The Data Mining usage in Production System Management

Pavel Vazan, Pavol Tanuska, Michal Kebisek

***Abstract*—**The paper gives the pilot results of the project that is oriented on the use of data mining techniques and knowledge discoveries from production systems through them. They have been used in the management of these systems. The simulation models of manufacturing systems have been developed to obtain the necessary data about production. The authors have developed the way of storing data obtained from the simulation models in the data warehouse. Data mining model has been created by using specific methods and selected techniques for defined problems of production system management. The new knowledge has been applied to production management system. Gained knowledge has been tested on simulation models of the production system. An important benefit of the project has been proposal of the new methodology. This methodology is focused on data mining from the databases that store operational data about the production process.

***Keywords*—**data mining, data warehousing, management of production system, simulation

I. INTRODUCTION

CURRENT databases are powerful sources of data, that it is possible to obtain the necessary information from. However not by conventional means such as languages based on SQL, but by processing type of ad hoc reports or OLAP (On-Line Analytical Processing) [4]. Analytics need to gather information, that they are able to model objects from, foresee trends and allow the responsible managers to make appropriate decisions dynamically.

Analysis and production management, business analysis, risk management and detection of failures and failed parts, are other examples of management activities spheres that require a perfect knowledge of the people who perform such activities.

We are in the context of knowledge management or knowledge management systems. Responsible people are becoming knowledge workers [3].

The process of knowledge discovery in databases, often also called data mining, is the first important step in knowledge management technology. End users of these tools and systems are at all levels of management operative workers and managers. And these are their demands on the processing and analysis of data and information that affect the development of these tools. The most important feature from all of the advantages of these instruments is multidimensionality, possibility to simultaneously monitor and analyze problems from several points of view, based on the needs of management tasks [8].

Work with business data can be divided into several areas, resulting in the possibility of using a particular technology. In the first phase it is necessary to handle business data management, efficiency and speed of their processing. It is also necessary to transform these data into usable form for analysis and presentation, accumulate amount of information, clean them and transform. Here it is important to decide from where these tools draw data and whether it is necessary to create some storage, e.g., data warehouse [10]. The task of the last layer is to access the information content of data by creating derived information. This layer, also known as data mining, includes data analysis and presentation of information in the most appropriate form for the end user [11]. The procedure for obtaining knowledge from the production databases is a nontrivial process of identifying valid, new, yet unfamiliar, potentially useful and well understandable knowledge in data [2].

The term data means a set of arranged facts (e.g., items in the database) and term knowledge means expression in some language, describing a subset of data or a model applicable on this subset. Therefore, by the term knowledge discovery in databases we also understand the design and fine-tuning of the studied reality model [7]. The way that best describes the data, searches for the data structure description, and last but not least to produce a description of the data at a higher level (metaknowledge). In the area of production systems management, knowledge discovery from production databases and state-driven process data is used very little currently [9].

II. POTENTIAL PROBLEMS IDENTIFICATION OF THE PRODUCTION SYSTEM

Production management must ensure the achievement of different production goals in a given timeframe. These objectives are often conflicting and their achievement depends on many factors. Many dependencies are so far very little explored, for example relationship of capacity utilization and value of lead time, depending on the size of the production batch. The problem of minimizing variable costs depending on the necessary operating supplies possession, alternatively with

This contribution was written with a financial support VEGA agency in the frame of the project 1/0214/11 „The data mining usage in manufacturing systems control“.

P. Vazan is with Slovak University of Technology in Bratislava, Faculty of Material Science and Technology in Trnava (phone: +421918892762; fax: +421335511028; e-mail: pavel.vazan@stuba.sk).

P. Tanuska. is with Slovak University of Technology in Bratislava, Faculty of Material Science and Technology in Trnava (e-mail: pavol.tanuska@stuba.sk).

M. Kebisek is with Slovak University of Technology in Bratislava, Faculty of Material Science and Technology in Trnava (e-mail: michal.kebisek@stuba.sk).

the possibility of increasing the value added percentage parameter (metric according to Lean Production). There are also few issues that are very little investigated, like the impact of priority rules for allocating the operations on the production goals. These problems and dependencies would be possible to solve by using data mining methods [1].

The process of knowledge discoveries from management of the production systems can be applied for solution of many problems. Here belong:

i. identification of production parameters influence on a production process,
ii. identification of breakdowns in production process,
iii. identification of times series of final production,
iv. deviations (divergences) detection from plan during the production,
v. failure states detection of production equipments,
vi. production process optimization,
vii. workstations layout optimization,
viii. optimization of storage subsystem,
ix. preventive (feed forward) control prediction of the production equipments,
x. failures prediction in production process,
xi. failures prediction in assembly process etc.

