J. A. Harding
e-mail: j.a.harding@lboro.ac.uk

M. Shahbaz
e-mail: m.shahbaz@lboro.ac.uk

Srinivas
e-mail: s.srinivas@lboro.ac.uk

Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire LE2 4LA, UK

A. Kusiak
Department of Mechanical and Industrial Engineering, The University of Iowa, Iowa City, IA 52242-1527
e-mail: andrew-kusiak@uiowa.edu

Data Mining in Manufacturing: A Review

The paper reviews applications of data mining in manufacturing engineering, in particular production processes, operations, fault detection, maintenance, decision support, and product quality improvement. Customer relationship management, information integration aspects, and standardization are also briefly discussed. This review is focused on demonstrating the relevancy of data mining to manufacturing industry, rather than discussing the data mining domain in general. The volume of general data mining literature makes it difficult to gain a precise view of a target area such as manufacturing engineering, which has its own particular needs and requirements for mining applications. This review reveals progressive applications in addition to existing gaps and less considered areas such as manufacturing planning and shop floor control. [DOI: 10.1115/1.2194554]

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. Knowledge exists in all business functions, including purchasing, marketing, design, production, maintenance and distribution, but knowledge can be notoriously difficult to identify, capture, and manage. Knowledge can be as simple as knowing who is best to contact if particular materials are running short, or can be as complex as mathematical formulas relating process variables to finished product dimensions. Spiegler [1] reviewed two models of knowledge. The first model follows a conventional hierarchy and transformation of data into information and knowledge with a spiral and recursive way of generating knowledge. The second model presents a reverse hierarchy where knowledge may appear before data and information processing. Knowledge discovery, knowledge management, and knowledge engineering are currently topics of importance to manufacturing researchers and managers intent on exploiting current assets. Database technology is central to all these knowledge-based research topics.

The use of databases and statistical techniques are well established in engineering [2]. The first applications of artificial intelligence in engineering in general and in manufacturing in particular were developed in the late 1980s [3,4]. The scope of these activities, however, has recently changed. The advancements in information technology (IT), data acquisition systems, and storage technology as well as the developments in machine learning tools have enticed researchers to move forward toward discovering knowledge from databases (KDD). Data from almost all the processes of the organization such as product and process design, material planning and control, assembly, scheduling, maintenance, recycling, etc., are recorded. These data stores therefore offer enormous potential as sources of new knowledge. Making use of the collected data is becoming an issue and data mining is a natural solution for transforming the data into useful knowledge. The extracted knowledge can be used to model, classify, and make predictions for numerous applications.

Contributed by the Manufacturing Engineering Division of ASME for publication in the Journal of Manufacturing Science and Engineering. Manuscript received April 4, 2005; final manuscript received December 9, 2005. Review conducted by C. J. Li.

The idea of finding patterns in manufacturing, design, business, or medical data is not new. Databases have been processed to derive the underlying relationships within the data for many years as evidenced by the developments in statistics. Traditionally, it was the responsibility of analysts, who generally used statistical techniques, but increasingly data mining, which is an emerging area of computational intelligence, is providing new systems, techniques, and theories for the discovery of hidden knowledge in large volumes of data. Data mining is a blend of concepts and algorithms from machine learning, statistics, artificial intelligence, and data management. With the emergence of data mining, researchers and practitioners began applying this technology on data from different areas such as banking, finance, retail, marketing, insurance, fraud detection, science, engineering, etc., to discover any hidden relationships or patterns. Data mining is therefore a rapidly expanding field with growing interests and importance and manufacturing is an application area where it can provide significant competitive advantage.

The use of data mining techniques in manufacturing began in the 1990s [5–7] 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. Since the importance of data mining in manufacturing has clearly increased over the last 20 years, it is now appropriate to critically review its history and application.

This paper presents a comprehensive overview of data mining applications in manufacturing, especially in the areas of production processes, control, maintenance, customer relationship management (CRM), decision support systems (DSS), quality improvement, fault detection, and engineering design. The remainder of the paper briefly describes pertinent manufacturing enterprise applications where data mining is applied to extract knowledge for improvement. The paper also approaches the challenging area

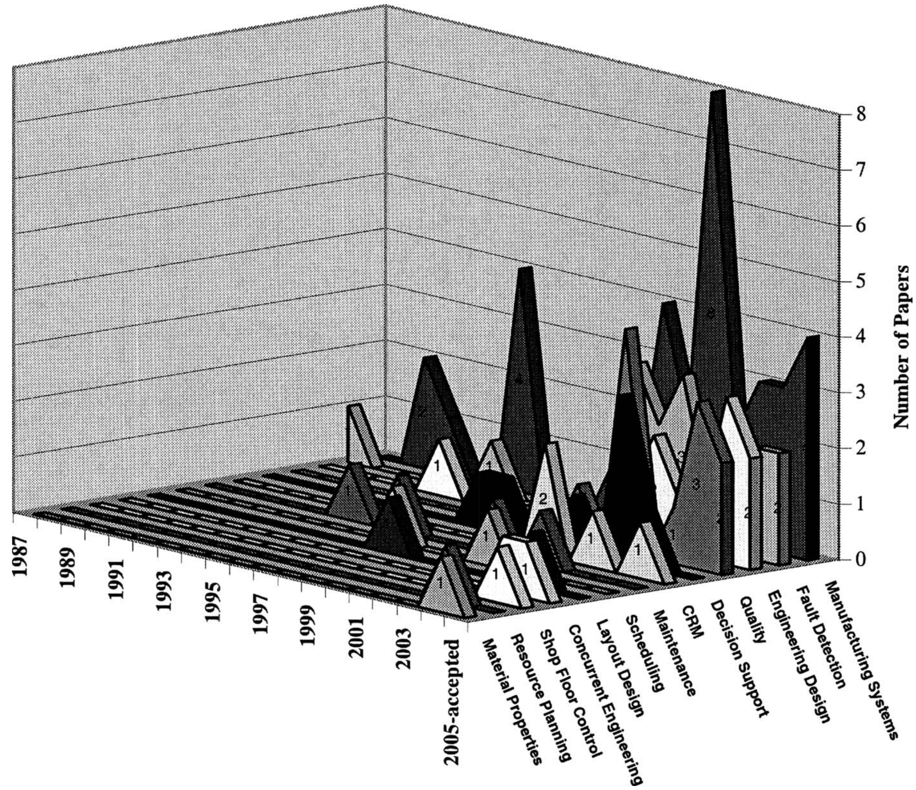

Fig. 1 History of manufacturing applications of data mining

of data mining system integration. Finally, the conclusions and future research directions outline the progress made by the ongoing research related to manufacturing control and quality improvement.

