ANNALS OF FACULTY ENGINEERING HUNEDOARA
- INTERNATIONAL JOURNAL OF ENGINEERING
Tome IX (Year 2011). Fascicule 3. (ISSN 1584 - 2673)

[1.] Michaela HORALOVA KALINOVA, [2.] German MICHALCONOK

APPLICATION DATA MINING IN AUTOMATIC CONTROL SYSTEMS

[1-2.] DEPARTMENT OF APPLIED INFORMATICS, UNIVERSITY OF SS. CYRIL AND METHODIUS, SLOVAKIA

ABSTRACT: The goal of the report is the estimation of a feasibility of methods data mining for control systems and automation. In the report it is underlined, that it is possible to put in a principle of design of the database method of spectral analysis of measurements of behavior of a separate variable management system.
KEYWORDS: Data mining, control system, spectral analysis, automation

❖ INTRODUCTION

With end of the last century, is actively developing a new information technology data mining, which is searching for patterns in vast amounts of data. Data mining widely are used in large companies, especially those working in e-commerce. Here is a partial list of applications data mining: Advertising; Bioinformatics; Communication with clients (CRM); Marketing; Fraud detection; e-commerce; Health; Investments/securities; Production management; Entertainment & Sports; Telecommunications; Exploring the Web. [1]

As can be seen from the list, there are areas such as process automation and optimization control systems domain complex interrelated objects. There is nothing surprising in this, because on the one hand the classical theory of regulation is still offering the necessary mathematical tools for all possible occasions. On the other hand computers for real-time control has always lagged for 10-15 years one of modern information technology. So in this case, the emergence in recent years database management systems-oriented work in real time (eg, Industrial SQL server 8.0 company Microsoft) offers opportunities to use the vast experience and practical technology data mining.

❖ METHODOLOGY AND DISCUSSION

The basis of the data analysis is the concept of OLAP (On Line Analytical Processing), which uses patterns to search for information in the form of a "cube OLAP". Moreover, the data cube is a three-dimensional table (array) of data, connected together by common attributes. OLAP cube does not necessarily have to be three dimensional. It can be two-, and multidimensional - depending on the task at hand. For especially complex analysis may need about 20 dimensions - and serious OLAP-products are precisely on such amount and calculated. More than simple desktop application support somewhere 6 measurements. OLAP cubes are, in fact, meta-statements. Cutting the meta-statements (cubes, that is) based on measurements, the analyst obtains, in fact, interested in its "normal" two-dimensional reports (not necessarily statements in the usual sense of the term - we are talking about data structures with similar functions). The advantages are obvious cubes - you must request data from a relational database only once - when building a cube. Because analysts generally do not work with information that is supplemented and changed "on the fly" generated cube is relevant for a sufficiently long time. Because of this, not only excluded outages server relational database management system (there is no request from the thousands and millions of lines of response), but also dramatically increases the speed of data access to the analyst.

For automatic control systems in such a system analyst can assume measurement and processing parameters (variables), which prepare the data for decision making in terms of control theory - the parameters of the regulator. Consider the formulation and solution of the problem of constructing a deterministic linear optimal regulator. Let the constant parameter linear system described by the matrix equation (1).

$$\begin{aligned} x(t) &= Ax(t) + Bu(t); \\ z(t) &= Dx(t); \\ u(t) &= -Kx(t), \end{aligned} \qquad (1)$$

where x (t) - n-dimensional state vector; u (t)- m-dimensional control vector, and A - (n x n) - dimensional matrix system; B - (n x m) - dimensional matrix input. The matrix D size l x n (l <n) establishes a link between the full state vector x (t) and l - dimensional vector of output variables z (t), traffic control and are designed controller. K - coefficient matrix of the feedback size (m x n).

Block diagram of control system, described by a system of equations in the complex plane p, is shown in Figure 1. Where $W_z(p)$ - matrix transfer function of a closed system of governance. For the performance criterion of value is the integral of the quadratic form

$$\int_{t_0}^{t_1}[z^T(t)R_1z(t)+u^T(t)R_2u(t)]dt, \qquad (2)$$

where R1 and R2 - symmetric weight matrix has a size of l x l and m x m, respectively. Solving the task with the help of the Riccati equation we can determine the matrix coefficients of the feedback K (t) and generate the optimal control in the form of a linear law [2]:

$$U(t) = - K(t) x(t). \qquad (3)$$

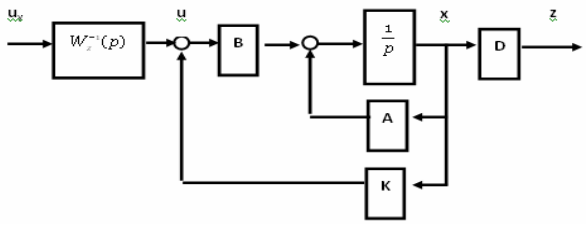

Fig.1. Block diagram of control system

Returning to the problem of optimal control for data mining must find a way to design a database structure adapted to technology, OLAP. It is clear that there needs to capture, filter and device information about the input vector control u (t) and the vector of state control object x (t). This information can then be used to calculate the control matrix K (t).

