National Conference on Advances in Computer Science and Applications with International Journal of Computer Applications (NCACSA 2012)
Proceedings published in International Journal of Computer Applications® (IJCA)

A Survey of Data mining techniques for quality improvement in process industries

V.Sharmila, PG Scholar
Computer Science and Engineering
Velammal Engineering College
Chennai, India

M.Shanmugasundaram, Asst. Prof
Computer Science and Engineering
Velammal Engineering College
Chennai, India

ABSTRACT
This paper presents a survey of quality improvement of various products in process industries and present how data mining can be used in manufacturing for improving the quality, maintenance, control and production of a product.

Keywords
Quality Improvement, Data Mining, Manufacturing, process industries, product

1. INTRODUCTION

Knowledge is the most valuable asset of a manufacturing enterprise, as it enables a business to differentiate itself from competitors and to compete efficiently and effectively to the best of its ability. Nowadays, manufacturing enterprises have to stay competitive in order to survive the competition in the global market. Quality, cost and cycle time are considered as decisive factors when a manufacturing enterprise competes against its peers. Among them, quality is viewed as the more critical for getting long-term competitive advantages. The development of information technology and sensor technology has enabled large-scale data collection when monitoring the manufacturing processes. Those data could be potentially useful when learning patterns and knowledge for the purpose of quality improvement in manufacturing processes. However, due to the large amount of data, it can be difficult to discover the knowledge hidden in the data without proper tools. Although traditional data analysis tools have been successfully used in improving quality of products and processes, now better tools exist to mine massive data sets collected through computerized systems in the industry.

Data mining refers to extracting or "mining" knowledge from large amounts of data. It is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories. The use of data mining techniques in manufacturing began in the 1990s and it has gradually progressed by receiving attention from the production community. Data mining is now used in many different areas in manufacturing engineering to extract knowledge for use in predictive maintenance, fault detection, design, production, quality assurance, scheduling, and decision support systems. Data can be analyzed to identify hidden patterns in the parameters that control manufacturing processes or to determine and improve the quality of products. A major advantage of data mining is that the required data for analysis can be collected during the normal operations of the manufacturing process being studied and it is therefore generally not necessary to introduce dedicated processes for data collection.

2. DATA MINING IN MANUFACTURING

Data mining can be used in manufacturing, especially in the areas of production processes, control, maintenance, customer relationship management, decision support systems, quality improvement, fault detection, and engineering design.
Giess et al. [1,2] mined a manufacturing and assembly database of gas turbine rotors to determine and quantify relationships between the various balance and vibration tests and highlight critical areas. This knowledge could then be fed back to the designers to improve tolerance decisions in the future design of components. They used a decision tree at the initial stage to determine appropriate areas of investigation and to identify problems with the data. At the next stage, a neural network was used to model the data.
M Perzyk, R Biernacki, and J Kozlowski [3] have used data mining in manufacturing to perform a significance analysis of process parameters. They have proposed and tested a methodology of determination of relative significances of process variables and possible interactions between them, based on interrogations of generalized regression models. The performance of several types of data mining tool, such as artificial neural networks, support vector machines, regression trees, classification trees, and a naive Bayesian classifier, is compared. Also, some simple non-parametric statistical methods, based on an analysis of variance (ANOVA) and contingency tables, are evaluated for comparison purposes. They found that performance of significance and interaction factors obtained from regression models, and, in particular, neural networks, is satisfactory, while the other methods appeared to be less accurate and/or less reliable. The proposed methodology exhibited some remarkable advantages over the other methods, already present in some DM and statistical software packages, and could be profitably included in such packages.

Chen-Fu Chien et al. [4] used Data mining for yield enhancement in semiconductor manufacturing. The target variable used in this study is the yield rate that is like a synthetic index of the performance of hundreds. of processes. They have developed a framework for data mining and knowledge discovery from database that consists of a Kruskal–Wallis test, K-means clustering, and the variance reduction splitting criterion to investigate the huge amount of semiconductor manufacturing data and infer possible causes of faults and manufacturing process variations. They validated this approach with an empirical study in a semiconductor foundry company and demonstrated the practical viability of this approach.