2 Data Mining Models for Manufacturing Applications

CRISP-DM™ (Cross Industry Standard Process for Data Mining), *SEMMA*, *SolEuNet* (Data Mining and Decision Support for Business Competitiveness: A European Virtual Enterprise), *Kensington Enterprise Data Mining* (Imperial College, Department of Computing, London, UK), and *Data Mining Group (DMG)* have established methodologies and developed languages and software tools for the standardization of industrial applications of data mining. However, most products focus on the implementation of data mining algorithms and application development rather than on the ease-of-use, integration, scalability, and portability. Most published research on data mining in manufacturing reports dedicated applications or systems, tackling specific problem areas, such as fault detection (see Fig. 1). Only limited research has been done to address the integration of data mining with existing manufacturing-based enterprise reference architectures, frameworks, middleware, and standards such as Common Object Request Broker Architecture (CORBA), Model-Driven Architecture (MDA), or Common Warehouse Metamodel (CWM). Neaga [8] explained the neglect of these issues and highlighted their importance and the investments already committed to the existing efforts for enterprise integration. Neaga and Harding [8,9] presented a holistic approach to a wide range of data mining applications suitable for manufacturing enterprises. The areas of manufacturing enterprise design, engineering and re-engineering, information modeling and the suitability of applying data mining techniques to use previous knowledge and information about an enterprise are examined in Refs. [8,9].

The CRISP-DM and SEMMA methodologies are most widely used by the data mining community. CRISP-DM and SEMMA provide a step by step guide for data mining implementation. CRISP-DM is easier to use than SEMMA as it provides detailed neutral guidelines that can be used by any novice in the data mining field. SEMMA is developed as a set of functional tools for SAS's Enterprise Miner software. Therefore those who use this specific software for their tasks are more likely to adopt this methodology. Using SEMMA, results may be found quickly by mining samples of data from the whole database, but if the discovered relationships do not follow in the whole database then new samples must be examined which means repeating the whole data mining process.

The details of each step of CRISP-DM [10] make it a reliable methodology that is easy to use and fast to implement. The detailed sub-stages are optional guidelines and can be skipped as required. The CRISP-DM methodology can therefore be fully or partially adopted depending on the problem and its requirements. This was the main reason that the CRISP-DM was used by the authors as a standard guide to implement the data mining research that is reported in Ref. [11].

Neaga and Harding also presented a framework for the integration of complex enterprise applications including data mining systems [12,13]. The presented approaches provide the definition and development of a common knowledge enterprise model, which represents a combination of previous projects on manufacturing enterprise architectures and Object Management Group (OMG) models and standards related to data mining.

3 Data Mining Applications Relevant to Manufacturing

This section details the contributions of researchers and practitioners in different areas of manufacturing from the late 1980s to date. The literature was searched extensively in different journals,

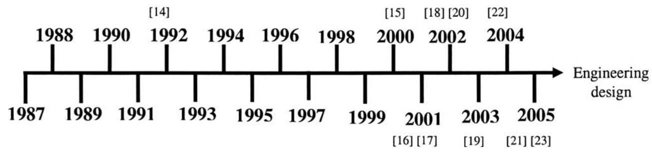

Fig. 2 Engineering design literature over time

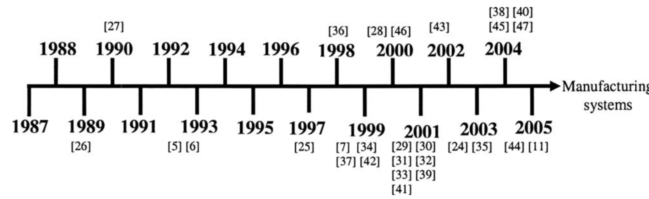

Fig. 3 Manufacturing systems literature over time

personal web pages, internet and citeseers web sites.[1] This review is particularly focused on data-mining applications and case studies in manufacturing and closely related fields.

The temporal stacked area chart in Fig. 1 shows the data mining research reported in different application areas of manufacturing. It clearly indicates the current trends of industry toward applications of data mining and shows that particularly since the beginning of the new century people have started to focus on solving their problems using historical databases. Areas such as manufacturing operations, fault detection, design engineering, and decision support systems have gained the attention of the research community, although there is still enormous potential for research in these areas. Other areas like maintenance, layout design, resource planning, and shop floor control require even greater attention and further exploration.

In each of the following subsections, a time series progress figure has been provided for quick reference to the history of data mining development and implementation in the particular area.

3.1 Engineering Design. Engineering design is a multidisciplinary, multidimensional, and non-linear decision-making process where parameters, actions, and components are selected. This selection is often based on historical data, information, and knowledge. It is therefore a prime area for data mining applications and although as yet only a few papers have reported applications of data mining in engineering design (see Fig. 2), this has been an area of increased research interests in recent years. These recently published papers form an important part of this review paper due to the essential synergies between design and manufacturing. The importance of considering how a product should be manufactured during the design stage and the constraints imposed on design by particular manufacturing processes and technologies have been accepted for many years. There is indeed great potential for data mined knowledge to integrate manufacturing, product characteristics, and the engineering design processes.

Sim and Chan [14] developed a knowledge-based system for the selection of rolling bearings. They used heuristic knowledge supported by a manufacturer's catalogue to optimize design specifications by matching the temporal data of the new product against the knowledge base. Kusiak et al. [15] proposed a rough-set theory approach to predict product cost. Ishino and Jin [16] used data mining for knowledge acquisition in design from the data obtained through observing design activities using a CAD system. They developed a method called Extended Dynamic Programming to extract the knowledge. Romanowski and Nagi [17] proposed a design system which supports the feedback of data mined knowledge from the life cycle data to the initial stages of the design process. Giess et al. [18,19] 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. Hamburg [20] applied data mining techniques to support product development by analyzing global environment aspects, market situation, strategy, philosophy, and culture of the manufacturing and customer behavior. He utilized a decision-tree algorithm to mine and integrate the enterprise data in the product development. Romanowski and Nagi [21,22] applied a data-mining approach for forming generic bills of materials (GBOMS), entities that represent the different variants in a product family and facilitate the search for similar designs and the configuration of new variants. By combining data-mining approaches such as text and tree mining in a new tree union procedure that embodies the GBOM and design constraints in constrained XML, the technical difficulties associated with a GBOM are resolved.

Kim and Ding [23] presented a data mining aided optimal design method capable of finding a competitive design solution with a relatively low computation cost. They applied the method to facilitate the optimal design of fixture layout in a four-station SUV side panel assembly process. The literature reviewed in this section is summarized in Fig. 2.

