Journal of Engineering and Applied Sciences 12 (15): 3884-3889, 2017
ISSN: 1816-949X

On Representation of Discrete Information of Temporal Databases in the Continuous Form

Gennadiy V. Averin, Anna V. Zviagintseva, Maria V. Shevtsova and Liliana N. Kurtova
Belgorod State University, Pobedy Street 85, 308015 Belgorod, Russia

Abstract: The representation problem in a continuous form of discrete quantitative data on the basis of development of the phenomenological models characterizing changing processes of difficult system's conditions is studied. The method of such model's creation is offered according to temporal databases using the idea of the description these states given in continual spaces. At the same time discrete quantitative information is considered in the form of multidimensional selection of data experience from the continuous hypothetical environment characterizing object's conditions on a set of indicators which in turn are considered as variable conditions of objects. Dependences for the phenomenological analysis of experimental data are formulated. For example, processing of social and economic information on a condition and development of Russian cities with the population more than 100 thousand people is executed. It is shown that for spaces of system's conditions the mathematical functions of entropy and potential can be offered. Their sense is connected with introduction of natural curvilinear coordinates in continual spaces. These values consider features of discrete representation of information and form the continuous regularities reflecting variable's interrelation according to experimental data. The received results allow to develop methods of empirical data processing.

Key words: Difficult systems, spaces of conditions, temporal data bases, models of data description, phenomenological dependences

INTRODUCTION

It is known that the theoretical informatics is mathematical discipline in which mathematical methods are applied to construction and studying the models of processing, transfer and use of information. It has historically developed that in most cases information messages are displayed by a discrete set. From the point of view of model's creation, the representation of information in the form of a continuous set is extremely important because all basic laws in natural sciences are connected with continuous values. Very often the results of experimental data or observations of processes and the phenomena in the nature and society are presented in the form of the data arrays characterizing set of the same objects. Such objects unite on the basis of belonging to a certain type of difficult systems, for example, substances, physical bodies, experimental samples, installations and equipment, biological organisms and individuals, populations, settlements and cities areas, regions, ecosystems, countries, etc. All such objects represent one class and have a certain number of characteristic properties which quantitatively are defined by the parameters changing eventually (Averin, 2014; Zviagintseva, 2016).

Similar information has the discrete form and can be presented in the form of temporal data arrays characterizing processes of change and development of objects and systems. Temporal data have table structure in the form of matrixes "objects-indicators" and the set of tables is ordered on time for example, years, months, etc. Many experimental and statistical data on observations of conditions of various difficult systems belong to temporal data.

The purpose of this study is to study the possibility of discrete quantitative information's representation in a continuous form on the example of analysis of temporal data arrays and receiving phenomenological ratios for the description of such data.

MATERIALS AND METHODS

Hypotheses and methods: For the solution of the assigned problem it is necessary to develop system of the principles, methods and means of the phenomenological data analysis for arrays of the experimental or statistical information reflecting processes of change and development of various difficult systems in a type of temporary series. Phenomenological models can be

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

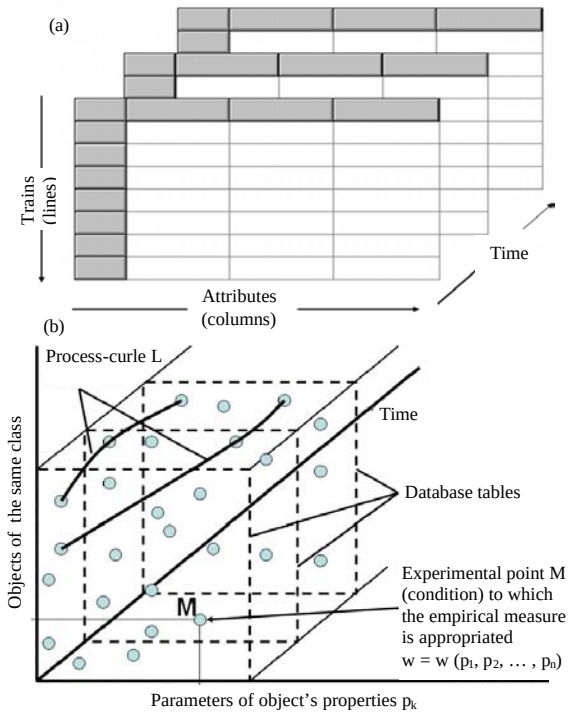

Fig. 1: Structure of temporal data arrays: a) structure of temporal databases and b) space of object's conditions

focused on the description of the structured quantitative information arrays in the most different applied areas (Zviagintseva, 2016; Averin *et al*., 2015a, b).

