International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:7, No:11, 2013

Knowledge Discovery from Production Databases for Hierarchical Process Control

Pavol Tanuska, Pavel Vazan, Michal Kebisek, Dominika Jurovata

***Abstract*—**The paper gives the results of the project that was oriented on the usage of knowledge discoveries from production systems for needs of the hierarchical process control. One of the main project goals was the proposal of knowledge discovery model for process control. Specifics data mining methods and techniques was used for defined problems of the process control. The gained knowledge was used on the real production system thus the proposed solution has been verified. The paper documents how is possible to apply the new discovery knowledge to use in the real hierarchical process control. There are specified the opportunities for application of the proposed knowledge discovery model for hierarchical process control.

***Keywords*—**Hierarchical process control, knowledge discovery from databases, neural network.

I. Introduction

As information technology is applied to more and more aspects of human life, produced and saved data grow proportionally. In the day-to-day operation of industry, administrations, offices, schools, hospitals, retail outlets, data is generated and saved. Similarly, industrial organizations increasingly save customer and supplier data. They administer orders, receipt cards, invoices and attempt to save as much data on the production process as is possible. Most organizations store data in their databases. However, they need access to useful information. The idea of an information society, or at the very least, the utilization of its strategic power found in data sources, requires not only new tools but also new way of thinking. The subject matter lies not only in the elaboration of new models. It is also about the acquisition of information on objects, their behavior, needs, covered relations, etc.

Production analysis and forecasting, prediction of production objectives, defects prediction in the production process, risk management and uncovering fraud are other examples of the fields of control that necessitate complete knowledge. Of relevance here is knowledge control or about knowledge systems control and the process whereby responsible individuals become knowledge operators. The process of knowledge discovery in databases, often referred to as data mining, represents the first important step in knowledge control technology.

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. Tanuska is with Slovak University of Technology in Bratislava, Faculty of Material Science and Technology in Trnava (phone: +421918646061; e-mail: pavol.tanuska@stuba.sk).

P. Vazan, M. Kebisek, and D. Jurovata are with the Slovak University of Technology in Bratislava, Faculty of Material Science and Technology in Trnava (e-mail: pavel.vazan@stuba.sk, michal.kebisek@stuba.sk, dominika.jurovata@stuba.sk).

II. Hierarchical Control Model

The current information and control systems come from well-known pyramid model of hierarchical process control [9]. Many of hierarchical control systems are built as a multiprocessors control systems with horizontal and mainly vertical communication. The intelligent elements as a sensors and actuators begin to apply in systems, whereby direct hierarchical relationship are changed into network oriented.
The emergent trends are coming through, i.e. connecting of independent systems. It can lead into creation of new behavior attributes for new system [5].

Therefore also process control is realized nowadays by implementation of hierarchical structure of control systems.

The pyramid model (that is shown on Fig. 1) is widely established model of complex process control.

There is possible to say that there is enormous volume of redundant data. These data are generated on all levels of control systems. Therefore is necessary to find new effective manners how these data at first to store and subsequently to extract useful and usable information for process control.

Extracting of useful, prospective usable information is very important on all levels of process control. Each level has got a specific assignment in the hierarchical control. On the top level there is the priority focused to production effectiveness, but on the control level there is the priority focused to important real time information suitable for operators.

On the basis of former statements about inter level communication is clear that conventional approach for data processing in the pyramid model of process control is not suitable to use. The real solution that can be used is exploitation of special data storages and modern methods application of data processing. One of the existing and suitable methods is KDD – knowledge discovery in databases.

III. The Proposal of Process of Knowledge Discovery For Hierarchical Process Control

The process of knowledge discovery in databases is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [2].

KDD is defined as a broad-ranging process that comprises a series of specific steps including data preparation, searches for regularities, verification, testing and the refinement of

discovered knowledge, where everything is repeated in numerous iterations. Non-triviality requires that the aforementioned steps include further partial tasks. It is not the direct calculation of advance known quantities. Legitimacy considers some rate of certainty and is not considered an absolute. The novelty and usefulness of the process relates to the user. It is aimed at mined data contributions and their possible transformation into knowledge.