3.2 Manufacturing Systems. Data collection in manufacturing is common but its use tends to be limited to rather few applications. Machine learning and computational intelligence tools provide excellent potential for better control of manufacturing systems (see Fig. 3), especially in complex manufacturing environments where detection of the causes of problems is difficult. Piatesky-Shapiro et al. [7] argued that the data mining industry is coming of age. However, this review of data mining in manufacturing shows that although there are several areas in manufacturing enterprises that have benefited from data-mining algorithms, there are still numerous areas that could benefit further [24]. In manufacturing environments the need and importance of data collection is ever present for statistical process control purposes. Lee [5] discussed and suggested several principles leading to a knowledge-based factory environment utilizing the data collected over several stages of the manufacturing-related processes. A comparative study of implicit and explicit methods to predict the non-linear behavior of the manufacturing process, using statistical and artificial intelligence tools, was discussed by Kim and Lee [25].

Semiconductor manufacturing is complex and faces several challenges relating to product quality, scheduling, work in process, cost reduction, and fault diagnosis. To overcome these problems several methods and systems have been developed, e.g., Rule-Based Decision Support Systems (RBDSS) [26], CAQ [27], Knowledge Acquisition from Response Surface Methodology (KARSM), and GID3 [6] or generalized ID3, a decision-tree algorithm for fault diagnostics and decision making have been developed and used. Gardener and Bieker [28] showed a substantial savings in the manufacture of semiconductors by applying decision-tree algorithms and neural networks to solve the yield problem in the wafer manufacture. Sebzalli and Wang [29] applied principal component analysis and fuzzy c-means clustering to a refinery catalytic process to identify operational spaces and develop operational strategies for the manufacture of desired products and to minimize the loss of product during system changeover. Four operational zones were discovered, with three for product grade and the fourth region giving high probability of producing off-specification product. Lee and Park [30] used self-organizing maps and Last and Kandel [31] applied information fuzzy networks for quality checks and extracted useful rules from their model to check the quality of the products. Kusiak [32]

[1]http://citeseer.ist.psu.edu/cs

proposed a rule-structuring algorithm that can handle data from different sources to extract rules, which is very helpful in semiconductor manufacturing. The algorithm formed relevant metastructures enhancing the utility of the extracted knowledge. Dabbas and Chen [33] proposed the consolidation and integration of data from different semiconductor manufacturing sources into one database to generate different factory performance reports. Their method can be further exploited to use data mining to extract information from these reports.

Different data mining tools for improvement in integrated circuit manufacturing were presented in Ref. [34]. Another successful application of a sophisticated data mining algorithm was reported by Fountain et al. [35]. They used the Naïve Bayes probabilistic model in their theoretic decision-making approach to optimize testing of dies (ICs in the wafer form) during a die-level functional test. Their results showed substantial savings in testing costs and hence reduced overall costs compared with other testing policies, such as "exhaustive," "package all," and "Oracle."

An interesting area of research in manufacturing enterprises has been determination of optimal machining parameters to minimize machining errors such as tool wear, tool breakage, and tool deflection, which could result in slower production rates and increased costs. Park and Kim [36] reviewed different techniques based on CAD systems, operational research, and computational intelligence to determine the optimal solutions to these errors and for online adaptive control using knowledge-based expert systems. Other knowledge-based systems have also been proposed in Ref. [37] for condition interpretation of tools and quality of the products.

Performance and quality issues have also been considered while applying data mining techniques in manufacturing process related areas. Gertosio and Dussauchoy [38] have used linear regression analysis to determine and establish the relationships between test parameters and the performance of truck engines. Their simple methodology showed up to 25% saving in test process time. A method to reduce the component testing time required before assembly was proposed by Yin et al. [39]. They applied genetic and rough-set algorithms on past test data to find the optimal test criteria to substantially reduce the overall testing time. Another successful application of a regression model is presented in Ref. [40] to predict the performance of the knurling process and the quality of the knurls. Similar results were achieved by using both regression and neural networks.

Efforts have also been made to develop models to study the entire factory or enterprise data altogether to discover the problem areas instantly affecting any subsequent processes. Maki et al. [41,42] developed an intelligent system in Hitachi for online data analysis using a data mining approach. Their system used a rule induction algorithm to extract rules using an automated data mining engine and delivered the results using an intranet for easy access. Adams [43] analyzed different software that can be used to mine a factory's data and compare the features of information sharing. Shahbaz's [24] integrated data mining model is the next level of knowledge sharing, as once the data are mined the relevant data and data mining results can be shared within the factory and beyond at other sites, by using a neutral data format. Shahbaz et al. [44,45] used association rules for product design improvement and applied supervised association rules for controlling the product dimensions by controlling the process variables using supervised association rules [11,45]. Their methodology can be used as an alternative and/or a support to the design of experiments methodology.

Finally, the last two papers to be included and reported in this section are in the area of material properties. Chen et al. [46] applied data mining in hyperspace to identify material properties. They used MasterMiner to build a hyperspace data mining model, which uses *n* principle factors or most relevant variables and built a mathematical model to find the solution equation in *n* dimensional space for a specific material property. This technique is useful in chemical and material industry where different variables affect one or more properties of the material or chemical reaction. Interesting research has also been done by Mere et al. [47] to determine the optimal mechanical properties of galvanized steel by using a combination of clustering and neural networks. Clustering was used in the first instance and then neural networks were applied to the clusters to predict the mechanical properties of the steel.

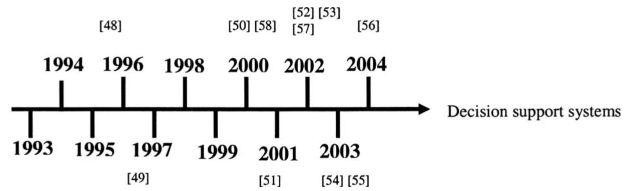

Fig. 4 Decision support systems literature over time

3.3 Decision Support Systems. Knowledge is the most valuable asset of an organization. Decisions are made based on a combination of judgement and knowledge from various domains. Decision support, knowledge management, and processing are interdependent activities in many organizations. Data mining applications related to Decision Support Systems are shown in Fig. 4. Ideally, all relevant knowledge should be available before making a decision. The knowledge extracted from databases (prescriptive data mining) can be integrated with existing expert systems. Grabot [48] used fuzzy logic to compliment the decision-support system to modify schedules. Koonce et al. [49] applied data mining to assist engineers in understanding the behavior of industrial data. They developed a software tool called DBMine using Bacons algorithm, decision trees, and DB learn. They applied the tool to find patterns in job shop scheduling sequences generated by a genetic algorithm [50]. Caskey [51] developed a general environment for providing the right knowledge at the right time. He used GAs and neural networks in identifying the structure of the data. The knowledge extracted was in the form of "actual control applied → performance obtained" and the knowledge generated could be used to increase the accuracy of the system or validate the performance model. Kusiak [52] applied data mining to support decision-making processes. Different data-mining algorithms were used to generate rules for a manufacturing system. A subset of these rules was then selected to produce a control signature of the manufacturing process. The control signature is a set of feature values or ranges that lead toward an expected output. Kusiak [53] used rough-set theory to determine the association between control parameters and the product quality in the form of decision rules and generated the control signature from those rules. Lee and Park [54] presented an agent-based customer centric electronic commerce model in a make-to-order semiconductor manufacturing environment. They used data mining for a decision support system by providing a set of recommendations reflecting domain knowledge. Knowledge-based systems can be used to enhance the application range of simulation. They offer the necessary knowledge required to make decisions in scheduling and rescheduling of manufacturing operations. Symeonidis et al. [55] applied data mining to make the ERP system more versatile and adaptive by integrating the knowledge extracted in companies' selling policies. Huang [56] presented an agent-based system forknowledge management focused on the decision support of modular and collaborative product design and manufacture. He used a neural network for the development of a decision support system.