Such problems are connected with creation of the temporary (temporal) databases representing the whole direction in the theory of DBMS. Temporal databases apply three-dimensional tables to storage of information. At the same time, relational tables DB can be considered as a special case of temporal data tables. Theoretically, the similar logical data model has a direct relation to multidimensional spaces of difficult system's conditions. In these systems the attributes of objects in the form of quantitative indexes correspond to attributive variables (state variables) in the accepted systems of coordinates. In turn, at commission by objects the processes a time acts as the general parameter in relation to these variables. The structure of temporal data in a form "objects indicators-time", characterizing difficult systems is shown in Fig. 1.

Let's assume that a certain group of uniform objects of the studied class of difficult systems in quantity m is characterized by n indicators p_1, p_2, ..., p_n. Then in n dimensional space of coordinates H^n {p_1, p_2, ..., p_n} n values of coordinates P_k will correspond to each studied object. We will define H^n as space of observed conditions for the studied group of uniform objects. The condition of any object in n-dimensional space in each timepoint will be displayed by a point M = M (p_1, p_2, ..., p_n), process of change of an object's condition in time-by a multidimensional curve which is described by a point M in this space (Fig. 1b).

Each object's conditions can be characterized not only by indicators p_1, p_2, ..., p_n but also by some observed events which reflect object's changes. Let's assume that for any point M in space of conditions H^n some event j and the corresponding probability of this event can be set to correspondence. Let's define that the object's condition in the set timepoint will be characterized by values p_1, p_2, ..., p_n which are in total displayed by a multidimensional point M_i and also by this observed event and its probability.

Statistical probabilities for an event j can be found with use of various algorithms of sorting, group and calculation of frequencies of favorable events in the general selection of all observations which are available in temporal data array (Averin, 2014; Zviagintseva, 2016a-c). At the same time let's consider events which we will be appropriate to one data table as joint events and the events appropriate to different data tables as not joint events.

In this research the probability of a joint event of several attributive indicator's observation was considered as probability of an object's condition (some condition M). We mean the attributive indicators as the most important indicators (variable states). Let's consider this event as indicative.

We will accept several hypotheses which have phenomenological character and can be confirmed or disproved on the basis of the analysis and processing of the available data.

The first fundamental idea of a research is in search of communications and regularities between probabilities of observation of various events appropriated to the studied temporal data array (Averin, 2014; Zviagintseva, 2016a-c; Averin *et al*., 2016a, b). It allows to consider hypothetically two values for the characteristic of a condition in each elementary area of space H^n: statistical probability of observation of indicative events w (P_1, P_2, ..., P_n) and metrics of space T = T (p_1, p_2, ..., p_n) the scalarvalue characterizing object's conditions and reflecting modeling result. This metrics is set in the form of dependences according to indicators p_1, p_2, ..., p_n. The metrics T in the area H^n can be represented in the form of functional dependences concerning all n indicators:

multiplicative, expert or other dependences or in the form of various measures of object's similarity: Euclidean, Manhattan, power distances, Chebyshev's distance, etc.

The second idea of research is connected with a possibility of creation the phenomenological models for the temporal data arrays. These models accept a hypothesis of existence the scalar fields of distributions of described event's probability. For this purpose it is supposed that the probability w (p_1, p_2, ..., p_n) in space H^n forms the scalar field. It is also considered that on the basis of a metrics T = T (p_1, p_2, ..., p_n) the approximate mathematical model can be constructed. This model forms one more scalar field called the modeling environment. Further for any process L near any point M communication of a form $dw = c_1.dT$ is postulated where c_1 the values allowing to receive more exact model for the data description. Values c_1 can be determined by the available statistical data on the basis of the regression analysis. Distinctive feature of this approach is the idea that the offered hypotheses can be accepted or rejected on the basis of processing of the available data. Also, this approach allows to study in general the processes of different nature: physical and chemical, ecological, social and economic, etc.