A. Real production system

The authors have assumed the real data obtained from operating databases of the production systems. No producer has been willing to provide the real data about its production process. The fragments of operative databases that we have obtained were non consisted and covered only small area of required data. Therefore the authors have decided to obtain relevant data from simulation models of real production systems.

On the base mentioned facts we propose all process to realize according to conceptual model on the Fig. 1.

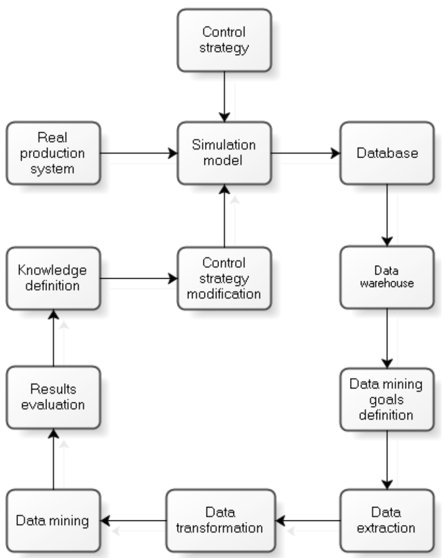

Fig. 1 The conceptual model of the project

B. Simulation model of production system

The authors have prepared several simulation models of productions systems according to real systems. The discrete-event simulation has been used for the building of simulation models as fundamental method. This approach allows to realize very detailed and accurate simulation model. The different control strategies have been implemented into the simulation model. These strategies have been oriented to reach the different production goals.

To these production goals belong:

i. to maximize capacity utilization,
ii. to minimize flow time of manufacturing,
iii. to minimize of production costs,
iv. to maximize the total number of finished parts etc.

The relevant data about production process have been collected during the simulation run. The connection between simulation model and relational database has been designed by authors as is shown on the Fig. 2.

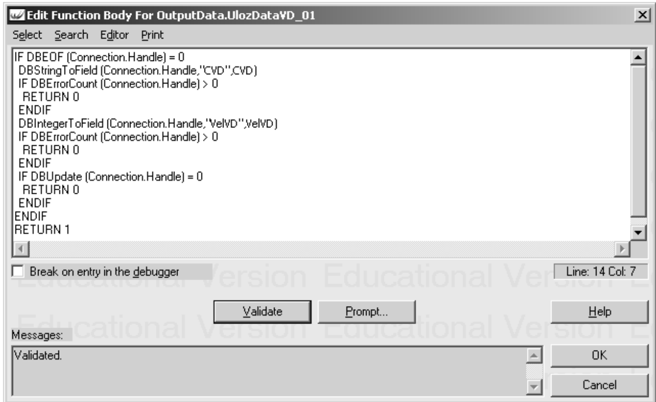

Fig. 2 Connection between simulation model and database

C. The creating of data warehouse

The designed simulation models of the production system generate all important data during the simulation run about the production process. These data need to be appropriately collected and stored for further analysis. The relational database, as a particular step has been designed to data storage from production process. The database server Oracle 11g has been used as a Database Management System (DBMS). The DBMS Oracle 11g provides sufficient tools for the management of the designed relational database. It is possible to use this database server as data source for data warehouse and simultaneously the database server can be used as an input data source for the further analyses in data mining tools [5].

Data stored in created relational database can be used directly in the data mining tool but also this data can be modified before their processing. The authors have decided to create the data warehouse for data preprocessing. The data warehouse has been designed and built by using Oracle Data warehouse Builder 11g. Preprocessed data stored in the data warehouse have been used in the further stages of data mining process [6].

III. The Proposal of the Data Mining Model

The tool STATISTICA Data Miner version 9.1 of StatSoft company has been used for data analysis from the production system. This tool allows the application of following stages of knowledge discovery: data collecting from production databases, data transformation and Modification (adaptation), data mining and knowledge discovery evaluation. All steps of this process can be performed (realized) by the graphical user interface of application, eventually some parts of the process may be precisely specified by entering the relevant source code e.g. SQL script for the specific select input data or by modification of individual parameters in source code of the data mining methods and techniques.

The following stages have been designed for the data mining process:

i. data mining goals definition,
ii. data extraction from data warehouse,
iii. data transformation and adaptation,
iv. data mining,
v. knowledge discovery evaluation.

We have defined goals that we try to obtain by using knowledge discovery from databases process in the first stage:

i. the influence analysis of production process parameters for production equipment utilization,
ii. the influence analysis of production process parameters for flow time of production batches,
iii. the influence analysis of production process parameters for number of finished parts.

