PERGAMON

Expert Systems with Applications 26 (2004) 81–85

Expert Systems with Applications

www.elsevier.com/locate/eswa

Analytic information technologies in oil refinery

A.A. Musaev*

Abstract

Problems of constructing analytical information systems and their application in controlling industrial processes at oil refining enterprise are considered. Systems being focused on control decisions support problem are regarded. Basic tasks, arising at constructing such systems and the ways to overcome them are listed. The appearance of multipurpose analytical information system (MAIS) is developed to serve as universal platform to create intellectual control tools.

The article carries arrangement character and is devoted to problems of elaboration MAISs and their implementation in the interests of improving industrial and technological control processes.

Keywords: Analytical information system; Data mining; On-line analytical processing; Situational production control; Data Warehouse; Data Mart

1. Introduction. Modern problems at oil refinery industrial and technological processes automated control

One of the most glaring characteristics of present state of oil refining enterprises development is active design and introduction optimization control methods, which have been named APC-technologies (*advanced process control*). Development of the optimization approach both at the level of *technological processes* (TP) control, and at the level of enterprise control in common, results in qualitative changes of the requirements for the production control methodology. Now we are to consider this question in detail.

In the majority of practical situations, common manufacture control as a whole supposes decomposition to a number of more concrete partial controls. The most typical examples of such concrete controls are:

- material streams (scheduling) control;
- raw and power resources control;
- operational control by manufacture technical means and its infrastructure;
- economic production control—from marketing to production scheduling;
- strategic production control including investment, technical and economic policy of the enterprise, etc.

All listed processes are obviously interconnected. Thereby they suppose qualitative optimization only in context of unified manufacture process. However, up to last time there were no methodological means allowing to solve the problem of integrated optimization of production processes control as a whole. Absence of formal tools of integrated production processes control was caused, first of all, by highest complexity of the problem to decide. As a matter of fact, refinery production, from the point of view of system approach, represents rather multiparameter dynamic system, described by phase vectors of large dimension. Thus the majority of production processes parameters are interdependent and interconnected, which essentially raises the number of control dimensions.

Another source of complexity of the problem to be decided is the fact that production processes represent, in essential, the open system. They have to function under the influence of large number of diverse external factors. Thus, optimization of production processes integrated control assumes the presence and account of the information on the state of the system environment, with which the system actively interacts. This, in turn, essentially raises dimension and connectivity of the control task.

As the third source of complexity of the problem to be decided, it is necessary to point the dynamic nature of production as a whole and its control systems, in particular. The state, both of production processes and environment, in their interaction, continuously evolves in time. Necessity to take this evolution into account for the interests of integrated optimization arises new problems of adaptive

* Corresponding author.*E-mail addresses:* amusaev@szma.com (A.A. Musaev).

0957-4174/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0957-4174(03)00109-X

synthesis of nonlinear dynamics models and their parametrical identification.

Thus, the complex analysis of production control represents extremely complicated problem for which decision is possible only recently due to the appearance of new generation of information systems—analytical. So, present article is devoted to the problem of such systems construction and their application in refinery production control.

2. Production processes control and analytical information technologies

The important stage of production control perfection is the switching from the concept of situational control to prognosis control technologies.

Situational approach is based on traditional information systems, providing accumulation, storage and displaying the information on current industrial situation. Thus previously the situation was understood as a set of quantitative and qualitative descriptions reflecting the current state of manufacture processes and environment in cooperation. The structure *of an information system of situational production control* (IS SPC) is shown in Fig. 1.

Formation of situational control decisions is made on the basis of the industrial situation monitoring. The results of monitoring are collected in *databases* (DB). On occasions, the results of monitoring may be used directly by executive information system (EIS) or by subject expert (SE). In the role of SE acts the specialist in the appropriate area of industrial knowledge. SE prepares a project of control decision. As a special mean for operative automated decision support *on-line transaction proceeding* (OLTP) may used.

