* For the reference see also the last page

DATA MINING APPLICATIONS IN INDUSTRIAL ENGINEERING: A PERSPECTIVE

T. W. Liao[1], J.-H. Chen[2], and E. Triantaphyllou[1]

[1]Industrial & Manufacturing Systems Engineering Department
[2]Computer Science Department
Louisiana State University, Baton Rouge, LA 70803

ASBSTRACT

Due to its power to discover otherwise hidden knowledge, data mining has become a new tool for processing large volumes of data stored in databases. This paper offers our perspective of potential data mining applications in industrial engineering. Several applications, which include parts similarity-based GT formation, interpretation of statistical data, diagnosis of muscle fatigue, and monitoring of machine condition, are detailed in this conference.

Key words: Data mining, Industrial engineering, Condition diagnosis and monitoring, Group technology, Pattern recognition.

DATA MINING

Data Mining (DM) and Knowledge Discovery in Databases (KDD) has emerged as one of the most significant and fast growing areas in computer science and other related disciplines. DM/KDD addresses the important issue of extracting meaningful patterns ("knowledge") from huge dataset, which is vital for intelligent decision-making in the face of explosive growth of data. Fayyad *et al.* (1996) defined knowledge discovery in databases as the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.

Here the data D is a set of facts stored in a database, which describes some aspects of the real-world. For example, in the detection of fraudulent cellular phone calls, the data set is a collection of records from a tele-communication company's billing database. Each record may have the four fields: domestic/international, duration, starting time, and status (fraud or not) of the call. A pattern of fraud calls can be "international calls with duration over 30 minutes". In principle, a pattern P is an expression in some language that captures a subset, D_i, of the data such that P is simpler than the enumeration of D_i. Obviously, not all patterns are of interest to KDD. The validity of patterns refers to the requirement that the patterns should be "correct" in predicting new data with some degree of certainty or accuracy. This is a reasonable requirement. The requirements of novelty, potential usefulness and understandability of patterns can also be easily justified. The notion of *interestingness* is usually taken as a comprehensive measure of pattern value, combing validity,novelty, usefulness and simplicity. 'Some KDD systems employ explicit interestingness measures to evaluate the merit of the patterns discovered.

Fayyad *et. al.* (1996) stated that KDD is an iterative, multi-step process with many decisions being made by the user. Before starting the KDD process, one must have a good understanding of the application domain, the relevant domain knowledge, the type of patterns sought, and the goals of mining with respect to the application. Following (Fayyad *et. al.*, 1996], the KDD process can be broadly outlined as consisting of the following steps:

(1) *Data selection.* This is the first step of the KDD process. A subset of data is selected from the massive amount of data stored in the entire database. Data mining will be performed on this selected data subset. This requires prior knowledge about the nature of data, and relevance of subsets of data to the knowledge discovery task. The selected data may be obtained from multiple, heterogeneous databases.

(2) *Data pre-processing and cleaning.* "Real-world" data is usually not clean, involving noise or outliers, errors, missing values, etc. Thus, a pre-processing and cleaning step is needed to prepare the data for the mining task. Statistical techniques are often used to handle noise removal and missing values. Prior knowledge about the application domain can also be useful here.

(3) *Data reduction and transformation.* Many real-world datasets have high dimensionality or large number of features. This step aims at selecting the most relevant features for the data mining task and thus reducing the number of features needed. Dimensionality reduction and transformation techniques are used in this step.

(4) *Data mining.* This step involves choosing suitable data mining algorithms and applying them to discover patterns in some forms. Patterns can be in the form of decision trees or rules, clusters in a multi-dimensional space, a Bayesian network, etc. Popular data mining algorithms include decision tree/rule learning methods, neural network techniques, statistical/Bayesian approaches, etc., and hybrid methods. A human user's knowledge in the application domain may also be incorporated to help the pattern finding process.

(5) *Result interpretation and evaluations.* During this step, the discovered patterns are interpreted and their merit assessed. According to (Fayyad *et. al.*, 1996), the primary goals of data mining are prediction and description. In prediction, we would like to predict or estimate the values of some target variables from values of other variables, or from the past known values of the target variable. Stock market index value prediction is an example of a prediction task. On the other hand, we may also be interested in finding descriptions and abstraction of huge amounts of data for a better understanding. Many data mining tasks involve a combination of prediction and description. Typical data mining tasks include: classification, clustering, summarization, dependency modeling, novelty and deviation detection, and regression.

KDD has found many successful real-world applications. In business and finance, KDD has been used for predicting investment returns (Apte and Hong, 1996). In science, successful applications include automatic catalog of sky surveys (Weir *et. al.*, 1995), gene discovery from DNA sequences and protein sequence matching (Kulp *et. al.*, 1996), and detections of earthquake from satellite imagery data (Stolorz *et. al.*, 1995), etc. In the following, we present our perspective of potential data mining applications in industrial engineering.

INDUSTRIAL ENGINEERING APPLICATIONS

The explosive growth in data gathering and the use of large databases has generated a need for new techniques and tools that can intelligently and automatically transform the processed data into useful information and knowledge. Some data mining applications in industrial engineering are described below. Our view is in no way complete. Nevertheless, we do hope to generate a higher interest in this exciting area.

