Data Mining in Manufacturing Environments: Goals, Techniques and Applications

Alex G. Büchner, Sarabjot S. Anand and John G. Hughes
Northern Ireland Knowledge Engineering Laboratory
University of Ulster
Shore Road, Newtownabbey, Co. Antrim, BT37 0QB
Northern Ireland, UK
e-mails: {ag.buchner, ss.anand, jg.hughes}@ulst.ac.uk

Abstract: The paper describes the concepts of data mining and their synergy with manufacturing environments. A generic process is introduced, which outlines data mining goals and techniques, supported by example scenarios. Various applications of manufacturing environments are shown in which data mining has been applied to successfully, and potential areas in which the outlined mechanisms are capable of being applied.

Keywords: Data mining, knowledge discovery, data mining process, manufacturing.

Alex G. Büchner received his MSc in Software Engineering from the University of Abertay Dundee, Scotland in 1993. He is currently employed by the Northern Ireland Knowledge Engineering Laboratory at the University of Ulster, Northern Ireland. In his current position as research officer he is mainly working in the areas of data mining, heterogeneous databases, internationalisation, and object-orientation.

Sarabjot S. Anand received a BA in mathematics from Hindu College, University of Delhi and an MSc in engineering computation from Queen's University of Belfast. He is a research fellow at the Northern Ireland Knowledge Engineering Laboratory. His research interests include database technology, high-performance computing, case-based reasoning, data mining, and reasoning under uncertainty.

John G. Hughes earned a BSc in mathematics and statistics and a PhD in applied mathematics, both from Queen's University of Belfast. He is currently the dean of the Faculty of Informatics and director of the Northern Ireland Knowledge Engineering Laboratory. His research interests include object-oriented databases, artificial intelligence, and knowledge engineering. He is a fellow of the British Computer Society and a member of the IEEE Computer Society.

1 Introduction

Data mining has been defined as "the efficient, semi-automated discovery of non-trivial, implicit, previously unknown, potentially useful and understandable information from large data sets" [1]. Within the last decade data mining mechanisms have been applied in various industrial and organisational sectors, and have initiated a wide range of research activities. Analogously, IT techniques have been added to all steps of manufacturing processes, and more recently with artificial intelligence mechanisms to improve the quality and quantity of the yields. The introduction of data and knowledge driven technologies has led to a situation in which the amount of data supersedes the quality of knowledge. The objective of this article is to demonstrate the potential of data mining in overcoming this information gap within manufacturing processes.

The outline of the paper is as follows. Section 2 describes the synergy of data mining and manufacturing, based on objectives and abilities of data mining, as well as objectives and drawbacks of current manufacturing environments. In Section 3, a generic data mining process is outlined and examples are used to show the applicability of each step. Section 4 describes a battery of manufacturing scenarios in which data mining has been applied successfully, before outlining, in Section 5, conclusions and further work, i.e. potential areas in which the outlined data mining mechanisms are capable of being harnessed.

2 The Data Mining and Manufacturing Synergy

As defined earlier, the overall objective of data mining is to discover knowledge from data. This is achieved through combining the disciplines of machine learning and database theory, and supported by techniques from related areas, such as mathematics, statistics, visualisation, high performance computing etc. That is, data mining is not a new invention; it is a synergy of the resulting amalgamate.

To enable data mining, several technologies are supportive, many of which operate in industrial environments already. One of the most important enabling technology is data warehousing, which is a technique for integrating legacy operational systems within a corporation to provide an enterprise-wide view for decision support purposes. One of the techniques supported by data warehousing is on-line analytical processing (OLAP), which has been defined as "the dynamic synthesis, analysis and consolidation of large volumes of multi-dimensional data" [2]. A data warehouse provides many components which facilitate data mining, especially the time-intensive task of data pre-processing. Another enabling technology is that of report generators used to display the contents of discovered knowledge. Since data mining operations can become computationally very expensive, parallel technologies are becoming a more popular enabling technology.

In modern manufacturing environments vast amounts of data are being collected in database management systems and data warehouses from all involved areas, such as product and process design, assembly, materials planning and control, order entry and scheduling, maintenance, recycling, etc. Many knowledge-based components have also been added to (semi-)automate certain steps in that process. Examples are expert systems for decision support, intelligent scheduling systems for concurrent production, fuzzy controllers, etc.