When you search for information by asking ourselves data mining to either confirm (deny) the model of the system, or optimize an objective function, based on a pre-selected model. As a result data mining we may find that a model or objective function is not fully complies with the original choice. Creative methods data mining can "correct" model or objective function, to provide forecasting and proactive means (predictive) control. Any structure (a dimension) of the database (DB) for Creative OLAP consists of data tables (cubes) and the fact table as a result of processing the data. Combining them into a dimension has to be possible on a "star" (star) or "snowflake" (snowflake). Consider the principles that could be put into the organization of data cubes.

For real-time control is clearly necessary to record data on the control actions and the status of the individual coordinate systems management. The frequency of measurements and records to the database should meet the dynamic performance management system [4]. If all data on one variable at a wide band of operating frequencies to record up to one database table with the frequency of measurement, it is obvious that we need to determine the total sampling time and sampling at a particular time of measurement and recording of data for one processing step on the technology OLAP. For this it is convenient to use the technique of spectral analysis and planning, which can be found in the literature on this issue [3]. Taking into account that the operating frequency band of control systems are usually more than three decades, for a reliable spectral analysis is necessary to divide it into several parts. Assume that the digital data x(t), t = 1, ..., N, correspond to the signal x (t), read by the time intervals Δ. In this case, the smoothed sample estimate of the spectrum is the sum of

$$\bar{c}_{xx}(f) = \Delta \sum_{k=-(L-1)}^{L-1} \omega(k) c_{xx}(k) e^{-j2\pi fk\Delta}, \quad -\frac{1}{2\Delta} \le f < \frac{1}{2\Delta} \qquad (4)$$

In (2) L = M / Δ, where M - the cut-off point. Since - is an even function, it must be calculated in the interval with a doubling of capacity. Commonly used formula takes the form (5) and is calculated only if:

$$\bar{c}_{xx}(f) = 2\Delta[c_{xx}(0) + 2\sum_{k=1)}^{L-1} \omega(k) c_{xx}(k) \cos 2\pi fk\Delta], \quad 0 \le f < \frac{1}{2\Delta} \qquad (5)$$

Dividing the entire strip of working frequencies of several bands M, 2M, ... NM, we can form a data cube for a database so that the third dimension will be the frequency band number N. For N> 1, data can be pre-filtered, using the principle of calculating the average.

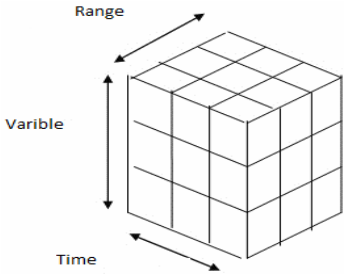

Figure 2. Organization of a cube is measuring one variable

Numbers of measurements are calculating the average of one variable determined by the number and range of the same L (N-1). An example of a cube measuring the organization for one variable is shown in Figure 2.

❖ CONCLUSIONS

1. Applying the principles of data mining is theoretical and practical interest.
2. For optimal control of OLAP technology can be transferred to the management of large systems.
3. The above schemes of measurement divided into frequency bands with the principles of OLAP and require the development of algorithms for combining these with a range of checks accepted model of management and correction of the objective function.

❖ ACKNOWLEDGMENT

This work was carried out with financial support of the University of Saints Cyril & Methodius in Trnava, project UCM 10/69

❖ REFERENCES

[1.] http://www.computerra.ru/. L. Levkovich-Maslyuk. Great digs and great challenges. Posted March 23, 2007.
[2.] Basharin A.V., Postnikov Y.V. Samples of automated electric computer. L.: Energoatomizdat, 1990. 512.
[3.] Jenkins, D. Watts. Spectral analysis and its applications. Moscow, Mir. 1972. 142.
[4.] Stremy M., Elias A. Virtual laboratory communication. In: Annals of DAAAM and Proceedings of DAAAM Symposium. - ISSN 1726-9679. - Vol. 20, No. 1 Annals of DAAAM for 2009 & Proceedings of the 20th international DAAAM symposium "Intelligent manufacturing & automation: Focus on theory, practice and education" 25 - 28th November 2009, Vienna, Austria. - Vienna: DAAAM International Vienna, 2009. - ISBN 978-3-901509-70-4, s. 0139-0140