
4) Producers of the key products – cluster leaders.

As a result, system features of industrial clusters of the region are formed, based on them, it is possible to evaluate the necessity and degree of control actions and to apply corresponding management models

$$\frac{\partial u_i}{\partial t} = c_i u_i + \sum_{j \neq i}^{n} d_{ij} u_j - \sum_{j \neq i}^{n} b_{ij} u_i u_j + D_i \Delta u_i i = \overline{1,n} \ ,$$

where the elements with the coefficients *dij* describe the dependence of production in the *i*–th element on the production in other elements of the cluster; the elements with the coefficients *bij* take into account the competition between the producers.
The dynamic modelling of social and economic systems on the territory allowed to:

- determine the steady states of the system being the target results of management;
- evaluate the state variables for the system in case of changing some of its parameters, i.e. monitor the management effect;
- evaluate the degree of approximation of the current state of the system to the present target values and choose the most efficient path for the particular conditions.

Despite the obvious complexity of models, the user-level use of application software allows to interpret the results and define management decisions fairly simple.
In order to guide a territorial system, e.g. an industrial cluster, into the area of sustainable development of competitiveness, it is necessary to adjust the parameters of cluster system. Furthermore, it was determined that some parameters are fairly inert. Such parameters include the length of a production cycle, staff rotation rate, tax liabilities etc. Other parameters have higher dynamics. These include extensive labor efficiency, cost per unit etc. The most efficient management parameters are nonlinear parameters, that is: intensive labor efficiency, which growth is ensured by innovations and introduction of new technologies, as well as employee displacement as a result of intensive growth of labor efficiency.
Application of dynamic models allows to balance the adjusted parameters, to evaluate the required degree of impact, target results and the rate of their achievement, and in the aggregate it gives an advantage in managing the competitiveness of a social and economic system.

Conclusion

The development of a model for forecasting the competitiveness of territories requires using large amounts of stream data in real-time. The aim of this study is to develop the models and methods of management decision-making based on forecasting of competitiveness of territories. The objectives of this study include the identification of competitiveness factors, the development of a method of managing SMB in the region, the development of a model of competitiveness of territories using expert estimates, the presentation of information from the experts using the BIG DATA technology. The results of the study include the models for making management decisions on competitiveness of territories using expert estimates and applying the BIG DATA technology. Practical outcomes include improving the quality and timeliness of making decisions on management of territories based on the model of forecasting of the region development.

References

1. Ramzaev VM, Kukolnikova EA, Khaimovich IN. Development of a functional model of active production elements in the regional management. Bulletin of Samara State University of Economics, 2014; 12: 87-99.
2. Ramzaev VM, Khaimovich IN. Integrated model of management of economic development of the region on the basis of increasing the competitiveness of companies. Contemporary issues of science and education, 2014; 6:136.
3. Ramzaev VM, Khaimovich IN, Chumak VG. Issues of data access in economic studies using the Big Data technology. Proceedings of the International Conference and School for Youth "Information Technology and Nanotechnology". Samara, Samara State Aerospace University, 2015: 147-152.
4. Ramzaev VM, Khaimovich IN, Chumak PV. Models for forecasting competitive growth of companies during energy modernization. Forecasting issues, 2015; 1: 67-75.
5. Bonacich P. Power and Centrality: A Family of Measures. American Journal of Sociology, 2007; 92(5): 1170-1182.
6. Hey T, Tansley S, Tolle K. The Forth Paradigm: Data-Intensive Scientific Discovery. Redmond, Microsoft Research, 2009.
7. Kalinichenko LA, Briukhov DO, Martynov DO, Skvortsov NA, Stupnikov SA. Mediation Framework for Enterprise Information System Infrastructures. Proc. of the 9th International Conference on Enterprise Information Systems (ICEIS-2007). Serial "Databases and Information Systems Integration", Funchal, 2007: 246-251.
8. Chumak PV, Ramzaev VM, Khaimovich IN. Models for forecasting the competitive growth of enterprises due to energy modernization. Studies on Russian Economic Development, 2015; 26(1): 49-54.
9. Chumak VG, Ramzaev VM, Khaimovich IN. Challenges of Data Access in Economic Research based on Big Data Technology. CEUR Workshop Proceedings, 2015; 1490: 327-337.
10. Drovyannikov VI, Khaimovich IN. Development of a set of models for managing the competitive development of social cluster of the region. Fundamental studies, 2015; 7(4): 822-827.
11. Drovyannikov VI, Khaimovich IN. Simulation modelling of managing a social cluster in the system Any Logic. Fundamental studies, 2015; 8(2): 361-366.
12. White T. Hadoop: The Definitive Guide. O'Reilly Media; Third edition, 2012.
13. Saracco C, Jain U. What's the big deal about Big SQL? Introducing relational DBMS users to IBM's SQL technology for Hadoop. IBM DeveloperWorks, 2013. URL: http://www.ibm.com/developerworks/library/bd-bigsql/bd-bigsqlpdf.pdf
14. Capriolo E, Wampler D, Rutherglen J. Programming Hive Data Warehouse and Query Language for Hadoop. O'Reilly Media, 2012.
15. Schaar P. The Internet and Big Data – Incompatible with Data Protection? Mind – Multistakeholder Internet Dialog. Berlin, Internet & Society Collaboratory, 2014; 7: 14-20.
16. Akyildiz IF, Jornet JM, Pierobon M. Nanonetworks: A New Frontier in Communications. Communications of the ACM, 2011; 54(11): 84-89.
17. Llatser I, Cabellos-Aparicio A, Alarcon E. Networking Challenges and Principles in Diffusion-based Molecular Communication. IEEE Wireless Communications, 2012; 19(5): 36-41.
18. Toporkov V, Tselishchev A, Yemelyanov D, Potekhin P. Metascheduling Strategies in Distributed Computing with Nondedicated Resources. Dependability Problems of Com-

plex Information Systems, Advances in Intelligent Systems and Computing (AISC). Switzerland, Springer International Publishing, 2014; 307: 129-148.
19. Toporkov V, Toporkova A, Tselishchev A, Yemelyanov D. Slot Selection Algorithms in Distributed Computing. Journal of Supercomputting, 2014; 69(1): 53-60.
20. Beyer KS, Ercegovac V, Gemulla R, Balmin A, Eltabakh M, Kanne C, Ozcan F, Shekita EJ. Jaql: A Scripting Language for Large Scale Semistructured Data Analysis. VLDB, 2011.
21. Hernandez M, Koutrika G, Krishnamurthy R, Popa L, Wisnesky R. HIL: a high-level scripting language for entity integration. Proceedings of the 16th International Conference on Extending Database Technology. EDBT, 2013: 549-560.