A. Control Level

The proposed and designed control system generally consists of the following subsystems on the control levels:

i. subsystem of measurement, collection and processing of information (including sensors) and transmission routes,
ii. control subsystem (including the transmission of the control signal to the actuators) [6],
iii. visualization subsystem and operator communication with control system (SCADA / HMI),
iv. real time databases (make the storage of data from technological process and data transmission into data analysis level),

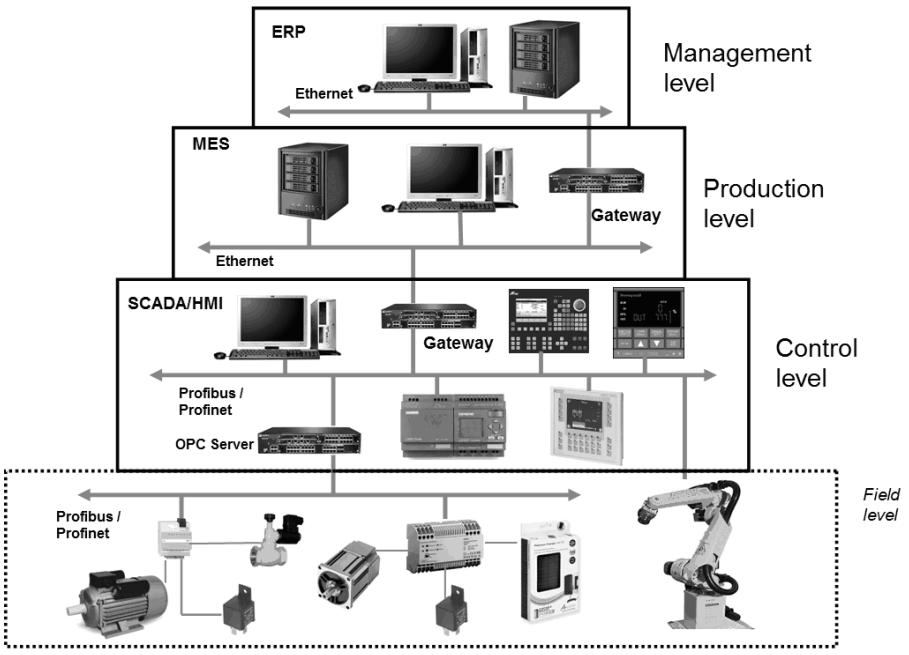

Fig. 1 Data communication in the pyramid model of hierarchical process control

Comprehensibility often lacks and is subject to further processing, e.g. via data visualization. The usefulness relates directly to interest and is taken as a total rate of pattern value combining its legitimacy, novelty, usefulness and simplicity. Functions of usefulness can be defined explicitly, or as the patterns that produce the answer to the requirements arranged due to the interest directed by KDD system.

The proposed solution for knowledge discovery for hierarchical process control is shown on Fig. 2.

The complete solution is divided into three levels. The basic level, which is information connected to the technology and / or manufacturing process, is called the control level. It goes out from the standard pyramid model.

v. information systems on the different control levels,
vi. subsystem of integration of information and control systems,
vii. support for control system (simulates various models of technological and/or production process and production strategies, etc.).

The proposed solution includes all relevant elements and subsystems that are indispensable for the process control of any industrial enterprise and provides, among others, the following functions:

i. measurement of immediate quantities values,
ii. monitoring of quantities (e.g. limit values, trends, etc.),
iii. control of actuators,
iv. manual operations,
v. input setpoints of controllers and logic control,
vi. data visualization,
vii. technological data storage,
viii. information transmission into the superior level,
ix. monitoring of production goals,
x. transfer of information into the management level.

Control systems for different types of processes are generally characterized by the following features i.e. technology of control system is generally distributed, the control of process requires rapid (dynamic) and failure-free measurements, the control of process also requires a fast and

reliable communications subsystem for information transfer, i.e. short reaction times for transfers, the control of actuators has to be fast (dynamic) and reliable, high operational reliability control system at hardware or software failures, the collection and processing of measured signals is executed cyclically at exactly defined intervals etc.

The mentioned characteristics of control systems currently proposed, designed and implemented by hardware and software components that are based on the local area network (decentralized control systems – digital control elements are interconnected by communication subsystem).

