International Journal of Soft Computing 10 (6): 458-461, 2015
ISSN: 1816-9503

The Development of Multi-Dimensional Data Models Based on the Presentation of an Information Space as a Continuum

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

Abstract: The topical issues related to the problem of multidimensional data modeling that characterize the processes and the system state of different nature. The physical and chemical, engineering, manufacturing, biological and socio-economic systems are considered the quantitative information about which may be presented as the arrays of tabular-time data. Each table of such an array has the structure "objects and settings" and many tables are ordered in time. The principles of creation and relational-field approach of multi-dimensional data model development are proposed for such information arrays. The peculiarity of the approach is the use of data presentation idea as a hypothetical solid information environment and the application of existence hypothesis concerning an information measure for a comprehensive assessment of tabular data array in the form of joint event probability field associated with the simultaneous observation of object parameters. It is also shown that using the hypothesis characterizing the features of data changes over time, you may reduce the initial modeling problems to the solution of mathematical physics equations. The obtained results enable to develop the methods of data modeling for the databases characterizing the condition and behavior of certain system classes.

Key words: Complex systems, table-time data, probable information spaces, data models, hypothesis

INTRODUCTION

One of the current ideas of a system analysis and informatics is associated with the ability to create adequate models for entire classes of different objects that could become the basic information technology, differing by a certain universality in a practical aspect. Today, traditionally with the use of models the algorithms for the set tasks are developed. However, the discussions take place on the prospects of computer technology development, the basis of which may be not the algorithmic information processing methods but the methods of a computer direct interaction with the models that will have a high level of formalization and the universality of model presentation. In Computer Science, this trend is related to the key word "constraint" (constraint programming, constraint propagation, etc.) (Narinyani, 2014). With regard to the quantitative data, such studies involve the processing of information not on the basis of direct work with the arrays of numbers but on the basis of the work with these array models that take into account the features and patterns of a particular subject area. It is known that a model is understood as a simplified representation of a real object that is displaying or reproducing a study object replaces it and provides new information about it. This definition also applies to data arrays, if they are considered as an object of modeling. The creation of data model requires a quantitative model development that describes the data with the required degree of accuracy. An own theory of data description each is used now a days for each subject area. For example, thermodynamics uses the methods of quantitative data storage and processing implemented at high levels of information understanding as the algorithms of presentation and data analysis use data models initially.

The appropriate models are based on hypotheses, theoretical statements and developed mathematical apparatus. After the works written by C. Caratheodory many researchers developed an axiomatic trend in thermodynamics (Gyarmati, 1962; Falk, 1959; Landsberg, 1970; Landsberg, 1961; Sears, 1964; Zemansky, 1970; Lionello and Berberan-Santos, 2000) and achieved a significant progress.

A natural question arises about the possibility of the specified approaches application during the description of the quantitative information if the data are regarded as a limited sample of solid information environment. If there are own theories of data description in a number of subject areas, it is not possible, yet to develop a universal model for a general case which formally would be applicable in a variety of fields of knowledge to describe

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

the data presented in a general structured way. The study proposed to use a common approach multidimensional data representation in the form of a hypothetical solid information environment that enables to offer a formal theory for certain types of data description.

MATERIALS AND METHODS

We will represent some system of a specific nature as the collection of same class objects such as agents, products, animals, people, countries, etc. All system objects have a certain number of specific properties that are determined quantitatively by measured or monitored parameters $z_1, z_2, ..., z_n$. Each object performs a natural process of development And therefore its parameters change over time $z_1(t), z_2(t), ..., z_n(t)$. Periodically (with an increment of a year, a month, an hour, a second, etc.) the monitoring of all system objects is performed. Therefore, an array of table-time data, in which each table has the following structure: "objects-parameters" and many tables ordered in time with a predetermined pitch. Then, we assume the existence of a vast data arrays for a studied system accumulated during a long observation of its behavior. Such structured data exist for the systems of diverse nature, both in sciences and in humanities. In order to formulate common simulation approaches we will use the materials of the researches (Averin, 2014; Averin and Zvyagintseva, 2013), so, the main results will be presented in a short description. In order to demonstrate the possibilities of data models in different subject areas let's make the following quite general assumptions.