3. DATA MINING FOR QUALITY IMPROVEMENT

Bashar Al-Salim, Mansour Abdoli [5] used data mining for decision support of quality improvement**.** The method advances the decision support system of the quality improvement process by grouping the related quality problems in two steps:. First, a soft grouping is performed using association rules as a data mining technique, and then, resulted groups are finalized by employing a costs minimization model. Moreover, to find the optimal groups, a mathematical programming language is used. Results show that this methodology is beneficial and attractive in making the quality improvement process more efficient and in providing support to managerial decisions for creating quality improvement teams.

Shu-guang He1, Zhen He, G. Alan Wang and Li Li [6]In their work on data mining for Quality Improvement in Manufacturing Processes, have proposed a knowledge based six-sigma model where DMAIC for six-sigma was used along with data mining techniques. The knowledge learned by the data mining techniques was helpful in identifying potential quality problems and assisting quality diagnosis. They also examined the problem of parameter optimization by applying data mining techniques to DOE (Design Of Experiments). A decision tree was built in order to dynamically adjust parameter optimization. In addition, they applied data mining to quality diagnosis where an association rule mining technique was used to analyze the electronic test data. The rules obtained by data mining provided a direct guidance in identifying the root causes of the quality problems. And the findings were beneficial for quality diagnosing that is still a difficult problem in six-sigma. Furthermore, a decision tree was also used in the service data analysis. The findings were valuable to improve product design. Finally, they presented a system infrastructure for quality improvement in manufacturing processes.

Gulser Koksal, Inci Batmaz, Murat Caner Testik [7] has reviewed data mining applications for quality improvement in manufacturing industry. They have done an extensive review covering the literature from 1997 to 2007 and several analyses on selected quality tasks are provided on DM applications in the manufacturing industry. The quality tasks considered are: product/process quality description, predicting quality, classification of quality, and parameter optimisation. The review provides a comprehensive analysis of the literature from various points of view: data handling practices, DM applications for each quality task and for each manufacturing industry, patterns in the use of DM methods, application results, and software used in the applications are analysed.

4. DATA MINING IN CHEMICAL PROCESS INDUSTRY

Advancements in sensors and database technologies have resulted in the collection of huge amounts of process data from chemical plants. A number of process quantities such as temperature, pressure, flow rates, level, composition, and pH can be easily measured. Chemical processes are dynamic systems and are equipped with hundreds or thousands of sensors that generate readings at regular intervals (typically seconds). In addition, derived quantities that are functions of the sensor measurements as well as alerts and alarms are generated regularly. Several commercial data warehouses, referred to as plant historians in chemical plants are in common use today around the world. These historians store large amount (weeks) of historical process operation data. This data is available for mining, analysis and decision support – both real-time and offline.

Given that large amounts of operational data are readily available from the plant historian, data mining can be used to extract knowledge and improve process understanding – both in an offline and online sense. There are two key areas where data mining techniques can facilitate knowledge extraction from plant historians, namely (i) process visualization and state-identification, and (ii) modelling of chemical processes for process control and supervision. Visualization techniques use graphical representation to improve human's understanding of the structure in the data. These techniques convert data from a numeric form into a graphic form that facilitates human understanding by means of the visual perception system. This enables post-mortem analysis of operations towards improving process understanding or developing process models or online decision support systems.