A persistent problem is the gathering of the required expert knowledge to implement these knowledge-based components. Data mining provides some solutions to minimise this knowledge acquisition bottleneck problem in that it sifts through relevant data, which contains most of the required knowledge implicitly and discovers patterns to be incorporated in the manufacturing process. Thus, on a more abstract level, data mining can be seen as a supporting vehicle from product data management to product data and knowledge management.

3 The Data Mining Process

Data mining is recognised to be a process, rather than a stand alone automated algorithm that discovers knowledge from data without human intervention. While such a system would clearly be ideal, it is far from possible using present data mining techniques. In this section a generic data mining process is being described (see Figure 1).

The start of the data mining process is the identification of a problem requiring IT support for decision making. The process that follows is comprised of a number of components beginning with the identification of the human resources required to carry out the data mining process.

To give an idea of how those steps can be applied in reality, examples from manufacturing environments are given. The scenarios are chosen from a virtual manufacturing unit which assembles parts of a larger component and tries to identify patterns, which could lead to faulty yields.

3.1 Human Resource Identification

After a problem has been identified at the management level of an enterprise, human resource identification is the first stage of the data mining process. In most real-world data mining problems the human resources required are: the domain expert, the data expert and the data mining expert. Normally, data mining is carried out in large organisations where the prospect of finding a domain expert who is also an expert in the data stored by the organisation is rare. The synergy of these human resources as early as possible within any data mining project is imperative to its success.

For example, in a production plant, the domain expert would belong to an engineering unit, while the data expert would belong to the IT department. The data mining expert would normally belong to an organisation outside the factory for the purpose of achieving the data mining goal.

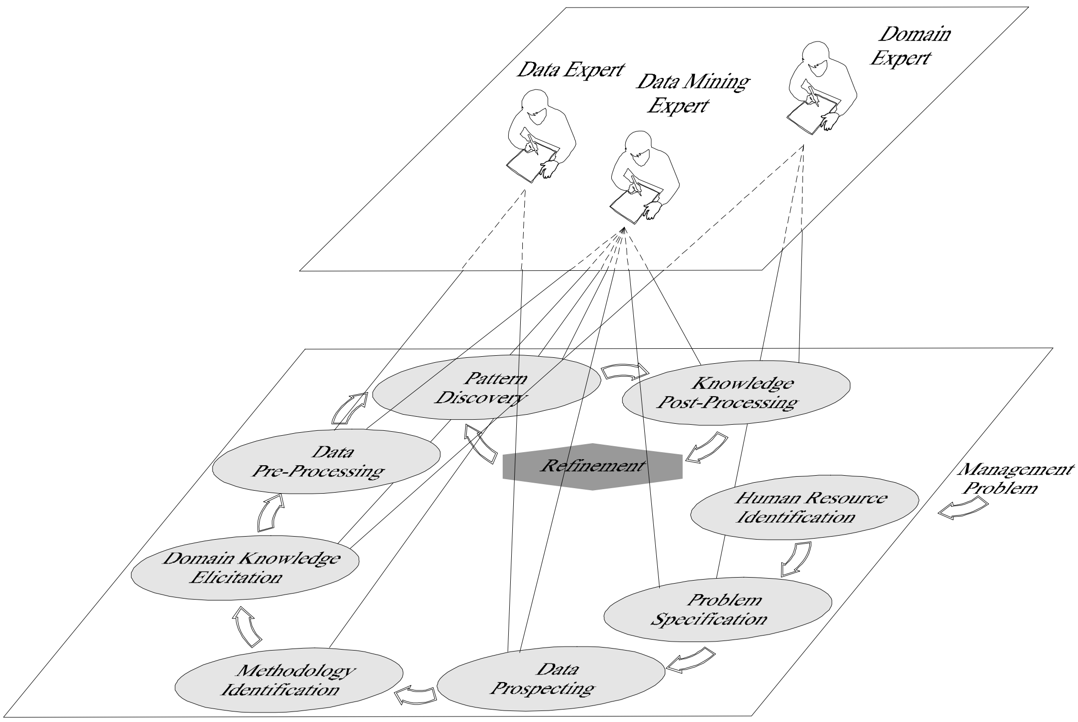

Figure 1. The Data Mining Process

3.2 Problem Specification

Problem specification is the second stage of the process. Here a better understanding of the problem is developed by the human resources identified in the previous stage of the process. The problem is decomposed into sub-problems and those tasks that can be solved using a data mining approach are identified. Each of these tasks is associated with a particular data mining goal.