Consequently we have picked the relevant data. We have realized the suitable data selection set from data warehouse with reference to defined goals and used methods and techniques of data mining.

We have adjusted and transformed the input data set. The next logical step has been investigation whether data set does not contain data which values are markedly different from others data (so-called outliers). It would be necessary to discover whether data set contains this kind of data and also the reason of their existence. It could be a random data and their extreme values could be caused by an error at the data recording. But it could be a significant data, too. Therefore it has been important to identify origin of these values correctly and decide whether these values will be included in used data set or will be removed. We have used the possibility of creation “Frequency” table for identification of markedly different data. Moreover the data set has been examined whether it does not contain missing data or kind of interference in some parameters. We have used transformed and adapted input data set in the next process steps.

The following methods and techniques of data mining have been involved into the designed model in Fig. 3:

i. Automated Neural Network Regression – Custom Neural Network,
ii. Automated Neural Network Regression – Automated Network Search,
iii. Automated Neural Network Classification – Custom Neural Network,
iv. Generalized K-Means Cluster Analysis,
v. Generalized EM Cluster Analysis,
vi. Generalized Additive Models,
vii. Standard Regression Trees (Classification and Regression Tree),
viii. Standard Regression CHAID,
ix. MARSplines,
x. SVM (Support Vector Machines).

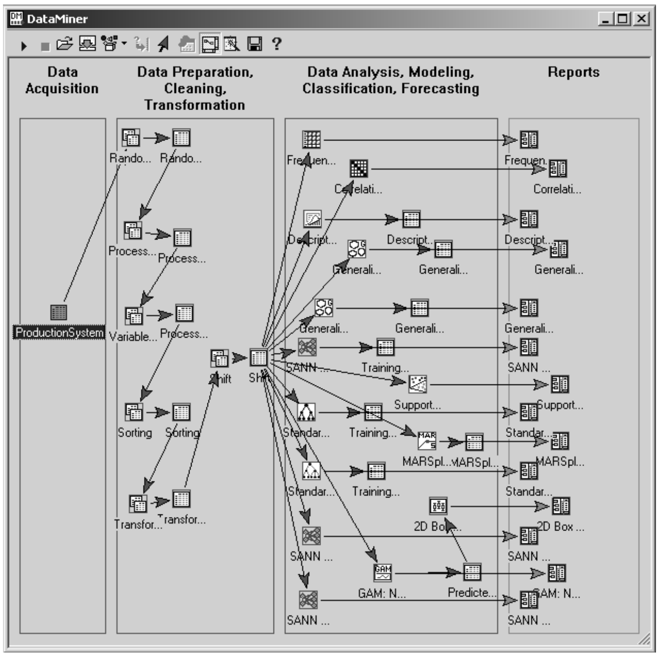

Fig. 3 Data mining model

We have obtained results from created data mining model in the form of output reports for particular data mining methods and techniques. The results are presented on the Fig. 4.

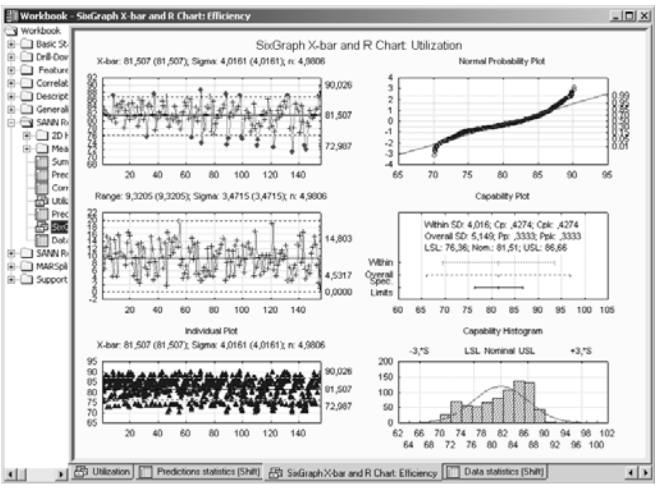

Fig. 4 Selection of data mining methods and techniques reports

Particular outcomes have been subsequently analyzed for the reason of obtaining new knowledge about surveyed production process.

IV. The Application of the Gained Knowledge

The defined goals have been evaluated individually according to gained results. The discovered knowledge for influence of parameters analyses of production process e.g. flow time, number of finished parts and capacity utilization are presented in table 1. The new and more suitable values of lot size have been gained by analyzing of reached results. The lot size value is very important input parameter for management strategies of the production system.

The new discovered knowledge has been applied to the designed simulation model. The knowledge discovery process from the production system could by verified by this way. We have compared the results of production system according to original management strategy with the results of production system that have been gained according to modified management strategy.

