Research Journal of Applied Sciences 10 (8): 411-414, 2015
ISSN: 1815-932X

Data Intellectual Analysis Means Use for Condition Indicators Assessment of the Territorial and State Formations

Gennadiy V. Averin, Anna V. Zviagintseva, Igor S. Konstantinov and Olga A. Ivashchuk
Belgorod State University, Pobedy Str. 85, 308015 Belgorod, Russia

Abstract: It is shown that search of communications and regularities in databases of the countries and regions development indicators can be carried out between probabilities of events connected with supervision of several main development indicators. The approaches to creating data models and algorithms for the intellectual analysis of information based on definition of statistical probabilities of such events are offered. The example of countries development assessment on the basis of UN Development program databases use is given. Proceeding from the offered methods, the level of human development in Russia in comparison with other countries is estimated and ranks of various countries are determined by the level and rates of development for the last 5 years. This approach is alternative to known approach which is accepted at calculation of an index of human development and differs in use of expert methods.

Key words: Countries and regions, state and development, strategic assessment, data intellectual analysis, index

INTRODUCTION

Now a days, many scientific branches involved themselves in extensive databases and information storages creation. In the field of climatology, geoecology, social and economic and human development the considerable volume of approved facts and statistical information, etc. has been gathered (Jon, 2014; Anonymous, 2010, 2014). Thus, development of computing science creates conditions for search of the hidden regularities in information to what methods of the Data Intellectual Analysis (DIA) are applied. Though application of the DIA methods for many tasks solution has no alternative yet, these methods have a number of shortcomings (Averin, 2014):

- Usually the DIA methods are poorly connected with features of applied area and the analysis of data most often logical methods of "blind" search of regularities are the cornerstone
- Search of regularities in data is connected with big intellectual work of researchers
- It is not enough algorithms of DIA which use high levels of containing data information understanding and also apply various models of the data studied

Nevertheless, today it is possible to speak about creation of models of the big databases given in relation to some types. For example, the Palantir Technologies Company in the products uses ontologic data models. Quid Inc. Company. applies model which received the name "technocenosis" in a known Quid data analysis product. In this study, applied areas, it is offered to use phenomenological models of data in the form of probabilistic environment of modeling; for those in each case model dependences and parameters are established.

TECHNIQUE AND DATA

Developments which are connected with creation of the theory of system dynamics and algorithmic methods of the statistical data analysis on countries development (Averin, 2014; Forrester, 1961; Goodman, 1989; Gallo and Ertur, 1980; Hanneman, 1988; Averin and Zvyagintseva, 2013) are the basis for this work. Such models may base on algorithms of statistical assessment of probabilities of system conditions. The corresponding probabilities are determined by data files for difficult events of simultaneous supervision of several indicators which indicatively reflect development of the countries. It can be both separate characteristic events and different combinations of several such events representing one complex event for example, a joint event of supervision of three or four development indicators. As a result, we come to need of development of the DIA methods which would consider probabilistic regularities of indicative data distribution. Application of similar approaches in nature and society sciences is of great importance since, it allows to offer objective methods of system research for n-dimensional dimension which all are public, economic, ecological and global systems.

Corresponding Author: Gennadiy V. Averin, Belgorod State University, Pobedy Str. 85, 308015 Belgorod, Russia

Let us illustrate this approach on the example of a technique development which is alternative to a known method of human development index calculation by UN Development Program (UNDP) (Jon, 2014; Anonymous, 2010, 2014).

In 2010, UNDP technique for calculation of human development index was changed (Anonymous, 2010). As attributive variables for calculation of an index the following indicators began to be used: average duration of training (p_2, years; the expected duration of training (p_1), years; the Gross National Income (GNI) per capita in recalculation at par Purchasing Power (PPS) in US dollars (p_3), a remaining life expectancy (p_4) years. For the objective solution we will use UNDP Reports databases. (Anonymous, 2010, 2014).