There is a whole battery of goals which can be achieved through the application of data mining techniques, e.g., association discovery, classification, cluster analysis, sequential pattern discovery, temporal modelling, deviation detection, regression, characteristics discovery, or dependency modelling. In this article only the goals most relevant to manufacturing problems are outlined briefly, namely the first six goals given in the above list. For more detailed information about other goal types, refer to [1].

Discovery of associations involves the discovery of rules that associate one set of attributes in a data set to other attribute sets. An association rule is in the form $A \rightarrow B$, where A and B are conjunctions of expressions on attributes of the database. A is referred to as the antecedent and B, the consequent. For example, given a table containing information about triggered alarms, a rule which associates temperature and material consistency to the type of alarm could be

Temp > 250 $\wedge$ MatConsistency = Weak
$\rightarrow$ (Alarm = Overheat)

Classification rules are rules that discriminate between different partitions of a database based on various attributes within the database. The partitions of the database are based on an attribute called the classification label. Each value within the classification label domain is called a class. Consider a data source which contains information about faulty components, and the classification labels ‘faulty’ and ‘good’. An example rule could be

If Device = Laser $\wedge$ Temp > 50 $\wedge$ BeltNo = 12
$\rightarrow$ Output = faulty

Cluster analysis or data segmentation, often referred to in machine learning literature as unsupervised learning, is concerned with discovering structure in data. This goal is also known as learning by observation and discovery. Cluster analysis differs from classification in that the classes, to which the data tuples in the training set belong, are not provided. The clustering algorithm has to identify the classes by finding similarities between different states provided as examples. A classification algorithm may then be used to discover the distinguishing features of these discovered classes within the data. Thus, often cluster analysis forms a precursor to the use of classification algorithms within data mining. A typical scenario in which cluster analysis would be performed is the detection of groups in a 3-dimensional search space in data about materials with the given attributes density, melting temperature and boiling point. Possible clusters to be detected (and named by a domain expert) would be heavy metal, light metal, composites, and others. These clusters can then be used as labels for classification.

Sequential pattern discovery is similar to discovery of associations. The difference here is that sequential pattern discovery techniques discover associations across time. A sequential pattern is in the form $(A)_{t_i} \rightarrow (B)_{t_j}$, where A and B are conjunctions of expressions on attributes of the database and the attributes in A appear in the database with an earlier time stamp than B i.e. $t_i < t_j$. This type of pattern can be interesting in predicting consequences or causes of faults in a manufacturing process. For example, the occurrence of a fault caused by high temperature on belt number 2 can lead to a similar problem on the connected belt 3 within the next 10 minutes.

If Temp = High $\wedge$ BeltNo = 2
$\rightarrow$ BeltNo = 3 ([0..10] minutes)

Temporal modeling is concerned with the discovery of rules that are based on temporal data. Its major objective is to find frequencies and relationships in data among intervals. An example rule to be discovered from process data of a machine tool over a certain period of time is

If $VibrationAtBearing_1$ > 3000Hz ($\geq$ 6 hours) $\wedge$
$VibrationAtBearing_2$ > 3500Hz (> 4 hours)
$\rightarrow$ $LatheTool_1$ = broken ([1..3] Days)

A deviation is defined as the difference between an observed value and a reference value. Deviations are of a number of types [3]: deviation over time, normative deviation and deviation from expectation. These three types of deviations differ in the norm used to calculate the deviation of the observed value. In deviation over time, the norm would be based on the value of the variable over a certain time period in the past. For example, the yield results of last years 3rd quarter could be the norm against which this years 3rd quarter values are compared. When a standard norm is available as a reference value, deviation from that value is referred to as normative deviation. An example standard norm is the ISO norm for screw measures, against which measures taken at the quality assurance step are compared. In deviation from expectation, the expected value may be generated from a model or may be based on a hypothesis provided by a domain expert. The calculated density of a material to be produced can be used as expected value and form the basis from which deviations are detected.