Mining of Parts Similarity Knowledge

Utilizing a similar part which has previously been designed in a new product could reap big benefits for the company such as shorter product development process (time to market), reduced number of parts, and lower purchasing or manufacturing costs. To achieve that, a mining tool capable of searching for similar parts at different stages of the product design process is necessary because of the massive part data resided internally or externally in the suppliers' databases. What makes it more complicated is that older parts might only be available in legacy data incompatible with the current system. Smith *et al.* (1997) described a neural design information retrieval system. Dowlatshahi and Nagaraj (1998) developed a classification and coding system for rapid design retrieval of all design data pertaining to the manufacturing of machine tools.

Parts that require similar processing operations are often identified to be manufactured by the same manufacturing cell to reduce handling, setup, work in process, and manufacturing lead time (the cellular manufacturing concept). From the manufacturing perspective, knowledge about parts similarity are often discovered from process plans or manufacturing-based GT codes. It is quite likely that, for a given part, more than one process plan coexist, developed by a different engineer at a different time. Please refer to Selim *et al.* (1998) for works done in this area.

Mining of Condition Interpretation Knowledge

In a manufacturing environment, the workhorses are operators, tools, and equipment. As the operation progresses, their condition unavoidably will undergo changes. For instance, an operator will fatigue. A tool will dull and break. A worn tool in turns will effect the process behavior such as chattering and subsequently the work quality. A broken tool must be stopped immediately to discontinue nonproductive operations and to avoid its damage to the workpiece. Therefore, the condition of the workhorse must be monitored and diagnosed in order to derive a decision as to restore the working ability of the workhorse or to replace it with a brand new one. The condition-based maintenance of critical components is key to prevent losses before a fault or failure actually occurs. To achieve that, a condition monitoring system must be in place, which should have the following components: sensors for data acquisition, data preprocessing to suppress noise, feature extraction to identify a small set of discriminate features, and decision-making knowledge to determine the condition. The trend of using multiple sensors to increase a system's reliability greatly increases the amount of data to be processed. Please refer to Dimla *et al.* (1997), and Mitchell (1993) for past research in the monitoring, diagnosis and interpretation of operator condition, tool condition, and machine condition.

Mining of Work Quality Knowledge

Once manufactured, each product must be inspected and tested to verify its quality against the specified standard. Depending upon the requirements, various inspections and tests must be conducted using the appropriate measuring equipment and methods. For machined components, quality characteristics of interest include dimensions, surface roughness, strength, etc. They can be measured by a coordinate measuring machine, a profilometer, and a universal testing machine. On the other hand, flaws such as cracks and cavities might be of particular interest for castings (Kehoe and Parker, 1992) and welds (Liao and Li, 1998). Their detection normally requires a nondestructive evaluation (NDE) method such as radiography and ultrasonic.

CONCLUDING REMARKS

Due to space limitation, many other applications including process fault diagnosis and failure analysis which are important to continuous process manufacturers cannot be covered. Please check our website at http://cda4.imse.lsu.edu for more details. We are also unable to get into the technical details of the data mining process here. We plan to do that in the near future. One issue that is quite clear though is that almost all past applications were limited to small data sets. In order to bridge the gap between current research and data mining research, there is a need to address the database issue with large volume of data.

REFERENCES

Apte, C. and S. J. Hong. In: *Advances in Knowledge Discovery and Data Mining*, U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (eds.), AAAI Press, 1996, pp. 541-560.

Dimla, D. E., Jr., Lister, P. M., and Leighton, N. J., *Int. J. Mach. Tools Manufact.*, Vol. 37, No. 9 (1997), pp. 1219-1997.

Dowlatshahi, S. and M. Nagaraj, *Computers & IE*, 34(1), 1998, pp. 235-255.

Fayyad, U. M., G. Piatetsky-Shapiro, and P. Smyth. In: *Advances in Knowledge Discovery and Data Mining*, U.M. Fayyad, *et al.* (eds.), AAAI Press, 1996, pp. 1-34.

Keheo, A. and G. A. Parker, *NDT&E Int.*, 25(1), 1992, pp. 23-36.

Kulp, D., D. Haussler, and F. Eeckman. *Proc. of Conf. on Intelligent Systems in Molecular Biology*, AAAI Press, 1996.

Liao, T. W. and Y. Li, *NDT&E Int.*, 31(3), 1998, pp. 183-192.

Mitchell, J. S., *Introduction to Machinery Analysis and Monitoring*, Pennwell: Tulsa, OK, 1993.

Selim, H. M., R. G. Askin, and A. J. Vakharia, *Computers & IE*, 34(1), 1998, pp. 3-20.

Smith, S. D. G., R. Escobedo, M. Anderson, and T. P. Caudell, *IEEE Trans. on Neural Networks*, 8(4), 1997, pp. 847-851.

Stolorz, P., C. Dean, R. Crippen, and R. Bolm. In: *Concurrent Supercomputing Consortium Annual Report*, Ed. T. Pauna, 1995, pp. 20-22.

Weir, N., U.M. Fayyad, and S.G. Djorgovski. *The Astronomical Journal*, 109-6: 2401-2412.

Proceedings of the 25th International Conference on

Computers & Industrial Engineering

March 29-31, 1999
New Orleans, Louisiana

Mohamed I. Dessouky
Conference Chair
Department of Industrial Engineering
Northern Illinois University
DeKalb, IL, 60115-2854, USA

Sherif M. Waly
Conference Co-Chair
Industrial & Manufacturing Systems
Engineering Department
Louisiana State University
Baton Rouge, LA, 70810, USA

M. Sadek Eid
Conference Co-Chair
The Manufacturing Technology Center
Universite de Moncton
Moncton New Brunswick
Canada EIA 3E9

Hosted and Organized
By
Department of Industrial and Manufacturing Systems Engineering
Louisiana State University