MAIN POINTS

On the basis of variables p_1, p_2, p_3, p_4, we will create four-dimensional space of coordinates in which possible conditions {p_1, p_2, p_3, p_4} of the studied system theoretically form a certain area Q_4 covering all observed database points. In this case, the condition of each country can be presented by a four-dimensional point, M (p_1, p_2, p_3, p_4) and processes of condition change for the countries for a certain period multidimensional lines. Let us assume an area Q_4 continuity. It means that in space Q_4, there is an infinite set of states for some population of objects (countries) and a point of states M (p_1, p_2, p_3, p_4), continuously fill this space. Also, we will consider that skilled points from database are limited selection of supervision of this population.

Let us consider a joint event of simultaneous supervision of four indicators stated above and we will define that the condition of each country of the world will be defined by this event. Let us find statistical probability, of this event on the basis of data which are available in database (Anonymous, 2010, 2014), with use of algorithms of search, group and calculation of favorable event frequencies (Averin, 2014; Forrester, 1961; Goodman, 1989; Gallo and Ertur, 1980; Hanneman, 1988; Averin and Zvyagintseva, 2013). Let us call this probability, the probability of a system condition. This probability w is counted in all group of objects. In the report, development of 169 countries of the world was analyzed.

For creating models of quantitative data we accept a hypothesis of a probability continuity of system condition in area Q_4. In other words, we assume existence of a scalar field of statistical probability w in space of Q_4 type w = w (M) Let us assume that in area Q_4, it is possible to set analytical continuous function T(p_1, p_2, p_3, p_4) on the basis of which the model of probabilistic space will be formed. At a known type of function T(p_1, p_2, p_3, p_4) and values of variables p_1, p_2, p_3, p_4 in area Q_4, it is possible to construct one more scalar field which we will call the modeling environment. When developing algorithms of DIA the environment of modeling was represented by the dependences entering classes of uniform or multiplicative functions of rather attributive variables. It is established that under these conditions variable T(p_1, p_2, p_3, p_4) in space Q_4 allows to use at phenomenological descriptions the quasilinear multidimensional equations in private derivatives of the first order which are closely connected with Pfaff's equations. For creation of probabilistic algorithms, it is supposed that in space of states Q_4 scalar fields of sizes w and T are unambiguously connected among themselves. This communication is presented in the form of phenomenological ratios $dw = c_1 \times dT$ where c_1 are empirical sizes which are development functions.

If hypotheses accepted above are true then phenomenological data descriptions presented by tabular and time arrays of information are closely connected with Pfaff's equations of type:

$$dw = c_1 \times \frac{\partial T}{\partial p_1} dp_1 + ... + c_4 \times \frac{\partial T}{\partial p_4} dp_4$$

For this equation of Pfaff in space Q_4, there is some field of the directions generated by a scalar field of statistical probability w and which is characterized by vector lines of this field (Averin, 2014). Also for this equation in space Q_4, there is the general potential of type $P(p_1, p_2, p_3, p_4) = C$ which represents a surface, orthogonal to vector lines of field. This potential P can be accepted as the generalized criterion for a complex assessment of condition of the countries of the world in multidimensional space Q_4. This size is function of a state is correct for living conditions of a scalar field of statistical probability w. Change of potential depends only on an initial and final condition of the country and does not depend on a way of its transition between these states.

With use of social and economic data of UN Development Program and World Bank as an example the technique of the information analysis and the DIA probabilistic methods was developed and also possibility of construction in space of many variable phenomenological models of the data characterizing development of the countries is shown. In the course of research performance on creation of data models, it is necessary to carry out large volume of calculations. It is connected with need of searching for various modeling environments, algorithmic definition of events statistical

probabilities, studying of various communications for arrays of information and definition of phenomenological constants, search of optimum models for the data description, etc. Therefore, one of the main objectives demanding the decision consists in development of an analytical platform in the form of the computing environment for modeling of the world countries development. Creating such platform will give the chance to increase efficiency of scientific activity, activity of analysts and experts in the analysis of processes of world and regional growth and also the solution of actual global studies problems.

EXAMPLE OF HUMAN DEVELOPMENT LEVEL ASSESSMENT OF THE COUNTRIES DURING 2008-2013

The received results allow to offer an objective method of countries development assessment of the world and to construct system of forecasting of their indicators, without leaning on expert methods. For example, taking into account a probabilistic assessment of the joint events connected with supervision of indicators ranks of development of the countries of the world were defined.