Thus, for creation the phenomenological models of data description we will use the following principles (Averin, 2014; Zviagintseva, 2016a-c; Averin *et al*., 2016a, b).

The continual principle is a principle of representation of quantitative data in space H^n according to which the continual environment in the form of multi-dimensional space of indicators is considered continuous and unstructured. Each element of space is connected with all neighbor elements according to regularities appropriated to the studied subject area. It allows to consider the available experimental data as some selection from the solid environment of an infinite number of conditions for objects of the same class. For model's creation each object's condition in continual space is characterized either by a set of object's indicators or by a joint event of this set's observation.

The second principle is based on a hypothesis that data form any "image" in continual space. It is considered that this "image" can be described on the basis of estimates of statistical probabilities of the indicative events characterizing location of each experimental point (a condition of M), concerning all group of the studied objects (all cloud of experimental points M). Proceeding from probabilistic representations, the possibility of creation the equations of object's conditions in the multidimensional distribution's form w = w (p_1, p_2, ..., p_n) is supposed.

For the assessment of probability of a joint event (an indicative event) methods of algorithmic determination of probabilities of events are used (Averin, 2014; Zviagintseva, 2016a-c). For this purpose, the statistical probability of an event w is defined by relative frequencies of events j using splitting whole space H^n into multidimensional parallelepipeds, proceeding from the quantity of intervals of grouping for each variable p_k (usually identical to all p_k). After that, relative frequencies of events which are equal to the ratio of quantity of experimental points, getting to the set fields of grouping, to total number of all observed points are calculated. The statistical probability is accepted in the form of cumulative relative frequencies of the studied events. Thus, statistical probabilities of an event j are estimated by an algorithmic way of direct calculation of probabilities.

The third principle is connected with a possibility of the phenomenological description of the scalar field of indicative event's j probabilities in space H^n using modeling functions in the form of metrics of condition's space. The principle of phenomenological data array's description allows to create in general the empirical continual space's models taking into account communications between probability of object's conditions and values of these object's indicators.

Main part: In general, the data processing technique in each case includes the following stages: the temporal database for a certain type of difficult system is formed; the list of attributive indicators (condition variables) which fully characterize conditions of the studied objects is formed. These indexes are defined by the ideas of behavior of this system's class which have developed in scientific community. They also defined by the correlation analysis of data or other methods of establishment of the most significant variables; indicative events of attributive indicator's observation are chosen. Probabilities of these events define probability of a condition of temporal database's subjects; features and regularities of this class of systems are investigated, hypotheses are offered and ways of definition the metrics of condition's space are developed; the regression dependences characterizing communication between statistical probability of a condition w and value T are established. The phenomenological constants c_1 are founded, proceeding from the regression data analysis on a basis of application the method of probit-analysis (Bliss, 1934); practical

calculations for test examples are executed, probabilistic models are developed for some types of systems, reliability and adequacy of data description's models are estimated. Creation the models of the temporal data description in space of conditions is connected with the solution of Pfaff (Eq. 1 and 2):

$$dw = c_1 \cdot \left(\frac{\partial T}{\partial p_1}\right) dp_1 + c_2 \cdot \left(\frac{\partial T}{\partial p_2}\right) dp_2 + ... + c_n \cdot \left(\frac{\partial T}{\partial p_n}\right) dp_n \quad (1)$$

where, c_k the phenomenological values characterizing the changing processes of object's conditions at changing the corresponding indicators P_k.