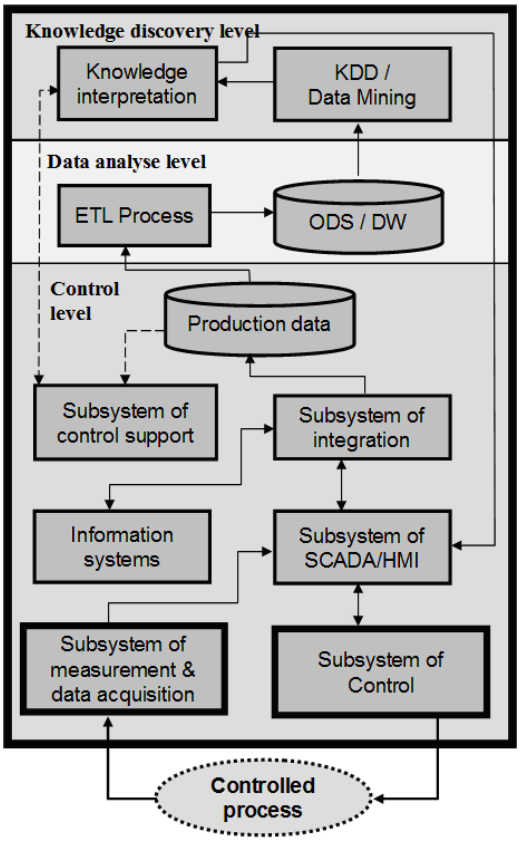

Fig. 2 Conceptual model proposal of knowledge discovery for hierarchical process control

B. Data Analysis and Knowledge Discovery Levels

The next level, that is necessary to understand as a superior level, is the data analysis level. This includes the subsystem for data collection, extraction, transformation and integration, including the data warehouse or operation data store [4]. This system includes OLAP technology.

The fundamental request for the data extracting and data processing from production databases is the data integrity retention. According to transformation and data processing from the different data structures in the data store we can say that this resolvable request is extremely complicated. The potential properties, which are hidden in huge amount of data, is necessary to prepare for the processing in several steps and besides inviolate the relations and connections in this data.

It is especially complicated, when data arise from systems, which are specified as safety-critical.

The safety-critical system is the system, whose incorrect functionality (failure) can have catastrophic consequences, e.g. serious injuries or casualties, enormous damage to property or environment [7].

One of the basic properties of safety-critical system is the real time, which is included in requests for almost all safety-critical systems. We do not only require to obtain reliable results of measuring and processing (technological data), but we have to obtain them at the right time.

Today we meet safety-critical processes significantly more than in the past. They are part of not only services e.g. in the chemical and pharmaceutical industry, nuclear energy, in the management of combustion process and many other services, but this already includes the management of basic processes in production systems, including machine engineering.

From the above, it can be concluded that properly designed data pump, often called the ETT process (ETL), it is the alpha and omega of good data warehouse design. The basic factor to ensure the correct functioning of the data pump is the verification and validation [3].

The last level, which should be considered as the highest is the knowledge discovery level. This level includes KDD subsystem and including of the subsystem for interpretation of knowledge.

Only correctly gained (by using methods and techniques of data mining) and interpreted knowledge can be effectively used to increase the quality management processes. Therefore the most important part of the system is knowledge interpretation module. Properly interpreted knowledge is then necessary to use a secure back transport into the production process. This is done through the feedback as a special function of SCADA system. This should be realized as an interface between the level of knowledge discovery and management level.

There is necessary to said that interventions from gained and correctly interpreted knowledge may be carried out in manual or automatic mode.

The information flow from the level of knowledge discovery on the control level can include the parameters of control algorithms, values of balance calculations, static and dynamic models parameters, parameters useful for diagnosis devices – maintenance support, documentation to ensure product quality etc.

IV. The Application of Knowledge Discovery For Hierarchical Process Control

There are many problems in the production process that can be resolved by the process of knowledge discovery in databases. Correct problem identification and a corresponding solution are important. The process of knowledge discovery in databases provides sufficient means to deal with the problems that arise in the field of production systems.

In relation to the subject matter there are two key perspectives on data mining – description and prediction.

Within the term of data mining we can associate all „more complex" activities over the database, or possibly data warehouse. Precisely, we can determine data mining as a specific process of acquiring in advance unknown information. Here advanced tools prepared by specialists have to be available and the end user is in a better position, when user obtains new and often in advance unknown information [8].

There are many definitions of data mining. The following is one of the key definition:

Fayyad: „Data mining is a single step in the process of knowledge discovery in databases that involves finding patterns in the data." [1]

The application of the proposed procedures was realized on the prediction problem of the goal parameters of the production management system. These problems are typical for the production management level.