Let's assume under any such system state a set of its observed parameters $z_1, z_2, ..., z_n$ that are formed under the influence of environmental conditions at a particular moment of time. Let's consider the n-dimensional space of states $\Omega\{z_1, z_2, ..., z_n\}$ when the points of this space correspond to n-dimensional set of values for all variables $(z_1, z_2, ..., z_n)$. This space will be considered as a multi-dimensional information space, which characterizes all observed states of a system. The state of system objects in n-dimensional space at any moment of time will be displayed by a multidimensional point $M = M(z_1, z_2, ..., z_n)$, the process of the system state changing is represented by a multidimensional curve described by the point M in this space. Let's introduce the concept of an information measure for the system state w. We will consider this measure as a scalar function of the state spaces, which comprehensively reflects the status of each point M in the space Ω_n. We believe that the value uniquely a system state, depends on the parameters $z_1, z_2, ..., z_n$ and cannot be one of the parameters in the system. Let us take as an information measure of a system state the probability of joint events related to the simultaneous observation of a set of object parameters $w = W(z_1, z_2, ..., z_n)$. This possibility exists for each point M in space Ω_n and, it can be algorithmically evaluated the available data monitoring (Averin, 2014) and may be evaluated algorithmically according to the available monitoring data. The introduction of an information measure allows to establish a functional relationship in a multidimensional space Ω_n between the position of the point M and the observable parameters of this point for any point in time.

Let's assume the continuity of the area Ω_n as well as the possible existence of a continuous scalar field of an information measure. This means that in the state space Ω_n has a lot of states for a general set of objects and the state point $M(z_1, z_2, ..., z_n)$ continually fill that space. We assume that the points $M_i(z_1, z_2, ..., z_n)$ of a table-time data array are a limited sample of a given general population.

Thus, the basic hypothesis is related to the ability of phenomenological model experience data in a table-time form that have a multidimensional field representation of the state space Ω_n as well as the existence of a scalar field of information measure w in the form of joint event observation probability for a set of parameters.

In order to develop a common data model, let's assume that in the area Ω_n of the existing dependence library one may set the function $\theta(M) = \theta(z_1, z_2, ..., z_n)$, on the basis of which we will create a data model. This function may be set from empirical or theoretical assumptions And allows you to create a scalar field of the value $\theta(M)$.

Just as in (Averin, 2014; Averin and Zvyagintseva, 2013), for any process 1 near the arbitrary point M, we postulate the relation of the following type $dw = c_1 \times d\theta$, where c_1 are phenomenological values which are the functions of the process and are determined by the available data. All this leads to a special type of Pfaff equations that are integrable for certain classes of functions $\theta(z_1, z_2, ..., z_n)$ and certain types of phenomenological variables c_1:

$$dw = c_1\left(\frac{\partial\theta}{\partial z_1}\right)dz_1 + ... + c_n\left(\frac{\partial\theta}{\partial z_n}\right)dz_n \qquad (1)$$

If we consider, the table-time data in the form of a continuous multi-dimensional environment, Eq. 1 will describe any elementary process in any area of the point M. In many cases, this equation in the multidimensional space Ω_n a general integral (potential) may exist like $U(z_1, z_2, ..., z_n) = C$, a general integral (potential) may exist like which may be determined by the approximation of the available data using set dependencies $\theta(z_1, z_2, ..., z_n)$.

In general case, the simulation environment in the area Ω_n may be represented in a variety of functional dependencies in respect of the following parameters $z_1, z_2, ..., z_n$: by multiplicative, grade, additive, expert or other dependencies belonging to the classes of homogeneous or multiplicative functions (Averin, 2014). During the study of probabilistic information space as a simulation environment a multi-dimensional geometric probability may be used.

The adequacy of the chosen simulation environments may be estimated on the accuracy of the data approximation data and according to the possibility of phenomenological quantities determination c_1, on the basis of the available experimental data.

RESULTS AND DISCUSSION

Main part: Thus, the problem of studied system simulation is associated with data arrays and modeling environments which may adequately describe these data. The search of various kinds of functions θ, the determination of variables c_1 and quality assessment of obtained dependencies leads to a significant amount of calculations, especially when there is a lot of experimental data. However, the processing of such data for the obtaining of such models may be carried out taking into account the assumptions that determine the patterns of an information measure development in the state space Ω_n.

The information measure as well as the system state parameters vary during time, i.e., the relation $w(t) = W(z_1(t), z_2(t), ..., z_n(t))$ is true one. Based on the hypothesis of a measure scalar representation in the vicinity of each point M, there are many values of w parameter changes, depending on the vector of the process l direction of the vector.