The second part of the problem specification stage is to identify the ultimate user of the knowledge. If the discovered knowledge is to be used by a human (e.g., a plant engineer), it must be in a format that the user can understand and is familiar with. However, if data mining is only a small part of a larger project and the output from knowledge discovery is to be interpreted by a computerised system, (e.g., a CNC machine or a statistical package) the format of the discovered knowledge will have to strictly adhere to the expected format.

3.3 Data Prospecting

Data prospecting is the next stage in the process. It consists of analysing the state of the data required for solving the problem. There are four main considerations within this stage:

- Identification of relevant attributes
- Accessibility of data
- Population of required data attributes
- Distributed and heterogeneity of data

The relevance of attributes differs from problem to problem. While the measurements of a component might be indispensable information in solving one data mining problem, it might be unessential for another. One type of information which is usually irrelevant are primary keys, since they are unique by nature and thus do not contain any patterns. In order to avoid unjustified biases, it is important to include all data attributes that could be related to the problem and not just those that are relevant according to the domain expert.

The accessibility of data can be revoked for several reasons. Data might not be stored electronically, for example, manually kept logs from an assembly belt. Data might also not be accessible physically, which can be caused by lack of infrastructure, e.g., disjoint factory units, or because of security reasons.

The population of relevant attributes is crucial to the quality of the discovered knowledge. Although null values can have some semantic meaning, they usually worsen the quality of the data mining outcome. For instance, a null value in a control unit can either mean that a measurement has not been taken because it was not necessary, or because the apparatus was broken. One major reason for non-existing values in manufacturing environments is that quality assurance at the testing stage of components or products is only carried out on a small sample, rather than every single element.

If data is distributed, its splitting topology, i.e. horizontal, vertical, or hybrid, has to be considered [4]. If data is heterogeneous, semantic inconsistencies have to be identified. Additionally, export schemata information has to be prospected to guarantee semantic equivalence among heterogeneous data sources, e.g., a parts database, an operational database, and related scheduling information [5].

During this stage the data mining expert, having gained a clear understanding of the problem in the previous stage, and the data expert work closely together to map the problem onto the data sources.

3.4 Domain Knowledge Elicitation

The next stage is that of domain knowledge elicitation. During this stage the data mining expert attempts to elicit any domain knowledge that the domain expert may be interested in incorporating into the discovery process. The domain knowledge may take the form of domain specific constraints on the search space as well as hierarchical generalisations defined on the various attributes identified during data prospecting [6]. The domain knowledge must be verified for consistency before proceeding to the next stage of the process.

Example domain specific constraints are rules which specify known behaviour of a drill, with respect to temperature, material and drill type. Another constraint is the specification of bandings of continuous variables, e.g., measurements can be classified in small, medium and large. A typical hierarchical generalisation is a parts structure of a component to be assembled. Instead of searching for patterns on the atomic level of each parts, knowledge can be found on every granularity level of the parts hierarchy.

3.5 Methodology Identification

The main task of the methodology identification stage is to find the best data mining methodology for solving the specified problem. The chosen methodology depends on the type of information required, the state of the available data (accessed at the data prospecting stage), the problem at hand and the domain of knowledge being elicited. Often a combination of methodologies is required to solve the problem.

The most commonly used methodologies to model the discovered knowledge are traditional statistics [7], neural networks (modelling neurological functionality found in brains [8]), genetic algorithms (based on Darwin's evolutionary principal of the survival of the fittest [9]), fuzzy logic and rough sets (extending crisp set theory [10]), Bayesian belief networks (modelling conditional probabilities [11]), evidence theory (generalisation of Bayesian probability [12]) , case-based reasoning (modelling memory functionality based on cognitive psychology [13]), and rule induction (facilitating heuristics [14]). Since a detailed description of each of these methodologies is beyond the scope of this article, it is referred to in the references given for each discipline. Also, many design and control techniques have used those methodologies to model uncertain aspects.

It is important to stress that there is no methodology panacea which can tackle all data mining problems. For example, if an explanation of the discovered knowledge is required neural networks would clearly not be an appropriate methodology. The selected technique may influence the format of the input data, whose preparation is part of the following knowledge discovery step. For example, when using neural networks, data transformation may be required to map input data into the interval [0, 1] or when association rule induction is used, the data may need to be discretised or converted into a binary format depending on the association algorithm used.