The change of required parameter (see Table I) i.e. capacity utilization has been reached. Capacity utilization has increased from the original value 73,31% to the new value 90,37%. This change has been realized by modification of lot size parameter. The discovered knowledge from the data mining model has been validated in this way. In fact the increase of capacity utilization has been ensured by changing of the lot size. However the next production goals have been changed at the same time. The number of finished parts has decreased and production flow time has increased simultaneously. The changes of these parameters can be evaluated from the point of view of the production system management negatively because both parameters got worse. This fact means that the production goals are conflicting – contradictory.

TABLE I
THE EVALUATION OF PARTIAL RESULTS FOR CAPACITY UTILIZATION

Parameter	Initial value	New value	Difference
Average flow time	1329 s	1674 s	–25,96%
Number of finished parts	8746 pcs	7358 pcs	–15,87%
Capacity utilization	73,31%	90,37%	23,27%

The similar results are presented in Table II. Outcomes document the applications of discovered knowledge for the next defined goals.

TABLE II
THE EVALUATION OF PARTIAL RESULTS FOR NEXT GOALS

Parameter	Initial value	New value	Difference
Average flow time	1329 s	1078 s	18,89%
Number of finished parts	8746 pcs	8311 pcs	–4,97%
Capacity utilization	73,31%	68,83%	–6,11%
Average flow time	1329 s	1537 s	–25,65%
Number of finished parts	8746 pcs	9891 pcs	13,09%
Capacity utilization	73,31%	72,35%	–1,31%

The discovered knowledge has resulted in the improvement of production process and simultaneously the defined goals have been fulfilled. It is important to mention that some parameters of production process got worse in some cases at the same time. Therefore it is necessary to determine the individual goals priority. Consequently the application of process knowledge discovery to the management strategies will be realized on the basis of priority system. Understandably the priority system will influence the parameters selection that will be monitored, collected and processed.

Because it is not possible to define generally validated priority system for the global objectives of production systems analyses, therefore it is necessary to solve each specific case individually. For example if it is important to increase the number of final production during the short time period then it will be needed to specify higher priority just to this parameter during the data mining process. The management strategy can be modified according to the new knowledge. The number of finished parts increases although some parameters of production systems can get worse simultaneously.

The very important part of this process is the determination of priority in the analysis of the production system. The priorities influence the stages of knowledge discovery process and also specifically gained knowledge.

V. Conclusion

The application of knowledge discovery process from databases in the production processes management will help to identify the impact of manufacturing parameters on the production process and the subsequent optimization of production process.

It can be used to predict failures, failed parts, emergency situations or states that can negatively affect the production process and thereby discovery knowledge that can help in the production process management. In the prediction scope, it can be further used to predict preventive controls of the production equipment, the cost of the production process, organization customers' behavior, etc.

The process implementation of knowledge discovery in the production systems management area can be used to achieve a better understanding of a production system. It can be also used to gain new and interesting knowledge for predicting future behavior of the production system. The new discovered knowledge will help managers in their decision making.

References

[1] P. Berka, *Knowledge discovering from databases.* Academia, 2003.
[2] U. M. Fayyad, G. Piatetski–Shapiro,G. P. Smyth, "From Data Mining to Knowledge Discovery: An Overview". *Advances in Knowledge Discovery and Data Mining*, MIT Press, pp. 1–37, 1996.
[3] P. Giudici, S. Figini, *Applied Data Mining for Business and Industry. 2nd Edition.* Wiley Computer Publishing, 2009.
[4] M. Kantardzic, J. Zurada, *Next Generation of Data-Mining Applications.* Wiley-IEEE Press, 2005.
[5] L. Lacko, *Oracle – Administration, programming and using database system.* Computer Press, 2007.

[6] L. Lacko, *Database: data warehouse, OLAP and data mining.* Computer Press, 2003.
[7] P. Schreiber, "Applications of Genetic Algorithms". *19th Central European Conference on Information and Intelligent Systems – CECIIS: Conference Proceedings*. University of Zagreb, 2008.
[8] D. Taniar, *Data Mining and Knowledge Discovery Technologies*. IGI Publishing, 2008.
[9] A. Trnka, "Classification and Regression Trees as a Part of Data Mining in Six Sigma Methodology". *WCECS 2010: World Congress on Engineering and Computer Science*, International Association of Engineers, 2010. pp. 449–453.
[10] C. Vercellis, *Business Intelligence: Data Mining and Optimization for Decision Making.* Wiley Computer Publishing, 2009.
[11] G. J. Williams, S. J. Simoff, *Data Mining – Theory, Methodology, Techniques, and Applications.* Springer, 2006.