Taking into account the previously adopted hypothesis about the relation of w and θ values changes of the type $dw = c_1 \times d\theta$, let's assume that there is a relation between the scalar fields of these variables in the following form for any process l in the arbitrary point M:

$$dw = \text{grad}_l W(M) = c(M) \times \text{grad}_l \theta(M) \qquad (2)$$

where, $c(M) = c(z_1, z_2, ..., z_n)$ is the coefficient of proportionality as the process function that determines the relationship between the values w and θ.

If we consider, the closed surface σ of the multidimensional volume υ, selected in the area Ω_n, then during the period dt through the surface σ the vector flow grad w(M) will be equal to:

$$dw = dt \iint_{(\sigma)} c \times \text{grad}\ \theta\ d\sigma \qquad (3)$$

One may put forward different hypotheses for the value w about its change over time which will be related with the essence of this magnitude in the studied subject area. For example, let's assume that the field of probability is subject to the preservation law, then using the balance method, we will obtain:

$$dw = dt \iiint_{(\upsilon)} \beta \frac{\partial \theta}{\partial t} d\upsilon = dt \iint_{(\sigma)} c \times \text{grad}\,\theta\ d\sigma \qquad (4)$$

$\beta(z_1, z_2, ..., z_n)$ is some function of proportionality. Applying the Ostrogradsky formula to Eq. 4, we obtain the equation of a parabolic type to select a simulation environment:

$$\begin{aligned} \beta \times \frac{\partial \theta}{\partial t} &= \text{div}\left(c \times \text{grad}\ \theta\right) \\ \beta \times \frac{\partial \theta}{\partial t} &= \frac{\partial}{\partial z_1}\left(c \times \frac{\partial \theta}{\partial z_1}\right) + ... + \frac{\partial}{\partial z_n}\left(c \times \frac{\partial \theta}{\partial z_n}\right) \end{aligned} \qquad (5)$$

This differential equation is similar to a non-stationary equation of convective diffusion at its generalization to the n-dimensional case.

The differential Eq. 5 and experimental data or observations in the form of table-time data allow to verify the original hypothesis and to determine the phenomenological values c(M) and β(M). This problem is reduced to the solution of inverse boundary problems for a parabolic type equation and the restoration the required quantities by data that are collected at the monitoring of the system parameter change processes over time.

Similarly, using different hypotheses in relation to the process of information measure change in time, one may come to various boundary-value problems of mathematical physics.

CONCLUSION

In this stusy, we showed that by accepting the hypothesis on the existence of an information measure, you may develop a data model for a system of any nature which have observation or experience data presented in the form of quantitative information arrays. Using various hypotheses, characterizing the change peculiarities in of an information measure in time, one may reduce an initial problem to the solution of various equations in mathematical physics. A special feature of this approach

is the transition to the idea of experimental data presentation in the form of a solid information environment.

REFERENCES

Averin, G.V., 2014. System dynamics. Donetsk, Donbass, pp: 405, 15.09.2015, http://www.chronos.msu.ru/ru/rnews/item/sistemodinamika.

Averin, G.V. and A.V. Zvyagintseva, 2013. The relationship of the thermodynamic and information entropy during the description of the ideal gas states. Syst. Anal. Information Technol. Sci. Nature Society, 1 (4)-2 (5): 46-55.

Falk, G., 1959. Die Rolle der Axiomatik in der Physik, erlautert am Beispiel der Termodynamik. Die Naturwissenschaften, 46 (16): 480- 486.

Gyarmati, I., 1962. On the Fundamentals of Thermodynamics. Acta Chim. Hung, 30: 147.

Landsberg, P.T., 1961. On Suggested Simplification of Caratheodory's Thermodynamics. Phys. Stat. Solidi, 1: 120.

Landsberg, P.T., 1970. Main Ideas in the Axiomatics of Thermodynamics. Pure Appl. Chem., 22: 215.

Lionello, P. and N.M. Berberan-Santos, 2000. Constantin Caratheodory and the axiomatic thermodynamics. J. Mathematical Chem., 28: 313-324.

Narinyani, A.S., 2014. Model or algorithm: the new paradigm of information technology. OLAP.ru, 15.09.2015. http://www.olap.ru/home.asp?artId=2384.

Sears, F.W., 1964. Simplified Simplification of Caratheodory's Treatment Thermodynamics. Am. J. Phys., 41: 2979.

Zemansky, M.W., 1970. Kelvin and Caratheodory a Reconciliation. Am. J. Phys., 22: 371.