3.6 Data Pre-processing

Depending on the state of the data this stage of the process may constitute the stage where most of the effort of the data mining process is concentrated. Data pre-processing involves removing outliers in the data, predicting and filling-in missing values, noise modelling, data dimensionality reduction, data quantisation, transformation, coding and heterogeneity resolution. Outliers and noise in the data can skew the learning process and result in less accurate knowledge being discovered. They must be dealt with before discovery is carried out. Missing values in the data must either be filled in or a paradigm used that can take them into account during the discovery process so as to account for the incompleteness of the data. Data dimensionality reduction is an important aid for improving the efficiency of the discovery algorithm as most of these have execution times that increase exponentially with respect to the number of attributes within the data set. Depending on the paradigm chosen the data may need to be coded or discretised.

Data pre-processing technologies can consist of a variety of tools, such as exploratory data analysis and thresholding for removal of outliers, interactive graphics for data selection, principal component analysis, factor analysis or feature subset selection for data dimensionality reduction, statistical models for handling noise in the data, techniques for filling in missing values, information theoretic measures for data discretisation, linear or non-linear transformation of data and semantic equivalence relationship handling for solving heterogeneity conflicts.

An example set of performed data pre-processing tasks is the selection of cases which have actually been tested at the quality assurance step, filtering out tuples with invalid values (e.g., negative widths), normalising continuous values (e.g., measurement deviations), and deriving new values (converting timestamps into numerical values).

3.7 Pattern Discovery

The pattern discovery stage follows the data pre-processing stage. It consists of using algorithms which automatically discover patterns from the pre-processed data. The choice of algorithm depends on the data mining goal. Due to the large amounts of data from which knowledge is to be discovered, the algorithms used in this stage must be efficient. It is usually better that the data mining task is not totally automated and independent of user intervention. The domain expert can often provide domain knowledge that can be used by the discovery algorithm for making patterns in the data more visible, pruning of the search space, or for filtering the discovered knowledge based on a user driven interestingness measure.

Different paradigms require different parameters to be set by the user. Example parameters are number of hidden layers, number of nodes per layer and various learning parameters like learning rate and error tolerance for neural networks, population size, selection, mutation and cross-over probabilities for genetic algorithms, membership functions in fuzzy systems, support and confidence thresholds in association algorithms and so on. Tuning these parameters is normally an iterative process and forms part of the refinement (see Section 3.9).

3.8 Knowledge Post-processing

The last stage of the data mining process is knowledge post-processing. Trivial and obsolete information must be filtered out and discovered knowledge must be presented in a user-readable way, using either visualisation techniques or natural language constructs. Often the knowledge filtering process is domain as well as user dependent. The most common way to filter discovered knowledge is to rank the knowledge and threshold based on the ranking. The ranking is often based on support, confidence and interestingness measures of the knowledge.

The support for a rule is the number of records in the database that satisfy the rule. The confidence in the rule is the belief that when the antecedent of the rule is true so is the consequent. Gebhardt [15] formalised interestingness by providing four facets of interestingness: the subject field under-consideration, the conspicuousness of a finding, the novelty of the finding and the deviation from prior knowledge. In general, measures of interestingness can be classified into objective measures and subjective measures. An objective measure depends on the structure of the pattern and the underlying data used. Subjective measures depend on the class of the users who examine the patterns. These are based on two concepts [16]: unexpectedness (a pattern is interesting if it is unexpected) and actionability (a pattern is interesting if the user can do something with it to his or her advantage) of the pattern.

Another aspect of knowledge post-processing is knowledge validation. Knowledge must be validated before it can be used for critical decision support. The most commonly used techniques here are holdout sampling, random resampling, n-fold cross-validation, and bootstrapping.

Due to the fact that the data used as input to the data mining process is often dynamic and prone to updates, the discovered knowledge has to be maintained. Setting up a knowledge maintenance mechanism may consist of re-applying the already set up data mining process for the particular problem or using an incremental methodology that would update the knowledge as the data changes keeping them consistent.