The following goals were defined:

i. the analysis of manufacturing process parameters influence on the capacity utilization,
ii. the analysis of manufacturing process parameters influence on the throughput times of production batches,
iii. the analysis of manufacturing process parameters influence on the number of finished parts.

We used Statistica Data Miner tool for the production system data analysis. The tool allows the application of the process of knowledge discovery in databases from data acquisition, through their transformation and modification, and data mining in evaluation of the achieved results.

Data from production database was recorded into the internal table after defining the selection query for retrieving information from database and switching into the Statistica Data Miner tool. The internal table lines correspond with individual lines obtained from the selected query and the columns represent the specific query attributes. This recorded data can be the modified according to the format required; the user can modify the names representing the particular variables, sorting etc.

The modification and transformation of data collected for database was the next step in applying the process of knowledge discovery in databases for the proposed production system. The basic parameters necessary for the analysis were identified: individual production batch sizes, production system input intervals, production capacity utilization, and throughput time when the production batch remained in the production system. All of these parameters, except throughput time, are recorded directly from the production system. Since the production process data is saved after one specific operation execution, it is possible to calculate the throughput time of this data. The continuous time is the difference between the end of the last production operation time and the beginning of the first production operation time on the specific production batch. To calculate the value, we utilized the possibility of complementing one variable directly in the Statistica Data Miner application table, where the function for continuous time of the production calculation was defined.

The next step was to determine whether the set of data did not comprise data whose values significantly differ from the other data (outliers). It was necessary to find not only whether the set of data comprised such data but also the cause of their existence. They might be random data, whose extreme value was, for example, caused by a mistake in assigning. However it could also concern significant values. Therefore it was necessary to correctly identify the origin of the values and decide whether the values would be included or eliminated in the set of data used. To identify the significantly different data a Frequency table was used (Fig. 3).

Frequency table: Throughput time (DataSource)

From To	Count	Cumulative Count	Percent	Cumulative Percent	100% - Percent	Logits	Probits	Normal Expected	Cumulative Normal
0,000000<=x<120,0000	0	0	0,00000	0,0000	100,0000			1,66799	2,8323
120,0000<=x<240,0000	0	0	0,00000	0,0000	100,0000			3,56379	6,3961
240,0000<=x<360,0000	0	0	0,00000	0,0000	100,0000			7,02594	13,4221
360,0000<=x<480,0000	0	0	0,00000	0,0000	100,0000			12,78121	26,2033
480,0000<=x<600,0000	56	56	7,14286	7,1429	100,0000	-2,56495	-1,46523	21,45436	47,6576
600,0000<=x<720,0000	35	91	4,46429	11,6071	92,8571	-2,03017	-1,19486	33,23044	80,8881
720,0000<=x<840,0000	47	138	5,99490	17,6020	88,3929	-1,54355	-0,93064	47,49352	128,3816
840,0000<=x<960,0000	71	209	9,05612	26,6582	82,3980	-1,01204	-0,62318	62,63414	191,0157
960,0000<=x<1080,000	71	280	9,05612	35,7143	73,3418	-0,58779	-0,36611	76,21962	267,2354
1080,000<=x<1200,000	63	343	8,03571	43,7500	64,2857	-0,25131	-0,15731	85,58582	352,8212
1200,000<=x<1320,000	60	403	7,65306	51,4031	56,2500	0,05614	0,03518	88,67810	441,4993
1320,000<=x<1440,000	90	493	11,47959	62,8827	48,5969	0,52719	0,32875	84,78335	526,2826
1440,000<=x<1560,000	126	619	16,07143	78,9541	37,1173	1,32216	0,80483	74,79702	601,0797
1560,000<=x<1680,000	71	690	9,05612	88,0102	21,0459	1,99340	1,17550	60,88879	661,9684
1680,000<=x<1800,000	61	751	7,78061	95,7908	11,9898	3,12490	1,72691	45,73717	707,7056
1800,000<=x<1920,000	14	765	1,78571	97,5765	4,2092	3,69544	1,97323	31,70149	739,4071
1920,000<=x<2040,000	4	769	0,51020	98,0867	2,4235	3,93704	2,07200	20,27532	759,6824
2040,000<=x<2160,000	3	772	0,38265	98,4694	1,9133	4,16408	2,16208	11,96554	771,6480
2160,000<=x<2280,000	0	772	0,00000	98,4694	1,5306	4,16408	2,16208	6,51588	778,1638
2280,000<=x<2400,000	0	772	0,00000	98,4694	1,5306	4,16408	2,16208	3,27407	781,4379
2400,000<=x<2520,000	1	773	0,12755	98,5969	1,5306	4,25238	2,19643	1,51802	782,9559
2520,000<=x<2640,000	1	774	0,12755	98,7245	1,4031	4,34899	2,23359	0,64944	783,6054
2640,000<=x<2760,000	2	776	0,25510	98,9796	1,2755	4,57471	2,31876	0,25637	783,8617
2760,000<=x<2880,000	2	778	0,25510	99,2347	1,0204	4,86497	2,42505	0,09338	783,9551
2880,000<=x<3000,000	3	781	0,38265	99,6173	0,7653	5,56196	2,66700	0,03139	783,9865
3000,000<=x<3120,000	2	783	0,25510	99,8724	0,3827	6,66313	3,01722	0,00973	783,9962
3120,000<=x<3240,000	0	783	0,00000	99,8724	0,1276	6,66313	3,01722	0,00279	783,9990
3240,000<=x<3360,000	1	784	0,12755	100,0000	0,1276			0,00074	783,9998
3360,000<=x<3480,000	0	784	0,00000	100,0000	0,0000			0,00018	783,9999
Missing	0	784	0,00000	100,0000	0,0000				