The prepared project of the decision, after necessary correction, comes to EIS. The head (or the *person accepting the decision* (PAD)) may correct and specify it during interactive dialogue with the SE and appropriate services and afterwards directs it to executive agencies of production processes.

The pointed approach of situational control underlies construction of the majority of modern information systems, used for manufacture and TP control. However, practice of its realization in real control systems exposed a whole number of problems, most important of which are:

- absence of program tools for operative search and visualization of multivariate interconnected information to display the current situation;
- insufficient volume of information, stored in operative DB, to reveal as basic tendencies of industrial situation development, so factors noticeably influencing its efficiency;
- low efficiency of traditional methods, algorithms and programs of the situation analysis applications.

Overcoming of listed problems resulted in appearance of *analytical information technologies* (AIT). In essence, given technologies are focused on elaboration of the new generation of intellectual automated *decision support system* (DSS). Now analytical technologies include some large directions in the field of information systems and applied mathematics. The most important of them are:

- Data Mining (DM or discovery knowledge in DB);
- On-line Analytical Processing (OLAP);.
- Data Warehouse (DW) and Data Mart.

Brief review of specified directions, forming, aggregately, a class of AIT, is carried out within the framework of the given article.

3. Data Mining—search of knowledge in databases

Preparation of control industrial decisions is inevitably connected with the development of industrial situation forecasting and with prior estimation of the efficiency of accepted decisions. The reason of that fact is obvious—realization of the administrative decision is always carried out in future. The briefest form of the given idea was stated still by Napoleon: “To control means to foresee!”

The traditional situational analysis, used in control systems on production processes, is focused, mainly, on the estimation of current situation, reflecting the current state of manufacture processes and environment. The further work on industrial situation forecasting and decisions preparation, as a rule, was assigned to SE. Particularly, in production tasks SE acts in a role of the expert in such fields as management, technology, power, economics, etc. Lacks of the situation expert analysis are known enough. The opinion of the expert always reflects only subjective vision of an industrial situation. The technologist, the expert in automated control or the economist can interpret the same situation differently.

Besides it is necessary to point out the limited opportunities of the expert in realization of the quantitative situation analysis. Even in conditions of full definition, the human brain is not capable to predict the behavior of nonlinear object which takes place under the influence of 2…3 factors even with known dynamic characteristics. Uncontrollability of human memory results in limited possibilities of using retrospective experience in forecasting and decision preparation. Consequences of it can lead to large, catastrophic failures in production processes control.

Classical mathematical technologies of forecasting up to last years also did not provide reliability of results. The reason was in using of extremely simplified models, which did not reflect real dynamic properties of production processes and environments. In turn simplification of models was caused by limited technical opportunities of former digital techniques, not allowing operative treatment of high dimension dynamic process. Development of computer facilities allows realizing qualitatively new level of approach to the problem of complex industrial situations forecasting. In particular, on the basis of newest information means there was created a new technology of applied mathematics—Data Mining, focused on problems of forecasting and decisions support (Gershberg, 2001b; Karpov, Musaev, & Sherstuk, 2000; Kiselev & Solomatin, 1997; Krechtov & Inanov, 1997; Musaev, 1998).

As examples of DM mathematical toolkit it is possible to point out such divisions of applied mathematics, as multivariate statistical analysis, nonlinear dynamics, neural networks, evolutionary programming, etc. (Fig. 2).

Main difference between DM and known means of technical data analysis consists in transition from OLTP technology to analytical methods of joint researches of current and retrospective data. It is necessary to note, that the majority of authors consider optimization methods only as auxiliary DM means. However, at preparation of optimal control decisions such means work as basic mathematical toolkit of DM.

DM main tasks are forecasting of situations development and preparations of decision projects on base of complex system analysis of the current situation and files of the retrospective data.

The data analysis includes such tasks as detection and identification of the latent dependence, tendencies and interrelations, revealing of significant influence factors, prognosis controllable process features, etc. Simultaneously traditional data processing includes visualization of results, preparation of reports and projects of allowable control decisions with estimations of their reliability and efficiency.