It is possible to represent a metrics of space of conditions H^n in the form of geometrical probability or in the form of relative change's measure because distribution of statistical probability is studied:

$$T = \frac{p_1 \cdot p_2 \cdot ... \cdot p_n}{p_{1max} \cdot p_{max} \cdot ... \cdot p_{n\,max}}; T = \frac{p_1 \cdot p_2 \cdot ... \cdot p_n}{p_{l_0} \cdot p_{2_0} \cdot ... \cdot p_{n_0}} \quad (2)$$

where, P_{kmax}, P_{k0} is respectively the maximum values or some basic values of attributive indicators. In research by Averin (2014) it is shown that for a metrics Eq. 2 it is possible to find entropy of a condition in relation to the point M in a form:

$$ds = \frac{dw}{T} = c_1 \frac{dp_1}{p_1} + c_2 \frac{dp_2}{p_2} + ... + c_n \frac{dp_n}{p_n} \quad (3)$$

$$s - s_0 = c_1 \cdot \ln\left(\frac{p_1}{p_{l_0}}\right) + c_2 \cdot \ln\left(\frac{p_2}{p_{2_0}}\right) + ... + c_n \cdot \ln\left(\frac{p_n}{p_{n_0}}\right) \quad (4)$$

Lines of entropy determine natural curvilinear coordinates in space of conditions H^n. In turn, for space of conditions there is a potential $P(p_1, p_2, ..., p_n) = C$ which can be presented in the form Averin (2014):

$$\frac{p_1}{c_1} \cdot dp_1 + \cdot \frac{p_2}{c_2} dp_2 + ... + \cdot \frac{p_n}{c_n} dp_n = 0 \quad (5)$$

$$P(p_1, p_2, ..., p_n) = \frac{1}{2}\left(\frac{p_1^2 - p_{l_0}^2}{c_1} + \frac{p_2^2 - p_{2_0}^2}{c_2} + ... + \frac{p_n^2 - p_{n_0}^2}{c_n}\right) \quad (6)$$

where, it is accepted that $P(P_{10}, P_{20}, ..., P_{n0}) = 0$. Potential is presented in the form of family of surfaces orthogonal to lines of entropy (s = const).

RESULTS AND DISCUSSION

Example of temporal data description: The analysis of experimental data shows that for many difficult systems of different nature the fundamental patterns can be determined on the basis of phenomenological descriptions of experimental data arrays (Averin, 2014; Zviagintseva, 2016a-c; Averin *et al*., 2016a, b).

Today the urban economy and living conditions in the cities for mankind have a great meaning (Tahmassebpour and Otaghvari, 2016; Tahmassebpour, 2017; Bailey, 2017). Feature of the present stage of social and economic development of the cities is growth of openness of their economy and amplifying globalization processes, against the background of extremely uneven development of regions in the world (Joss, 2012; Esfahani *et al*., 2013; EIU, 2013). According to the famous architect Jan Gehl "we form the cities and they form us" (Tahmassebpour and Otaghvari, 2016).

That's why for example we will present the possibility of creation the descriptions of statistical data, characterizing a condition and development of the cities, using the offered method.

Studying the information on Russian cities was based on data of service of the state statistics of Russian Federation. Selection of information on statistics of 159 Russian cities with the population >100 thousand people is made from this database. For each city information on 63 indicators was used. The available data covered time period from 2003-2015, data tables "the cities-indicators" were formed with a step one year. As a result of the works the temporal array of statistical data is created which included 13 data tables (2003-2015), containing in each table information on 63 indicators for 159 cities.

For illustration the possibilities of creation the models for the analysis, we will choose several indicators reflecting the economic potential of citie's development: the volume of goods and services of industrial production p_1, amount of works executed in construction p_2 and retail trade turnover p_3. We will define a joint event of observation of these indicators as the main indicative event, characterizing the city's condition in economic aspect. The statistical probability of this event was founded by the algorithmic way (Averin, 2014; Zviagintseva, 2016a-c) and calculated cumulatively in all group of objects (159 cities).