The first ten countries having the highest level of development in 2013 are: Qatar, Liechtenstein, Kuwait, Singapore, Brunei, Norway, Luxembourg, United Arab Emirates, Switzerland and Hong Kong. From the Great Twenty (G20) of the world countries in this list there is no country, and from the EU there is only Luxembourg, and Germany on 14 is the USA on the 11th place. On rates of development during 2008-2013 there are ten quickly developing states: Qatar, Kuwait, Singapore, Brunei, Saudi Arabia, Switzerland, Liechtenstein, Luxembourg, Hong Kong and Iceland.

Guinea, Guinea-Bissau, Mozambique, Niger, Liberia, Malawi, Burundi, Ethiopia and the CAR are among ten countries having the lowest level of development. All these countries are in Africa. On rates of development in 2008-2013, the most lagging behind countries were: Uganda, Niger, CAR, Malawi, Solomon Islands, England, Equatorial Guinea, Greece, Bahamas and Barbados. In this list, there are England and Greece entering into group of EU countries. England and Greece iare the two European countries which during the studied period had a decrease in specific GNI.

Russia in the level of development in 2013 rating took the 44th place in turn, Kazakhstan 53, Belarus 62, Ukraine the 97th place. On rates of development in 2008-2013 Russia was on the 24th place, Kazakhstan on 25, Belarus on 52 and Ukraine on the 95th place. For the same period Russia advanced Italy, France, Japan, England, Canada, Estonia in rates of development, however lagged behind the USA, Germany, Sweden, Lithuania, Latvia.

SUMMARY

This method can be applied to any set of initial indicators, however with increase in number of indicators over 5-7 operating time of algorithms of DIA increases in the analysis of information in multidimensional space. In general labor input of the method is defined by need for creating a set of models for assessment of probabilities of the most different events. If the UNDP database contains some tens indicators demanding the analysis, the database of the World Bank contains already more than one thousand indicators almost for 200 countries of the world.

Thus, databases of indicators along with methods of a complex assessment and algorithms of DIA by definition of probabilities of events allow to determine consistent patterns of development of the countries.

CONCLUSION

The ratings of the world countries on level of development determined by method of UNDP human development index calculation significantly differ from the similar rating calculated on the basis of this approach. For the majority of developed countries the technique of UNDP gives the overestimated ratings of development level and for many developing states the underestimated ratings. Among the countries which on the potential of development P got to the first ten, only Singapore and Switzerland are available in a rating of Top-10 of UN Development program. All this says that the index of human development expertizes more favorable integrated for the countries of “gold billion” (the USA, Canada, Australia, Japan, the European Union countries) and less favorable for all other countries.

REFERENCES

Anonymous, 2010. Human Development Report. The real wealth of Nations: pathways to human development, pp: 228.

Anonymous, 2014. Human Development Reports (1990-2013 years) 25.11.2014. Search the Human Development Report Office’s Library. http://hdr.undp.org/en/reports/.

Averin, G.V. and A.V. Zvyagintseva, 2013. Application of methods of the data intellectual analysis for assessment of development of Ukraine. Collection of Scientific Works Geotechnical Mechanics, 112: 257-270.

Averin, G.V., 2014. System dynamics. Donetsk: Donbass, pp: 405. http://www.chronos.msu.ru/ru/rnews/item/sistemodinamika.

Forrester, J.W., 1961. Industrial Dynamics. MIT Press. New York, pp: 363-386.

Goodman, M., 1989. Study Notes in System Dynamics. Pegasus.

Gallo, J. and C. Ertur, 1980. Exploratory Spatial Data Analysis of the distribution of regional per capita GDP in Europe, 1980-1995. University of Burgundy 04.03.15. http://www.u-bourgogne.fr/LATEC.

Hanneman, R., 1988. Computer-assisted theory building. Modeling dynamic social systems. SAGE. N.Y.

Jon, H., 2014. Human Development Report "Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience" 25.11.2014. http://hdr.undp.org/en/.