4. OLAP—on-line analytical processing

The presence of enough data volume, reflecting the current situation (the state of production processes in interaction with environment) does not mean the opportunity of rational system control. It is not enough to have the data—it is necessary to interpret them correctly. In turn, realization of expert interpretation of huge raw data files assumes necessity of ordering, integration, aggregation and visualization in totality with opportunity of operative complex inquiries processing. Conception of OLAP, offered by E. Codd (Codd, Codd, & Salley, 1993) in 1993, was devoted to the decision of specified problems.

A basis of this conception was founded on the idea of data representation as a multivariate orthogonal cube on which axes measurements of the situation parameters are pointed. For dynamic subject domains the basic axis of hypercube is time. For production processes every possible characteristics of material streams, technological parameters, economic characteristics, etc. can be chosen as axes.

Each of the axes gives one or several data consolidation directions, allowing thus to choose the most convenient dimension for visualization and interpretations of the data. So, for example, for a time axis directions of consolidation generalization levels are 'year–season–month–day' or 'year–week', etc. For a structural axis of production processes generalization levels can be chosen as 'enterprise–department–group–employer'. It is obvious, that similar approach allows getting detailed information level for each parameter of the situation under investigation. Codd has formulated 12 rules of quality estimation of OLAP-products (Gershberg, 2001a):

- multivariate conceptual data presentation allowing analysts to operate easy with directions of consolidation;
- transparency, i.e. discharging the user from direct work with technological means of data storage and processing;
- stability of productivity (independence of productivity from parameters number and from DB volume);
- orientation to 'client–sever' architecture;
- equality of parameters used at the description of situations;
- presence of dynamic processing with rare matrices;
- support of multiuser technology;
- unlimited support of relations between given data cells (cross-measure operations);
- maintenance of user-comfortable interface (an intuitive way of manipulation with the data);
- opportunity of flexible generation of reports;
- absence of restrictions on number of the parameters describing production situation, and on number of aggregation levels.

Practical realization of OLAP-technology was carried out within the framework of three basic directions differing by a way of data presentation (Fig. 3).

The most natural matches with OLAP ideology are provided by the technology of MOLAP (multivariate databases). Thus all data are structured as hypercube (or polycube). Multivariate DB allows carrying out search and extracting the necessary information most quickly. Besides, they rather simply adapt for inclusion in information models of new built-in functions.

However, specified DB has restrictions on data volume (up to units of gigabytes), are closed (allow to work only with similar information systems) and are rather expensive. Besides unified hypercube representation inevitably results in rare structure of data storage, that essentially decreases the efficiency of utilization of the external memory.

For relational data storage model, technology ROLAP (relation OLAP) assumes presence of additional data transformation to the multivariate structure using an intermediate metadata layer. Application of ROLAP allows rather simple coordination of analytical researches with already existing technologies of corporate information systems (relational DB and OLTP data processing). Program ROLAP tools enable work with large information storehouses. Technology of ROLAP, as against MOLAP, does not demand physical reorganization of DB at changing of investigated situation dimension. At last, technology of ROLAP is more constructive in protection of the information.

The payment for relational 'comfort' is the reduction of productivity at processing the inquiries.

The third variant of OLAP-systems construction is connected with using hybrid technologies (HybridOLAP or HOLAP). The purpose of such systems is in association of advantages of both above-mentioned approaches. Unfortunately this, as well as in other cases of conciliatory proposals, unites also lacks of technologies-prototypes. Some most known program OLAP-products and their developers are given in Table 1.

5. Data Warehouses and Data Marts

Spectrum of problems connected with construction of analytical data storage systems, was in detail considered by Gershberg (2001a), Musaev and Sherstuk (2000), Saimon (1997), and Saharov, 1996 and makes rather impressive list.