Fig. 3 Frequency table

The set of data was then subjected to investigation as to whether it comprised the missing or noised data in some of its parameters.

This modified set of data produces the input set of data for further progress in the process of knowledge discovery in databases.

The next step is data mining. Proposed data mining models contain several data mining methods and techniques as neural networks, classification and regression trees, cluster analysis etc. The selected model is shown on Fig. 4.

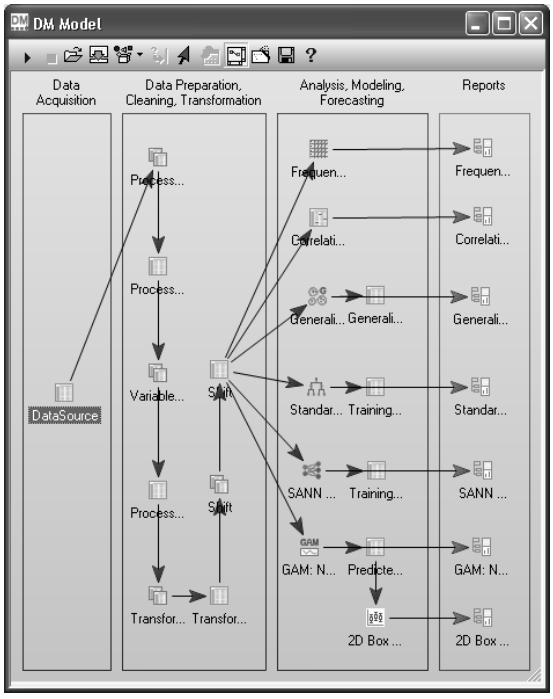

Fig. 4 Data mining model

All of the reports for each data mining model were independently saved again for analysis of the achieved results. The results were evaluated with respect to the data mining outcome reports on particular data mining methods and techniques. The evaluation was carried out independently of defined objectives.

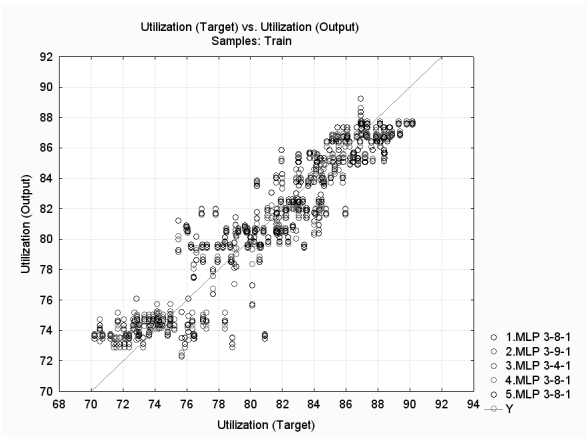

Fig. 5 Results obtained from data mining process – neural networks results

The prediction of production capacity utilization by neural network usage is presented on Fig. 5. It is obvious that the neural network is able to very accurately predict the targeted value. The gained predictions are possible directly to apply as knowledge for hierarchical process control on the base of gained results.