3.9 The Refinement Process

It is an accepted fact that the data mining process is iterative. After the knowledge post-processing stage, the knowledge discovered is examined by the domain expert and the data mining expert. This examination of the knowledge may lead to the refinement process of data mining. During the refinement process the domain knowledge as well as the actual goal posts of the discovery may be refined. Refinement could take the form of redefining the data used in the discovery, a change in the methodology used, the user defining additional constraints on the mining algorithm, refinement of the domain knowledge used or refinement of the parameters of the mining algorithm. Once the refinement is complete the pattern discovery stage and the knowledge post-processing stages are repeated. Note that the refinement process is not a stage of the data mining process. Instead it constitutes its iterative aspects and may make use of the initial stages of the process i.e. data prospecting, methodology identification, domain knowledge elicitation and data pre-processing.

4 Manufacturing Applications

There is a wide range of scenarios within manufacturing environments in which data mining has been applied successfully. Fault diagnosis is certainly the area in which data mining has been applied most often. Three case studies are described briefly. Other relevant areas are outlined in the sequel, which include process and quality control, process analysis, and machine maintenance[1].

4.1 Fault Diagnosis

Texas Instruments has isolated faults during semiconductor manufacturing using automated discovery from wafer tracking databases [17]. Firstly, associations (called classes of queries) are generated which are based on prior wafer grinding and polishing data. These classes have the potential to identify interrelationships among processing steps, which can isolate faults during the manufacturing process. Secondly, domain filters are incorporated to minimise the search space of the discovered associations. Thirdly, the interestingness evaluator tries to detect outliers, clusters (using the minimum description length) and trends (using Kendall's t-coefficient) which are feed back to the query generator. Lastly, another domain filter has been implemented to set interestingness thresholds, before finally a list of detected patterns is output.

Apté et al. facilitated five classification methods (k-nearest neighbour, linear discriminant analysis, decision trees, neural networks and rule induction) to predict defects in hard drive manufacturing [18]. Error rates at a critical step of the manufacturing process were used as input to identify knowledge (classes fail or pass) for further assistance of engineers. In the particular environment, none of the methods achieved outstanding results. The best outcome was achieved by rule induction, in order to minimise the high dimensionality of given data and thus, to improve the performance of the manufacturing quality control bottleneck.

The authors were involved in a project in which data mining has been applied in a wafer fabrication plant, namely Seagate Technology (Ireland) Ltd. The problem was to identify a lapse in the production of recording heads and discover its causes. The given data consisted of production process data including production parameters, test results and parameters, as well as some date and time stamps. After the data pre-processing was carried out (converting datetime fields into numerical values) and the classification labels were identified (pass and fail), a model was build using [19]'s C4.5 algorithm. The results indicated that from a certain date failed yield was far higher than usual. Based on that observation the data was refined in that two new fields 'before_date' and 'after_date' were derived which, formed the basis for an outcome with far higher accuracy. The new set of rules gave the participating engineer strong evidence of the cause of assembling failure.

4.2 Other Application Areas

Process and quality control is concerned with the correct performance of the entire life cycle of a product. During and/or after every product life cycle phase a control step is being carried out and measurements are taken to be compared against a norm. Deviations from that norm form a lucrative basis for data mining. Patterns being discovered can be classifications (types of deviations), sequential rules (intra-deviations), or temporal patterns (inter-deviations).

[1] Strongly related areas such as distribution, supply forecasting, and delivery are not considered in here, because they are more subject of logistics and marketing, and relevant literature can be found elsewhere, e.g., [1].

Process analysis is concerned with optimising interrelated processes within an enterprise. These are usually a combination of a primary process (company goals), control processes (strategic, adaptive, as well as operational), and support processes (assistance through human resources, means and knowledge). Identifying interrelationships among those processes has been proven a valuable field of data mining technology [20]. A related aspect of process analysis is concurrent engineering, which aims to perform internal and external requirements simultaneously. Again, data mining can provide mechanisms to identify interrelationships, and thus optimise the analysed process.

Machine maintenance is concerned with the correct timing of preservation of tools and instruments. Too early maintenance will be costly, whereas delayed maintenance can result in major productivity loss and customer dissatisfaction. Identifying patterns which indicate the potential failure of a component or machine is another potential exercise of data mining.

5 Conclusions and Outlook

The article has shown the capabilities of data mining, and how this technology can overcome several problems in manufacturing environments. It can even be argued that in the near future data mining has the potential to become one of the key components of manufacturing scenarios

Various problems still haven't been solved, which will certainly form further research in that synergetic area. The most relevant problem fields are real-time processing, which is indispensable in control theory, data quantity and quality collected by manufacturing units, as well as the degree of automaton in control environments.