Its necessary to specify questions of coordination of various data formats from own operational DB and diverse external sources, the verification of the data, the problem of binding to unified time scale, the problem of superlarge volumes of information (hundred gigabytes and more) control, problem of multidimensionality of inquiries structure, arbitrariness in experts inquiries, etc.

The basic conclusion from the given (and not full) list is unequivocal enough: traditional DB, focused on the use in corporate information systems with technologies of OLTP processing, cannot serve as effective means of storage and access to analytical information. The decision of razed problems demands new technology of the storage of analytical data and operative search of the required diverse information.

As possible variants of decisions were offered:

- the concept of *Data Warehouse* by Inmon
- the concept of Information Warehouse, developed by IBM in 80th years.

In their basis were relied principles of presence of subject orientation, presence of data integration means, data invariance and chronology. The offered concepts carried in themselves both positive nucleus and a number of disputable moments. However, as well as any other concept, they had especially ideological character. Technical realization of the proclaimed principles inevitably conducted to concrete changes in technology of data storage and processing the inquiries.

The natural approach to DW construction for the corporations, already having own information system, is application of relational technology. One of the development directions of relational DB in interests of DW construction is the design of the tables with denormalized structure (an update of typical logic data structure, known as 'star'). Other ways to increase the efficiency of relational DB are also known: application of Bitmap-indexes, subdivision of large factologycal tables into separate fragments (horizontal fragmentation), other variants of horizontal or vertical division of tables, application optimization procedures of treating the inquiries in DB with data storage circuits such as 'star', etc.

Alternative approaches to decision of a DW construction problem, as was already marked, are application of multivariate and hybrid formats. However, even in the modified form, the variant of a unified, monolithic and constant data source is not quite constructive. In particular, the analysis of changes in production processes in various conditions assumes generation of various virtual scripts and adequate data. Blocking up the unchangeable and all remembering DW by such data would be, at least, irrational. Besides it is difficult to expect any satisfactory efficiency of situational analysis when every inquiry of the expert is immersed in gigabyte depths of DW.

The presence of given contradictions has resulted in the idea of two-level data storage structure formation. At the first level, constant super-large DW is used with limited efficiency of inquiries processing. The information in such DW is catalogued thematically, covering all spectra of probable situations of production processes control regardless to any specific target. At the second level less capacious, but more operative data storage is assumed, focused on the decision of specific targets. Given approach was partly used at elaboration of the concept of Data Mart, offered by Forrester Research. Concept of Data Mart assumes, that the expert works only with such data, that are necessary for the decision of a specific target. Thus target Data Mart is approached best possible to terminal user and does not demand the superlarge memory, that is a characteristic of DW. Such two-level circuit of data storage gradually becomes the de facto standard at construction of analytical systems.

6. The multipurpose analytical information system

Despite of intensive development of the above-stated directions, a problem of their association in interests of analytical information systems development is still in a research stage. Realization of known DM-servers is frequently guided to elementary analysis circuits such as trees of decisions. Integration of OLAP-technology with

DM products is also an open question. A special problem is the construction of industrial OLAP tools.

At the same time, practical activities at utilization of analytical technologies also met certain difficulties. The market of program products offers a wide choice of mathematical packages of the DM-analysis. However, their effective utilization assumes professional mathematical preparation of the expert. In practice, in overwhelming majority of cases the expert-mathematician, being well familiar with restrictions and opportunities of DM applications, is insufficiently prepared in concrete subject domain.

Thus a complex problem of integration AIT arise with the purpose of creation automated DSS, free from the specified lacks. Search for the decision of the given problem resulted in idea of multipurpose analytical information system (MAIS) construction, which is capable to carry out functions of a base platform to form object-focused (in particular, industrial) automated DSS (Karpov et al., 2000; Musaev & Sherstuk, 2000). The structure of such system is shown in Fig. 4.

As the central link which carries out the analysis of an industrial situation, the forecasting and control decisions projects preparation, it is offered to use interactive subsystem covering *automated workplaces* (AWP) of SE and *expert-analytic* (EA).