Data-based models are also frequently used for process supervision – fault detection and identification (FDI). The objective of FDI is to decide in real-time the condition – normal or abnormal – of the process or its constituent equipment, and (ii) in case of abnormality, identify the root cause of the abnormal situation. It has been reported that approximately 20 billion dollars are lost on an annual basis by the US petrochemical industries due to inadequate management of abnormal situations. Efficient data mining algorithms are hence necessary to prevent abnormal events and accidents. In the chemical industry, pattern recognition and data classification techniques have been the popular approaches for FDI. When a fault occurs, process

variables vary from their nominal ranges and exhibit patterns that are characteristic of the fault. If the patterns observed online can be matched with known abnormal patterns stored in a database, the root cause of a fault can generally be identified.

Jayanthi Ranjan [8] has analysed the applications of data mining techniques in pharmaceutical industry. She explains the role of data mining in pharmaceutical industry. She shows how data mining on large sets of data uses tools like association, clustering, segmentation and classification for helping better manipulation of the data help the pharma firms compete on lower costs while improving the quality of drug discovery and delivery methods.

Dan Braha and Armin Shmilovici [9] have shown how data mining can be used for improving a Cleaning Process in the Semiconductor industry. They have presented a comprehensive and successful application of data mining methodologies to the refinement of a new dry cleaning technology that utilizes a laser beam for the removal of micro-contaminants. They suggest that data mining methodologies may be particularly useful when data is scarce, and the various physical and chemical parameters that affect the process exhibit highly complex interactions. Another implication is that on-line monitoring of the cleaning process using data mining may be highly effective.

Mohamed Azlan Hussain has reviewed the applications of neural networks in chemical process control using simulation and online implementation. He provides an extensive review of the various applications utilizing neural networks for chemical process control, both in simulation and online implementation. He has highlighted the broad, extensive and continuing increase in the application of neural network in many chemical process control applications, both online and in simulation.

5. CONCLUSION

In process industries, data historians continuously collect process data from most of the process variables through the whole process chain for several months and sometimes for years. But the information contained in this data is very rarely used. Using a combination of machine learning, statistical analysis, modelling techniques and database technology, data mining finds patterns and subtle relationships in data and infers rules that can be beneficially used for quality improvement in process industries.

6. REFERENCES

[1] Giess, M. D., Culley, S. J., and Shepherd, A., 2002, "Informing Design Using Data Mining Methods," *ASME DETC*, Montreal, Canada, pp. 98–106

[2] Giess, M. D., and Culley, S. J., 2003, "Investigating Manufacturing Data for Use Within Design," *ICED 03*, Stockholm, Sweden, pp. 1408–1413

[3] Perzyk, M., Kochanski, A., and Kozlowsk.J, 2008, "Data mining in manufacturing: significance analysis of process parameters" Proc. IMechE Vol. 222 Part B: J. Engineering Manufacture pp 1503 – 1516.

[4] Chen-Fu Chien, Wen-Chih Wang, Jen-Chieh Cheng, Data mining for yield enhancement in semiconductor manufacturing and an empirical study" Expert Systems with Applications 33 (2007) 192–198

[5] Bashar Al-Salim, Mansour Abdoli, "Data mining for decision support of the quality improvement process" Proceedings of the Eleventh Americas Conference on Information Systems, Omaha, NE, USA August 11th-14th 200 .pp 1462-1469

[6] Shu-guang He1, Zhen He, G. Alan Wang and Li Li, "Data Mining and Knowledge Discovery in Real Life Application", Book edited by: Julio Ponce and Adem Karahoca, pp. 438, February 2009, I-Tech, Vienna, Austria

[7] Gulser Koksa, Inci Batmaz, Murat Caner Testik, "A review of data mining applications for quality improvement in manufacturing industry" Expert Systems with Applications 38 (2011) 13448–13467

[8] Jayanthi Ranjan "Applications of data mining techniques in pharmaceutical Industry "Journal of Theoretical and Applied Information Technology (2005-2007).,pp(61-67)

[9] Dan Braha and Armin Shmilovici, "Data Mining for Improving a Cleaning Process in the Semiconductor industry" IEEE transactions on semiconductor manufacturing, VOL. 15, NO. 1, FEBRUARY 2002 pp (91-101)