The number of finished products prediction on dependence of Product2 lot size is shown on Fig. 6.

We can state that the results are directly usable in hierarchical process control on the base of verification in the existing simulation model.

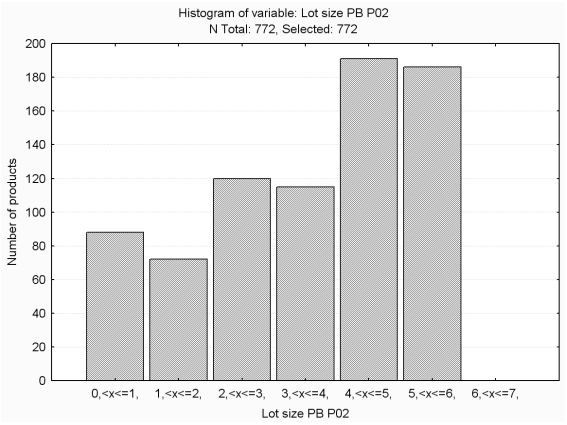

Fig. 6 Results obtain from data mining process – histogram of review for lot size

The rest of defined goals have been evaluated individually according to gained results.

V. Conclusion

The new discovered knowledge is possible to apply on the base of proposed model of knowledge discovery into the real hierarchical process control. The gained knowledge is obtained on the management level and then they are used on the lower levels.

The application of the proposed solution in practice can be useful for solution the following problems:

i. The critical states prediction of controlled process bases on the principle of finding of analogical situations by processing of large amounts of data in real time. The solution is taking shape in the creation of a typical critical states library in off-line mode.
ii. The prediction of production devices preventive controls. This prediction has a significant relationship to maintenance. The periodical maintenances are expensive and furthermore the plan of downtime accurately does not match with lifetime of individual components of the devices.
iii. The identification of production parameters influence on the production process.
iv. The identification slightly incorrect information sources (sensors). Usually the standard techniques for range estimation of the alarms fail.
v. The diagnostic of production systems with respect to total lifetime of the systems.
vi. The identification and optimization of relevant parameters control that have the impact on the increasing of industrial processes control safety.

vii. The fail operations of actuators like insufficient realization of the calculated action.
viii. More precise specification of non-linear dynamic models of controlled processes with the objective to optimization of parameters.
ix. The continuous monitoring of the control process quality on the base of quality evaluation from online gained knowledge.
x. The detection of failed states of production devices and products – to reveal occurrence of defective products.
xi. The identification various non-standard states that have the influence on the production process and that the operator has to solve by unplanned shutdown of a machine or technology.
xii. The problem solution by using the gained knowledge without pre defined goal.
xiii. The predictions for needs enterprise management and different ad hoc reports.
xiv. The effective implementation and especially innovation of the management systems at all levels.

REFERENCES

[1] U. M. Fayyad, "Data Mining and Knowledge Discovery: Making Sense Out of Data". *IEEE Expert/Intelligent Systems & Their Applications*, pp. 20–26, 1996.
[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] R. Halenar, "Matlab Routines Used for Real Time ETL Method". *Applied Mechanics and Materials*, pp. 2125-2129, 2012.
[4] R. Halenar, "Real Time ETL Improvement". *International Journal of Computer Theory and Engineering*. vol. 4, no. 3, pp. 405-409, 2012.
[5] J. Jadlovsky, "Proposal of distribution control system of FMS" in *International Conference Cybernetics and Informatics*, Vysna Boca 2010.
[6] P. Mydlo, T. Skulavik, P. Schreiber, "The fuzzy PI controller`s chosen parametres influence on the regulation process" in *Process Control 2010*, University of Pardubice, pp. C055a1-8, 2010.
[7] M. A. Schwarz, *Introduction to software engineering for secure and reliable software – Einführung in die Softwaretechnik für sichere und verlässliche Software*. Institut für Informatik im Paderbom, 2004.
[8] A. Trnka, "Classification and Regression Trees as a Part of Data Mining in Six Sigma Methodology" in *World Congress on Engineering and Computer Science*, Hong Kong: International Association of Engineers, pp. 449-453, 2010.
[9] A.-W. Sheer, *CIM Computer Integrated Manufacturing*. Berlin: Springer-Verlag, 2011.