Bolloju et al. [57] suggested integrating decision support systems and knowledge management processes across organizations using OLAP (On Line Analytical Processing). Based on this approach, a general framework was proposed for an enterprise decision support system using model marts and model warehouses for structured repositories of knowledge obtained through various

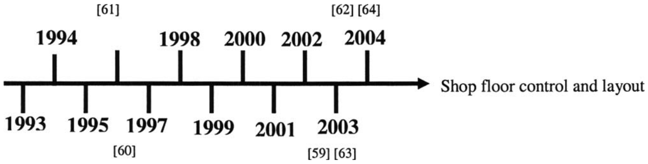

Fig. 5 Shop floor control and layout literature over time

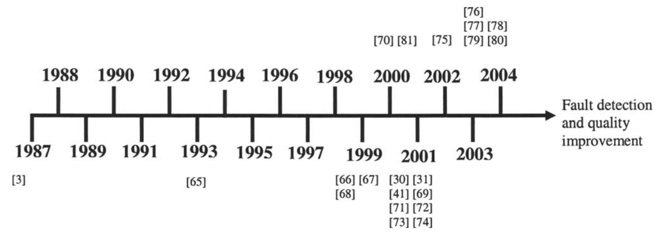

Fig. 6 Fault detection and quality improvement literature over time

sources. They assume that decision makers combine different types of data (e.g., internal data and external data) and knowledge (both tacit and explicit knowledge) available in various forms in the organization.

An interesting combination of neural networks and OLAP, called Neural On-Line Analytical Processing System (NOLAPS), was developed to enhance the decision support functionality of a network of enterprises. NOLAPS used a neural network for extrapolating probable outcomes based on available patterns of events and OLAP for converting complex data into new information and knowledge. The areas addressed are the selection of business partners, coordination in the distribution of production processes, and the prediction of production problems. The adoption of NOLAPS in real industrial situations was also suggested [58].

3.4 Shop Floor Control and Layout. The shop floor control and layout problems are concerned with the efficient and effective utilization of resources, at the lowest level of control in manufacturing. A vast amount of data is recorded during the operation of a shop floor, often to ensure that parts and production steps can be traced. These data can also be used to optimize the process itself, since the knowledge generated from mining historical work-in-process data helps in characterizing process uncertainty and parameter estimation of the system concerned. Data Mining literature related to these topics is shown in Fig. 5.

Chen [59] used association rules for cell-formation problems. Associations among the machines are found from the process database, which leads to the identification of the occurrences of other machines with the occurrence of a machine in the cell. This approach also clusters the parts and machines into families and cells simultaneously and hence requires minimal manual judgement. Chao et al. [60] presented an intelligent system to generate associative data for input in layout generation tools. They used an expert system, object oriented database, and cluster analysis, which ensures data consistency and determines the strength of relationship between the two items under consideration.

Knowledge generated from data mining can be used to analyze the effect of decisions made at any stage. Belz and Mertens [61] used SIMULEX coupled with a knowledge based system to model the plant and evaluate the results of various rescheduling measures. They used MANOVA for statistical analysis. The collected data can be analyzed to identify the normal and abnormal patterns in it. Kwak and Yin [62] presented a data-mining based production-control system for testing and rework in dynamic CIM. Their system analyzes the present situation and suggests dispatching rules to be followed and also how data mining can be used to evaluate the effect of those decisions. The knowledge generated can be used as the intelligence in a multi-agent system, Mitkas et al. [63] presented a multi-agent system for concurrent engineering equipped with a data-driven inference engine. The behavior and intelligence of each agent in the system is obtained by performing data mining on available application data and the respective knowledge domain. Srinivas et al. [64] presented a multi-agent based control architecture which uses data mining for decision support systems.

3.5 Fault Detection and Quality Improvement. Fault diagnosis is an area that has seen some of the earliest applications of data mining, e.g. Ref. [3], (see Fig. 6). A common and intuitive approach to problem solving is to examine what has happened in the past to better understand the process, then predict and improve the future system performance. Hence, the error rates in manufacturing are commonly used for knowledge acquisition to assist the quality control engineers. Data mining can help in identifying the patterns that lead toward potential failure of manufacturing equipment. This methodology helps in identifying not only the defective products but can also simultaneously determine the significant factors that influence the success or failure of the process. The knowledge thus generated by searching large databases can be integrated with the existing knowledge-based systems to enhance process performance and product improvement.

Data mining can be used to improve quality control, for example Apté et al. [65] used computational techniques for quality control in manufacturing. They deployed it in a disk-drive manufacturing line to reduce the number of expensive tests while meeting the performance criteria. They applied rule induction, neural network, decision tree, and *k*-nearest neighbor in their experimentation. Lee and Park [30] applied self-organizing maps to determine the optimal areas of inspection for a manufactured wafer. This technique can save considerable time that is used in carrying out a 100% inspection of the semiconductor wafer.

An important aspect of quality improvement is accurate fault diagnosis, and determining types of fault and failures. Malkoff [3] introduced a methodology which uses temporal data in performing fault diagnosis in a subsystem of a Navy Ship propulsion system. The methodology used patterns of the binary tree to generate the corrective actions. Liao et al. [66] presented fuzzy clustering based techniques for the detection of welding flaws. They also presented a comparison between two fuzzy clustering methods, i.e., fuzzy *k*-nearest neighbors and fuzzy c-means. Liao et al. [67,68] presented an integrated database and expert system for assisting the human analyst in identifying the failure mechanism of mechanical components. Liao et al. [69] discussed a multi-layer perceptron neural network to model radiographic welding data. Last and Kandel [31] used information fuzzy networks to build a prediction model for quality checks and then used this model for the extraction of rules. Shen et al. [70] applied rough-set theory to diagnose more than one category of faults in a generic manner, since it was used to extract the rules leading to the failure. These rules were used to distinguish the fault type or to inspect the dynamic characteristics of the machinery. They demonstrated their approach for the identification of valve faults in a multi-cylinder diesel engine. Lu [71] mined enterprise data to improve the quality of the product. They decreased the dimension of the data and then applied different data mining tools for quality improvement.