Additionally, the data mining process and manufacturing processes are very loosely coupled, and minimising gaps between the cycles should improve the interaction of the two parts. A result of such a process re-engineering exercise would be the potential synergy of more advanced manufacturing disciplines, such as hybrid control systems or concurrent production.

REFERENCES

1. ANAND, S.S. and BÜCHNER, A.G.: **Decision Support using Data Mining**, Pitman Publishers, forthcoming, 1997.
2. CODD, E.F., Codd, S.B., Salley, C.T., **Providing OLAP to User-Analysts: An IT Mandate**, White Paper produced by Codd and Date Inc., 1993.
3. PIATETSKY-SHAPIRO, G., MATHEUS, C.J., **The Interestingness of Deviations**, Proc. Knowledge Discovery in Databases Worshop, pp. 25–36, 1994.
4. BÜCHNER, A.G., ANAND, S.S., BELL, D.A. and HUGHES, J.G., **A Framework for Discovering Knowledge from Distributed and Heterogeneous Databases**, Proc. IEE Coll. On Knowledge Discovery and data mining, Digest No 96/198a, pp. 8/1-8/4, London, 1996.
5. BÜCHNER, A.G., YANG, B., RAM, S., BELL, D.A. and HUGHES, J.G., **A Holistic Architecture for Knowledge Discovery in Multi-Database Environments**, Proc. ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'97), Tucson, AZ, pg. 87, 1997
6. ANAND, S.S., BELL, D.A., HUGHES, J.G., **The Role of Domain Knowledge in Data Mining**, Proc. 4th Int. Conf. on Information and Knowledge Management (CIKM95), pp. 37-42, 1996.
7. GLYMOUR, C., MADIGAN, D., PREGIBON, D., SMYTH, P., **Statistical Inference and Data Mining**, Comm. of the ACM, 39(11):35-41, 1996.
8. BIGUS, J.P., **Data Mining with Neural Networks**, McGraw-Hill, 1996.
9. GOLDBERG, D.E., **Genetic Algorithms in Search, Optimization, and Machine Learning**, Addison-Wesley, New-York, 1989.
10. LIN, T.Y., CERCONE, N., **Rough Sets and Data Mining**, Kluwer Academic Publishers, 1997.
11. HECKERMANN, D., **Bayesian Networks for Data Mining**, Data Mining and Knowledge Discovery, 1:79-119, 1997.
12. ANAND, S.S., BELL, D.A., HUGHES, J.G., **A General Framework for Data Mining based on Evidence Theory**, Data and Knowledge Engineering, 18:189-223, 1996.
13. REIDER, R, **Troubleshooting CFM 56-3 Engines for the Boeing 737 – Using CBR and Data-Mining**, Lecture Notes in Computer Science 1168, pp. 512-5??, 1996.
14. QUINLAN, J.R., **Induction of Decision Trees**, Machine Learning, 1:81-106, 1986.
15. GEBHARDT, F., **Discovering interesting statements from a database**, Applied Stochastic Models and Data Analysis, 10:1-14, 1994.

16. SILBERSCHATZ, A., TUZLHILIN, T., **What Makes Patterns Interesting in Knowledge Discovery Systems**, IEEE Trans. on Knowledge and Data Engineering, 8(6):970-974.

17. SAXENA, S., **Fault Isolation during Semiconductor Manufacturing using Automated Discovery from Wafer Tracking Databases**, Proc. Knowledge Discovery in Databases Workshop, pp. 81-88, 1993.

18. APTÉ, C., WEISS, S., GROUT, G., **Predicting Defects in Disk Drive Manufacturing: A Case Study in High-Dimensional Classification**, Proc. 9th Conf. Artificial Intelligence on Applications, pp. 212-218, 1993.

19. QUINLAN, J.R., **C4.5 Programs for Machine Learning**, Morgan Kaufman, San Mateo, 1993.

20. MULVENNA, M.D., BÜCHNER, A.G., HUGHES, J.G. and BELL, D.A., **Re-engineering Business Processes to facilitate Data Mining**, Proc. of the Workshop on Data Mining in Real World Databases at the Int. Conf. on Practical Aspects of Knowledge Management, Vol. 1, Basel, Switzerland, 1996.