Regression probability dependences of a joint event of two or three indicator's observation for a certain

Table 1: The condition equations of cities for 2003, 2013 and 2015

Years	Citie's indicators	Condition equation	Corr. coeff.
2003	p_1,p_2, p_3	$Pr = -4.161+0.333. \ln(p_1/p_{1min})+0.327. \ln(p_2/p_{2min})+0.250. \ln(p_3/p_{3min})$	0.97
2013	p_1, p_2, p_3	$Pr = -4.881+0.245. \ln(p_1/p_{1min})+0.222. \ln(p_2/p_{2min})+0.392. \ln(p_3/p_{3min})$	0.97
2015	p_1, p_2, p_3	$Pr = -4.960+0.246. \ln(p_1/p_{1min})+0.196. \ln(p_2/p_{2min})+0.406. \ln(p_3/p_{3min})$	0.96
2003	p_1, p_3	$Pr = -4.309+0.441. \ln(p_1/p_{1min})+0.465. \ln(p_3/p_{3min})$	0.98
2013	p_1, p_3	$Pr = -4.543+0.295. \ln(p_1/p_{1min})+0.510. \ln(p_3/p_{3min})$	0.97
2015	p_1, p_3	$Pr = -4.684+0.279. \ln(p_1/p_{1min})+0.526. \ln(p_3/p_{3min})$	0.97
2003	p_2, p_3	$Pr = -3.416+0.448. \ln(p_2/p_{2min})+0.409. \ln(p_3/p_{3min})$	0.98
2013	p_2, p_3	$Pr = -4.530+0.334. \ln(p_2/p_{2min})+0.561. \ln(p_3/p_{3min})$	0.98
2015	p_2, p_3	$Pr = -4.623+0.300. \ln(p_2/p_{2min})+0.585. \ln(p_3/p_{3min})$	0.96
2003	p_1, p_2	$Pr = -3.842+0.461. \ln(p_1/p_{1min})+0.446. \ln(p_2/p_{2min})$	0.98
2013	p_1, p_2	$Pr = -4.312+0.431. \ln(p_1/p_{1min})+0.357. \ln(p_2/p_{2min})$	0.98
2015	p_1, p_2	$Pr = -4.354+0.455. \ln(p_1/p_{1min})+0.317. \ln(p_2/p_{2min})$	0.96

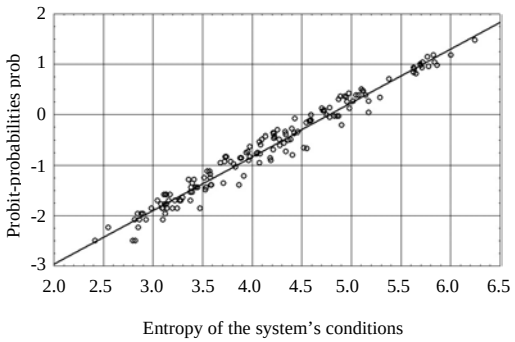

Fig. 2: Distribution of probabilities of the joint events characterizing citie's conditions in 2013 for indicators p_1, p_2, p_3

temporal database table (for the chosen observation's year) were defined on the basis of a probit-analysis method in a form:

$$w(Pr) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Pr} exP\left(-\frac{t^2}{2}\right) dt;\ Pr = c_0 + s$$

$$s = c_i \cdot \ln\frac{P_i}{P_{i_0}} + c_q \cdot \ln\frac{P_q}{P_{q0}} \qquad (7)$$

$$s = c_1 \cdot \ln\frac{P_1}{P_{1_0}} + c_2 \cdot \ln\frac{P_2}{P_{2_0}} + c_3 \cdot \ln\frac{P_3}{P_{3_0}}$$

Where:

c_1 = Empirical constants

p_1, p_2, p_3 = The chosen indicators, i, q = 1, 2, 3

The minimum indicator's values $p_{k0} = p_{kmin}$ in a group of the cities (in a column of data) which were observed in 2003 were accepted as basic values.

Some of the received equations of the citie's conditions for various indicator's combinations are given in Fig. 2 and Table 1. Correlation coefficients for regression dependences have high values (from 0.96-0.98). There by, the hypothesis of a possibility of creation the condition equations is confirmed. These equations have a form of multidimensional empirical functions of joint event's distribution. At the same time a certain dependence of distribution's coefficients on time in a form Eq. 7 (Table 1) is observed. And that is natural to this class of the objects, differing in the expressed dynamics of processes.