Maki and Teranishi [41] developed an automated data mining approach for data analysis in manufacturing and used it on an LCD production line. Their system consisted of three main features. First, it defined the data feeding and mining that were automated concurrently with the production process. They used an induction method for mining and also determined its statistical significance. Second, their system had the facility to store the generated rules in the intranet of the company and the third feature of their system was that it could also predict the temporal variance in the process. Zhou et al. [72] applied the C4.5 algorithm for drop test analysis of electronic goods. Kusiak and

Kurasek [73] used data mining to solve the quality engineering problems (solder ball defects) in the manufacture of printed circuit boards (PCB). They applied rough set theory to determine the causes of defects which needed further investigation. Oh et al. [74] presented an intelligent control system using a data mining architecture for quality improvement in the process industry. They used a neural network modeling method to establish the relationships between process and quality variables and identified the main causes of defects, which also provided optimized parameter adjustments. Skormin et al. [75] applied data mining for accurate assessment and forecasting of the probability of failure of hardware, such as avionics based on the historical data of environmental and operational conditions. They developed a heuristic for the determination of informative subspace in low dimension and then used a decision tree to model the data. Chen et al. [76] generated association rules for defect detection in semiconductor manufacturing. They determined the association between different machines and their combination with defects to determine the defective machine. They used the Piatetsky-Shapiro formula to determine the statistical significance of the identified association. Tseng et al. [77,78] used rough set theory to resolve quality control problems in PCB manufacturing by identifying the features that produce solder ball defect and also determined the features that significantly affect the quality of the product. Tseng et al. [79] used rough-set theory on machining data to identify the relationships between the features of the machining process and surface roughness. Shi et al. [80] applied a neural network to model nonlinear cause and effect relationships and applied it in the chemical and PCB manufacturing process.

Another interesting work is reported by Yuan et al. [81] in determining the toxicity (Microtox) in the process effluents from a chemical plant using neural networks and principal component analysis. Their software analyzer predicts the toxicity level and helps in developing strategies in process operations for toxicity reduction in the effluents.

3.6 Data Mining in Maintenance. Preventive maintenance is of key importance in process and manufacturing engineering. Databases containing the events of failure of the machines and the behavior of the relevant equipment at the time of the failure can be used in the design of the maintenance management systems. Batanov et al. [82] researched knowledge-based maintenance systems and developed a prototype system called EXPERT-MM, which works on historical failure data and provides suggestions for an appropriate preventive maintenance schedule. A data-based design of optimal maintenance methods has also been proposed by Hsu and Kuo [83]. They started working on this project at the beginning of the 1990s and they suggested that 100% of the inspection should start after the manufacture of a certain number (n) of parts and when the percentage of bad parts reaches a certain threshold value. Preventive maintenance should then start to bring the process under control again. When the process has been controlled and an additional n parts have been manufactured, the procedure can be repeated. Maintenance operations and quality control are interrelated and quality control databases can therefore be used to design preventive maintenance plans. Sylvain et al. [84] used different data mining techniques, including decision trees, rough sets, regression, and neural networks to predict component failure based on the data collected from the sensors of an aircraft. Their results also led to the design of preventive maintenance policies before the failure of any component. Romanowski and Nagi [85] applied data mining in a maintenance domain to identify which subsystems were responsible for low equipment availability. They recommended a preventive schedule and found that sensors and frequency response provide the most information about faults. They used a decision tree to model the data.

Although only a few reports of data mining have been identified in maintenance applications, this was one of the first areas of manufacturing to take advantage of data mining based solutions (see Fig. 7).

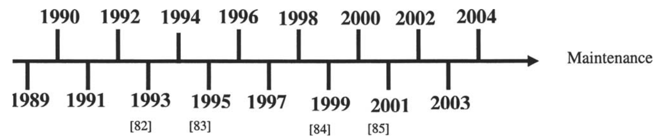

Fig. 7 Maintenance literature over time

3.7 Customer Relationship Management. The marketing model has shifted from product-focused to customer-focused. Customer Relationship Management (CRM) is concerned with increasing the value of interaction with customers and maximizing the profit. In this competitive and global business environment, the application of data mining in CRM related to manufacturing industry has attracted research interest (see Fig. 8).

CRM is as important as producing high quality and low cost products and is complementary to demand management which may be defined as a set of practices aimed at managing and coordinating a demand chain, starting from the end customer and working backwards to raw material and suppliers. To collect appropriate information, customer demand data are collected and analyzed and then the product design features are changed to meet the customer's demands. Similarly, in service industries, data from customers is the only source of knowledge about their satisfaction with the product. Morita et al. [86] applied data mining for customer segmentation to determine which customers are likely to shift from one cellular company to another. They used a rule induction algorithm on the transformed data to build rules and then predict the potential moves of the customers. Hui and Jha [87] used DBMiner to develop a decision-support system using a customer service database and they used neural networks and case based reasoning to mine the unstructured customer data to identify the machine faults. Rygielski et al. [88] discussed different data-mining techniques and provided an overview of customer relationship management. They presented two case studies, one using neural networks and the other using Chi-Square Automation Interaction Detection (CHAID) to improve the business by targeting customer's data. They reviewed both the models, comparing the simplicity of implementation of CHAID against the accuracy of neural networks. Agard and Kusiak [89] applied data mining to customer response data for its utilization in the design of product families. They used clustering for customer segmentation, i.e. to group the customers. The requirements from the product were then analyzed using association rules for the design of the product. Padmanabhan and Tuzhilin [90] described different ways in which optimization and data mining can help another for certain customer relationship management applications in e-commerce.

4 Future Directions and Conclusions

This paper has surveyed numerous applications of data mining in manufacturing. In recent years there has been a significant growth in the number of publications in some areas of manufacturing, such as fault detection, quality improvement, manufacturing systems, and engineering design. In contrast, other areas such as customer relationship management and shop floor control have received comparatively less attention from the data mining community. An exponential growth of data mining applications in the

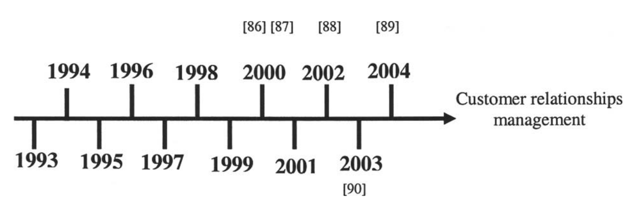

Fig. 8 Customer relationships management literature over time

semiconductor industry has been observed. The reasons for this may be that large volumes of data are generated during manufacture and that small improvements can have a significant impact in this industry. No other sector of manufacturing industry reports such large increases of data mining applications. This is rather surprising as other industries such as aerospace routinely collect huge quantities of data during product manufacture and hence are good potential environments for data mining studies.