Interface of SE is formed in conceptual environment, typical for underlying subject domain, and, thus, is extreme adapted for SE. In problems of production processes control in a role of such tool can act AWP, the chief of technological system, the technologist of working shop, the dispatcher, the mechanics of the shop, the chief of the shop etc.

Basic SE tools are interactive dialogue means for EIS and AWP EA, means to get inquiries from DW on initial data (the current situation and retrospective data), the advanced means of information visualization and aggregation (OLAP-server, OLAP-tools), and also means of preparation reports. Besides SE should have an opportunity to address to help and contemporary records sources and to external information sources (corporate, branch, national and world) within its competence framework.

It is important to note, that at such organization of work, SE is released from functions of the quantitative data analysis and the formalized situations forecasting. Work with mathematical researches of the data is carried out by EA. Main tool of EA is the DM-server. It includes the analytical subsystem (AS), allowing operating with DM-tools. Besides it is supposed to use the specialized DM-programs of data analysis and forecasting of non-stationary industrial situation development (including processes with breakdown changes of a state). Presence of AS, integrated in AWP AE, allows EA to manipulate with various mathematical methods and algorithms and also to verify the results by statistical analysis tools.

Besides AWP EA is equipped with means of interactive dialogue with SE, means of work with specialized DB (Mart Data, DB of precedents, DB of abnormal situations, etc.), means of visualization and preparation the reports.

The organization of data storage is based on the two-story circuit (Data Warehouse–Mart Data), which allows to minimize analytical inquiries time expedience. As a result of distributed character of storage and data processing in the MAIS and specificity of the organization of storehouses of the data of large volume, the computing environment and telecommunication platform of MAIS is realized on the basis of local computer network, equipped by tools of information exchange with external sources and consumers of the information.

7. Conclusion

Development of analytical information systems of industrial assignment represents the complex, expensive problem, including a joint of information, mathematical and telecommunication technologies. Cost of expansion of information storehouse is large enough; nevertheless, the majority of leading world firms creates, introduces and applies systems of a similar class, or plans their introduction in immediate prospects. It is explained by the huge prize received due to improvement of the quality of production control and accuracy in forecasting the development of industrial situations. According to consulting company Meta Group, the profit from introduction analytical system with DM-technology can achieve 1000%. As a rule such systems pay off for some months of operation.

References

Codd, E. F., Codd, S. B., & Salley, C. T. (1993). *Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate*. E.F. Codd and Associates.

Gershberg, A. F. (2001a). *Modeling refinery technological processes automated control*. St Petersburg: New Petersburg School.

Gershberg, A. F. (2001b). *Integration and intellectualization of control system of the oil refining enterprise*. St Petersburg: New Petersburg School.

Karpov, E. A., Musaev, A. A., & Sherstuk, Ju. M. (2000). *Multi-purpose analytical information system. Methodology of creation and the basic design decisions*. St Petersburg: SPb University of Communication.

Kiselev, M., & Solomatin, E. (1997). Means of extraction of knowledge in business and the finance. *Open Systems*, *4*, 41–44.

Krechtov, N., & Inanov, P. (1997). Products for the intellectual data analysis. *Computer Week*, *14*, 32–39.

Musaev, A. A. (1998). The intellectual analysis of the data: Klondike or Babylon? *Bank Technologies*, *11–12*(43), 73–84.

Musaev, A. A., & Sherstuk, Ju. M. (2000). *Architectural and technological aspects of creation of analytical information systems* (*Vol. 4*). *Proceedings of the works international conference MMTT-2000 (31–33), St Petersburg*.

Saharov, A. A. (1996). Concepts of construction and realization of the information systems focused on the data analysis. *Control System of Data Bases*, *4*, 55–64.

Saimon, A. (1997). Data Warehouses. *Control Systems of Data Bases*, *3*, 87–96.