CONCLUSION

According to these methods it is offered to use the general approach of representation the research object's observations in the form of the structured multidimensional selection of experimental data from the continuous hypothetical environment characterizing difficult system's conditions on a set of indicators. For such idea's formalization the hypothesis of empirical measure's existence for the complex description of system's conditions in the form of probability of indicative events is also used. It allows to offer the theory for the description of some types of quantitative discrete data in a continuous form.

So, in space of conditions H^n it is possible to enter natural curvilinear coordinates in the form of entropy and potential, at the same time each object in the course of change and development will hold some position concerning these coordinates. These values differ in additive property. They are condition functions at justice of existence of the scalar field of statistical probability w. Entropy and potential can be used at creation the phenomenological models of temporal data description.

Upshots: Thus, the continual principle of representation the space of object's condition, probabilistic approach to the description of "image" of all object's conditions as clouds of geometrical points and a phenomenological research method give the chance to present discrete information on object's conditions in the form of continuous models of the data description. It allows to find the regularities appropriated to the studied difficult systems and to use this approach at developing algorithms of empirical data processing, forecasting difficult system's conditions, multiple-parameter ranging of objects, etc.

REFERENCES

Averin, G.V., 2014. [System Dynamics]. Donbass Publisher, Donetsk, Ukraine, Pages: 405 (In Esperanto).

Averin, G.V., A.V. Zviagintseva, I.S. Konstantinov and O.A. Ivashchuk, 2015a. Data intellectual analysis means use for condition indicators assessment of the territorial and state formations. Res. J. Applied Sci., 10: 411-414.

Averin, G.V., I.S. Konstantinov, A.V. Zviagintseva and O.A. Tarasova, 2015b. The development of multi-dimensional data models based on the presentation of an information space as a continuum. Int. J. Soft Comput., 10: 458-461.

Averin, G.V., A.V. Zviagintseva, M.V. Shevtsova and L.N. Kurtova, 2016a. Probabilistic methods of a complex assessment of quantitative information. Res. J. Appl. Sci., 11: 415-418.

Averin, G.V., I.S. Konstantinov and A.V. Zviagintseva, 2016b. About continual approach to model data presentation. J. Comput. Inf. Technol., 10: 47-52.

Bailey, L., 2017. The wealth report global cities survey. Knight Frank, London, UK. http: //www.knightfrank.com/blog/.

Bliss, C.I., 1934. The method of probits. Science, 79: 38-39.

EIU., 2013. Hot spots 2025: Benchmarking the future competitiveness of cities. Economist Intelligence Unit, London, UK. http://www.citigroup.com/citi/citiforcities/pdfs/hotspots 2025.pdf.

Esfahani, M., M. Emami and H. Tajnesaei, 2013. The investigation of the relation between job involvement and organizational commitment. Manage. Sci. Lett., 3: 511-518.

Joss, S., 2012. Tomorrows City Today: Eco-city Indicators, Standards and Frameworks; Bellagio Conference Report. University of Westminster, London, England, ISBN:978-0-9570527-2-7,.

Tahmassebpour, M. and A.M. Otaghvari, 2016. Increase efficiency big data in intelligent transportation system with using IoT integration cloud. J. Fundam. Appl. Sci., 8: 2443-2461.

Tahmassebpour, M., 2017. A new method for time-series big data effective storage. IEEE. Access, 1: 1-110.1109/ACCESS.2017.2708080.

Zviagintseva, A.V., 2016a. About probabilistic analysis of observational data about of observation the natural and antropogenic systems in multidimensional spaces: Belgorod State University Scientific Bulletin. Econ. Inf. Technol., 2: 93-100.

Zviagintseva, A.V., 2016b. Probabilistic Methods of a Complex Assessment of Natural and Anthropogenic Systems. Cpektr Publishing House, Moscow, Russia, Pages: 257.

Zviagintseva, A.V., 2016c. Russian cities air pollution simulation on the basis of determining the adverse events probability: Belgorod State University Scientific Bulletin. Econ. Inf. Technol., 16: 107-114.