Many reported applications are related to the causes of malfunctioning of different types of manufacturing systems or processes and hence the discovered knowledge leads toward the better functioning of the manufacturing enterprise. Developments in data mining are generally directed at the refinement of algorithms and their application in manufacturing, their integration with existing systems, standardization, the use of common methods and tools, and the definition of repeatable projects. Recent trends indicate an increasing awareness as more and more people are using data mining for problem solving in manufacturing. It is expected that future research will be directed at analyzing data related to design, shop floor control, scheduling, ERP, supply chain, and in developing a generic system where these can be integrated with existing knowledge based systems to enhance their capability.

The research reviewed in this paper has mainly concentrated on applications of the algorithms. The quality of the data and data preparation issues, particularly relating to manufacturing databases have not been discussed. Major effort is needed in the data preparation process, as this is often simply based on practitioner's instinct and experience. A more generic process for data cleaning is essential to enable the growth of data mining in manufacturing industry.

The manufacturing data-mining research often does not consider the quality of the rules or knowledge discovered. The knowledge generated is sometimes cumbersome and the relationships obtained are too complex to understand. Future research effort is therefore also needed to enhance the expressiveness of the knowledge.

The CRISP-DM methodology provides high level step-by-step instructions for applying data mining in engineering. Further research is needed to develop generic guidelines for a variety of different data and types of problems, which are commonly faced by manufacturing engineering industry.

References

[1] Spiegler, I., 2003, "Technology and Knowledge: Bridging a 'Generating' Gap," Inf. Manage., **40**, pp. 533–539.

[2] Han, J., and Kamber, M., 2001, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, New York, 550 pp.

[3] Malkoff, D. B., 1987, "A Framwork for Real-Time Fault Detection and Diagnosis Using Temporal Data," Artif. Intell. Eng., **2**(2), pp. 97–111.

[4] Ramamoorthy, C. V., and Wah, B. W., 1989, "Knowledge and Data Engineering," IEEE Trans. Knowl. Data Eng., **1**(1), pp. 9–16.

[5] Lee, M. H., 1993, "Knowledge Based Factory," Artif. Intell. Eng., **8**, pp. 109–125.

[6] Irani, K. B., Cheng, J., Fayyad, U. M., and Qian, Z., 1993, "Applying Machine Learning to Semiconductor Manufacturing," IEEE Expert, **8**(1), pp. 41–47.

[7] Piatetsky-Shapiro, G., 1999, "The Data Mining Industry Coming of Age," IEEE Intell. Syst., **14**(6), pp. 32–34.

[8] Neaga, E. I., and Harding, J. A., 2002, "A Review of Data Mining Techniques and Software Systems to Improve Business Performance in Extended Manufacturing Enterprises," Int. J. Adv. Manuf. Syst. (IJMAS), Spec Issue Decis. Eng., **5**(1), pp. 3–19.

[9] Neaga, E. I., and Harding, J. A., 2001, "Data Mining Techniques for Supporting Manufacturing Enterprise Design," *International Conference on Industrial and Production Management*, Quebec City, Canada, pp. 232–241.

[10] SPSS, 1999, *CRISP-DM 1.0 Step-By-Step Data Mining Guide*.

[11] Shahbaz, M., 2005, "Product and Manufacturing Process Improvement Using Data Mining," Ph.D. thesis, Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire, 194 pp.

[12] Neaga, E. I., and Harding, J. A., 2005, "An Enterprise Modelling and Integration Framework Based on Knowledge Discovery and Data Mining," Int. J. Prod. Res., **43**(6), pp. 1089–1108.

[13] Neaga, E. I., 2003, "Framework for Distributed Knowledge Discovery Systems Embedded in Extended Enterprise," Ph.D. thesis, Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, UK, 192 pp.

[14] Sim, S. K., and Chan, Y. W., 1992, "A Knowledge-Based Expert System for Rolling-Element Bearing Selection in Mechanical Engineering Design," Artif. Intell. Eng., **6**(3), pp. 125–135.

[15] Kusiak, A., Kernstine, K. H., Kern, J. A., McLaughlin, K. A., and Tseng, T. L., 2000, "Data Mining: Medical and Engineering Case Studies," *Industrial Engineering Research Conference*, Cleveland, OH, pp. 1–7.

[16] Ishino, Y., and Jin, Y., 2001, "Data Mining and Knowledge Acquisition in Engineering Design," *Data Mining for Design and Manufacturing: Methods and Applications*, D. Braha, ed., Kluwer Academic, Dordrecht, pp. 145–160.

[17] Romanowski, C. J., and Nagi, R., 2001, "A Data Mining for Knowledge Acquisition in Engineering Design," *Data Mining for Design and Manufacture: Methods and Applications*, D. Braha, ed., Kluwer Academic, Dordrecht, pp. 161–178.

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

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

[20] Hamburg, I., 2002, "Improving Computer Supported Environment Friendly Product Development by Analysis of Data," *2nd European Conference on Intelligent Systems and Technologies*, Iasi, Romania, 6 pp.

[21] Romanowski, C. J., and Nagi, R., 2005, "On Comparing Bills of Materials: A Similarity/Distance Measure for Unordered Trees," IEEE Trans. Syst. Man Cybern., Part A. Syst. Humans, **35**(2), pp. 249–260.

[22] Romanowski, C. J., and Nagi, R., 2004, "A Data Mining Approach to Forming Generic Bills of Material in Support of Variant Design Activities," ASME J. Comput. Inf. Sci. Eng., **4**(4), pp. 316–328.

[23] Kim, P., and Ding, Y., 2005, "Optimal Engineering System Design Guided by Data-Mining Methods," Technometrics, **47**(3), pp. 336–348.

[24] Shahbaz, M., and Harding, J. A., 2003, "An Integrated Data Mining Model for Manufacturing Enterprises," *International Conference on Manufacturing Research*, Glasgow, pp. 539–545.

[25] Kim, S. H., and Lee, C. M., 1997, "Non Linear Prediction of Manufacturing Systems Through Explicit and Implicit Data Mining," Comput. Ind. Eng., **33**(3-4), pp. 461–464.

[26] Adachi, T., Talavage, J. J., and Modie, C. L., 1989, "A Rule Based Control Method for a Multi Loop Production System," Artif. Intell. Eng., **4**(3), pp. 115–125.

[27] Whitehall, B. L., Lu, S. C. Y., and Stepp, R. E., 1990, "CAQ: A Machine Learning Tool for Engineering," Artif. Intell. Eng., **5**(4), pp. 189–198.

[28] Gardner, M., and Bieker, J., 2000, "Data Mining Solves Tough Semi Conductor Problems," *KDD 2000*, Boston, pp. 376–383.

[29] Sebzalli, Y. M., and Wang, X. Z., 2001, "Knowledge Discovery From Process Operational Data Using PCA and Fuzzy Clustering," Eng. Applic. Artif. Intell., **14**, pp. 607–616.

[30] Lee, J. H., and Park, S. C., 2001, "Data Mining for High Quality and Quick Response Manufacturing," *Data Mining for Design and Manufacturing: Methods and Applications*, D. Braha, ed., Kluwer Academic, pp. 179–205.

[31] Last, M., and Kandel, A., 2001, "Data Mining for Process and Quality Control in the Semiconductor Industry," *Data Mining for Design and Manufacturing: Methods and Applications*, D. Braha, ed., Kluwer Academic, pp. 207–234.

[32] Kusiak, A., 2001, "Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing," IEEE Trans. Electron. Packag. Manuf., **24**(1), pp. 44–50.

[33] Dabbas, R. M., and Chen, H. N., 2001, "Mining Semiconductor Manufacturing Data for Productivity Improvement—An Integrated Relational Database Approach," Comput Ind., **45**, pp. 29–44.

[34] McDonald, C. J., 1999, "New Tools for Yield Improvement in Integrated Circuit Manufacturing: Can They be Applied to Reliability?," Microelectron. Reliab., **39**(6-7), pp. 731–739.

[35] Fountain, T., Dietterich, T., and Sudyka, B., 2003, "Data Mining for Manufacturing Control: An Application in Optimizing IC Test," *Exploring Artificial Intelligence in the New Millennium*, G. Lakemeyer and B. Nebel, eds., Morgan Kaufmann, San Francisco, CA, pp. 381–400.

[36] Park, K. S., and Kim, S. H., 1998, "Artificial Intelligence Approaches to Determination of CNC Machining Parameters in Manufacturing: A Review," Artif. Intell. Eng., **12**, pp. 127–134.

[37] Liao, T. W., Chen, J. H., and Triantaphyllou, E., 1999, "Data Mining Applications in Industrial Engineering: A Perspective," *Proceedings of the 25th International Conference on Computers and Industrial Engineering*, New Orleans, LA, pp. 265–276.

[38] Gertosio, C., and Dussauchoy, A., 2004, "Knowledge Discovery From Industrial databases," J. Intell. Manuf., **15**, pp. 29–37.

[39] Yin, Z. L., Pheng, K. L., and Cheong, F. S., 2001, "Derivation of Decision Rules for the Evaluation of Product Performance Using Generic Algorithms and Rough Set Theory," *Data Mining for Design and Manufacturing: Methods and Applications*, D. Braha, ed., Kluwer Academic, Dordrecht, pp. 337–353.

[40] Feng, C. X. J., and Wang, X. F., 2004, "Data Mining Techniques Applied to Predictive Modelling of the Knurling Process," IIE Trans., **36**, pp. 253–263.

[41] Maki, H., and Teranishi, Y., 2001, "Development of Automated Data Mining System for Quality Control in Manufacturing," *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 93–100.

[42] Maki, H., Maeda, A., Morita, T., and Akimori, H., 1999, "Applying Data Mining to Data Analysis in Manufacturing," *International Conference on Advances in Production Management Systems*, pp. 324–331.

[43] Adams, L., 2002, "Mining Factory Data," Quality Magazine, Software and Analysis, May, pp. 5.

[44] Shahbaz, M., Srinivas, Harding, J. A., and Turner, M., "Product Design and

Manufacturing Process Improvement Using Association Rules," Proc. Inst. Mech. Eng., Part B (in press).
[45] Shahbaz, M., Srinivas, and Harding, J. A., 2004, "Knowledge Extraction from Manufacturing Process and Product Databases using Association Rules," *PDT Europe*, Stockholm, Sweden, pp. 145–153.
[46] Chen, N., Zhu, D. D., and Wang, W., 2000, "Intelligent Material Processing by Hyper Space Data Mining," Eng. Applic. Artif. Intell., **13**, pp. 527–532.
[47] Mere, J. B. O., Marcos, A. G., Gonzalez, J. A., and Rubio, V. L., 2004, "Estimation of Mechanical Properties of Steel Strip in Hot Dip Galvanising Lines," Ironmaking Steelmaking, **31**(1), pp. 43–50.
[48] Grabot, B., Blanc, J. C., and Binda, C., 1996, "A Decision Support System for Production Activity Control," Decision Support Sys., **16**, pp. 87–101.
[49] Koonce, D. A., Fang, C. H., and Tsai, S. C., 1997, "A Data Mining Tool for Learning From Manufacturing Systems," Comput. Ind. Eng., **33**(1-2), pp. 27–30.
[50] Koonce, D. A., and Tsai, S. C., 2000, "Using Data Mining to Find Patterns in Genetic Algorithm Solutions to a Job Shop Schedule," Comput. Ind. Eng., **38**, pp. 361–371.
[51] Caskey, K. R., 2001, "A Manufacturing Problem Solving Environment Combining Evaluation, Search and Generalisation Methods," Comput Ind., **44**, pp. 175–187.
[52] Kusiak, A., 2002, "Data Mining and Decision Making," *SPIE Conference on Data Mining and Knowledge Discovery: Theory, Tools and Technology IV*, Orlando, FL, pp. 155–165.
[53] Kusiak, A., 2002, "A Data Mining Approach for Generation of Control Signatures," ASME J. Manuf. Sci. Eng., **124**, pp. 923–926.
[54] Lee, J. H., and Park, S. C., 2003, "Agent and Data Mining Based Decision Support System and its Adaptation to a New Customer-Centric Electronics Commerce," Expert Sys. Applic., **25**, pp. 619–635.
[55] Symeonidis, A. L., Kehagias, D. D., and Mitkas, P. A., 2003, "Intelligent Policy Recommendations on Enterprise Resource Planning by the Use of Agent Technology and Data Mining Techniques," Expert Sys. Applic., **25**, pp. 589–602.
[56] Huang, C. C., 2004, "A Multi-Agent Approach to Collaborative Design of Modular Products," Concurr. Eng. Res. Appl., **12**(1), pp. 39–47.
[57] Bolloju, N., Khalifa, M., and Turban, E., 2002, "Integrated Knowledge Management Into Enterprise Environment for the Next Generation Decision Support," Decision Support Sys., **33**, pp. 163–176.
[58] Lau, H. C. W., Chinb, K. S., Punb, K. F., and Ninga, A., 2000, "Decision Supporting Functionality in a Virtual Enterprise Network," Expert Sys. Applic., **19**, pp. 261–270.
[59] Chen, M. C., 2003, "Configuration of Cellular Manufacturing Systems Using Association Rule Induction," Int. J. Prod. Res., **41**(2), pp. 381–395.
[60] Chao, K. M., Guenov, M., Hills, B., Smith, P., Buxton, I., and Tsai, C. F., 1997, "An Expert System to Generate Associativity Data for Layout Design," Artif. Intell. Eng., **11**, pp. 191–196.
[61] Belz, R., and Mertens, P., 1996, "Combining Knowledge Based Systems and Simulation to Solve Rescheduling Problems," Decision Support Sys., **17**, pp. 141–157.
[62] Kwak, C., and Yih, Y., 2004, "Data Mining Approach to Production Control in the Computer Integrated Testing Cell," IEEE Trans. Rob. Autom., **20**(1), pp. 107–116.
[63] Mitkas, P. A., Symeonidis, A. L., Kehagias, D., and Athanasiadis, I., 2003, "Application of Data Mining and Intelligent Agent Technologies to Concurrent Engineering," *International Conference on Concurrent Engineering: Research and Applications*, Madeira, Portugal, pp. 11–18.
[64] Srinivas, Harding, J. A., and Shahbaz, M., 2004, "Agent Oriented Planning Using Data Mined Knowledge," *10th International Conference on Concurrent Engineering, Adaptive Engineering for Sustainable Value Creation*, Seville, Spain, pp. 301–307.
[65] Apte, C., Weiss, S., and Grout, G., 1993, "Predicting Defects in Disk Drive Manufacturing: A Case Study in High Dimensional Classification," *IEEE Annual Computer Science Conference on Artificial Intelligence in Application*, Los Alamitos, pp. 212–218.
[66] Liao, T. W., Li, D. M., and Li, Y. M., 1999, "Detection of Welding Flaws From Radiographic Images With Fuzzy Clustering Methods," Fuzzy Sets Syst., **108**, pp. 145–158.
[67] Liao, T. W., Khan, Z. H., and Mount, C. R., 1999, "An Integrated Database and Expert System for Failure Mechanism Identification: Part II—The System and Performance Testing," Eng. Failure Anal., **6**, pp. 407–421.
[68] Liao, T. W., Khan, Z. H., and Mount, C. R., 1999, "An Integrated Database and Expert System for Failure Mechanism Identification: Part I—Automated Knowledge Acquisition," Eng. Failure Anal., **6**, pp. 387–406.
[69] Liao, T. W., Wang, G., Triantaphyllou, E., and Chang, P. C., 2001, "A Data Mining Study of Weld Quality Models Constructed With MLP Neural Networks From Stratified Sample Data," *Industial Engineering Research Conference*, Dallas, TX, p. 6.
[70] Shen, L., Tay, F. E. H., Qu, L. S., and Shen, Y., 2000, "Fault Diagnosis Using Rough Set Theory," Comput Ind., **43**, pp. 61–72.
[71] Lu, J. C., 2001, "Methodology of Mining Massive Data Sets for Improving Manufacturing Quality/Efficiency," *Data Mining for Design and Manufacturing: Methods and Applications*, D. Braha, ed., Kluwer Academic, Dordrecht, pp. 255–288.
[72] Zhou, C., Nelson, P. C., Xiao, W., Tirpak, T. M., and Lane, S. A., 2001, "An Intelligent Data Mining System for Drop Test Analysis of Electronic Products," IEEE Trans. Electron. Packag. Manuf., **24**(3), pp. 222–231.
[73] Kusiak, A., and Kurasek, C., 2001, "Data Mining of Printed Circuit Board," IEEE Trans. Rob. Autom., **17**(2), pp. 191–196.
[74] Oh, S., Han, J., and Cho, H., 2001, "Intelligent Process Control System for Quality Improvement by Data Mining in the Process Industry," *Data Mining for Design and Manufacturing: Methods and Applications*, D. Braha, ed., Kluwer Academic, Dordrecht, pp. 289–310.
[75] Skormin, V. A., Gorodetski, V. I., and PopYack, I. J., 2002, "Data Mining Technology for Failure of Prognostic of Avionics," IEEE Trans. Aerosp. Electron. Syst., **38**(2), pp. 388–403.
[76] Chen, W. C., Tseng, S. S., and Wang, C. Y., 2004, "A Novel Manufacturing Defect Detection Method Using Data Mining Approach," Lecture Notes in Artificial Intelligence, **3029**, pp. 77–86.
[77] Tseng, T. L., JothiShankar, M. C., and Wu, T., 2004, "Quality Control Problems in Printed Circuit Board Manufacturing—An Extended Rough Set Theory Approach," J. Manuf. Syst., **23**(1), pp. 56–72.
[78] Tseng, T. L., JothiShankar, M. C., Wu, T., Xing, G., and Jiang, F., 2004, "Applying Data Mining Approaches for Defect Diagnosis in Manufacturing Industry," *IERC 2004*, Houston, TX, p. 7.
[79] Tseng, T. L., Leeper, T., Banda, C., Herren, S. M., and Ford, J., 2004, "Quality Assurance in Machining Process Using Data Mining," *IERC 2004*, Houston, TX, p. 6.
[80] Shi, X., and Boyd, P. S. D., 2004, "Applying Artificial Neural Network and Virtual Experimental Design to Quality Improvement of Two Industrial Processes," Int. J. Prod. Res., **42**(1), pp. 101–118.
[81] Yuan, B., Wang, X. Z., and Morris, T., 2000, "Software Analyser Design Using Data Mining Technology for Toxicity Prediction of Aqueous Effluents," Waste Manage., **20**, pp. 677–686.
[82] Batanov, D., Nagarur, N., and Nitikhumkasem, P., 1993, "Expert—MM: A Knowledge Based System for Maintenance Management," Artif. Intell. Eng., **8**, pp. 283–291.
[83] Hsu, L. F., and Kuo, S., 1995, "Design of Optimal Maintenance Policies Based on Online Sampling Plans," Eur. J. Oper. Res., **86**, pp. 345–357.
[84] Sylvain, L., Fazel, F., and Stan, M., 1999, "Data Mining to Predict Aircraft Component Replacement," IEEE Intell. Syst., **14**(6), pp. 59–65.
[85] Romanowski, C. J., and Nagi, R., 2001, "Analysing Maintenance Data Using Data Mining Methods," *Data Mining for Design and Manufacture: Methods and Applications*, D. Braha, ed., Kluwer Academic, Dordrecht, pp. 235–254.
[86] Morita, T., Sato, Y., Ayukawa, E., and Maeda, A., 2000, "Customer Relationship Management Through Data Mining," *Informs-Korms 2000*, Seoul, pp. 1956–1963.
[87] Hui, S. C., and Jha, G., 2000, "Data Mining for Customer Service Support," Inf. Manage., **38**, pp. 1–13.
[88] Rygielski, C., Wang, J. C., and Yen, D. C., 2002, "Data Mining Techniques for Customer Relationship Management," Technol. Soc., **24**, pp. 483–502.
[89] Agard, B., and Kusiak, A., 2004, "Data-Mining Based Methodology for the Design of Product Families," Int. J. Prod. Res., **42**(15), pp. 2955–2969.
[90] Padmanabhan, B., and Tuzhilin, A., 2003, "On the Use of Optimization for Data Mining: Theoretical Interactions and eCRM Opportunities," Manage. Sci., **49**(10), pp. 1327–1343.