Data Mining Methodology in Perspective of Manufacturing Databases

Muhammad Shahbaz[1], Muhammad Shaheen[1], Muhammad Aslam[1], Syed Ahsan[1], Amjad Farooq[1], Junaid Arshad[1], Syed Athar Masood[2]

[1] Department of Computer Science & Engineering, University of Engineering and Technology, Lahore, Pakistan
[2] Department of Engineering Management, NUST College of E&ME, Rawalpindi, Pakistan
Muhammad.Shahbaz@gmail.com, M.Shahbaz@uet.edu.pk,

Abstract: In recent years data mining has become a very popular technique for extracting information from the database in different areas due to its flexibility of working on any kind of databases and also due to the surprising results. This paper is an attempt to introduce application of data mining techniques in the manufacturing industry to which least importance has been given. A taste of implement-able areas in manufacturing enterprises is discussed with a proposed architecture, which can be applied to an individual enterprise as well as to an extended enterprise to get benefit of data mining technique and to share the discovered knowledge among enterprises. The paper proposes conceptual methods for better use of different data mining techniques in product manufacturing life cycle. These techniques include statistical techniques, neural networks, decision trees and genetic algorithms. An integrated and unified data mining platform is anticipated then to improve overall manufacturing process.

[Muhammad Shahbaz, Muhammad Shaheen, Muhammad Aslam, Syed Ahsan, Amjad Farooq, Junaid Arshad, Syed Athar Masood. **Data Mining Methodology in Perspective of Manufacturing Databases.** Life Science Journal. 2012;9(3):13-22] (ISSN:1097-8135). http://www.lifesciencesite.com. 3

Keywords: Data Mining, Manufacturing, Industrial application, Data Mining methodologies, Data Warehousing

1. Introduction

Computer integrated manufacturing systems as well as more simply controlled enterprises, generate huge amounts of data daily but even when companies appreciate the importance and value of this information, people seldom try to explore their databases thoroughly, because of other ‘urgent’ commitments towards their engineering and technical duties. The usual approaches for addressing and controlling problems that arise in manufacturing areas is through the application of engineering knowledge and experience. Yet these problems may also be tackled and solved by analysing the available operational information directly. This involves sifting through past data that exists in the enterprise’s databases to see if any trends exist, or are emerging that may be causing specific defects or faults. Therefore, data mining techniques could be applied to improve exploitation of valuable information and knowledge sources, to better control the system and also to check the strategic gains and losses during the manufacturing.

In the section two we will describe, at a high level, areas in manufacturing enterprises where benefit can be gained from data mining technology. We shall also provide some details of design constraints and discuss how product life cycle data can be used to explore and discover knowledge by using data mining techniques. Section three covers data warehousing, as this plays an important part in implementing data mining and saves time spent on the initial steps required for data pre-processing. The fourth section is about the real technology of data mining, a brief background and description of some of its popular tools are given, in the perspective of manufacturing enterprises. In the last section an integrated data mining model is proposed, keeping in mind the requirements of manufacturing enterprises and how a generic data mining system can fit within a manufacturing system.

2. Manufacturing System Design and Performance Improvement:

The importance of data oriented knowledge discovery techniques cannot be denied in any industry. Manufacturing enterprises in particular normally generate quantities of data at every step of the manufacturing processes from design through to the disposal of the product, and generally, most of this data is not fully exploited. Currently there is very little research being carried out in the manufacturing sector into the application of data-oriented knowledge discovery techniques.

Efforts have been made in the recent past to utilize the databases from manufacturing enterprises for design and quality control processes for example the factory data model [2][3] and data warehouses. The factory data model proposed in [2] promotes better exploitation of the information residing in the databases. It is however very difficult to design a data supported manufacturing enterprise which can gain the benefits of their historical as well as their current databases.

In manufacturing systems involving small sized products in very large volumes, for example in the semiconductor industry, quality checking has always been a problem and detecting any process abnormality as early as possible becomes more crucial than ever. Similarly in the production of very large sized products with small volumes it is desirable to maintain the non-conformities level to a minimum. The aim of current manufacturing systems is to decrease the time from occurrence of a fault to its detection. The shorter the identification time the more controlled the manufacturing system is. It is therefore important to learn from both past problems and successes, and to use this existing knowledge to improve product designs.

Design and Re-usability of knowledge:

Design can be defined as "effective allocation of resources"[4]. Alternatively, Pahl and Beitz [5] describe it as "a process of synthesis and integration". The most comprehensive definition is given in [6] which states, "Systematic, intelligent generation and evaluation of specifications for artefacts whose form and function achieve stated objectives and satisfy specified constraints." In their opinion, an engineering design does not directly result in a physical product compared with other design domains but rather, it provides a set of specifications to construct or fabricate the products.

Conceptual design, layout design, drafting and design analysis all create data, (which in turn may be used as the basic raw material for a data mining process). Product and process faults may occur at any stage of the product's life cycle, hence, it is possible that certain faults can be traced back, using the data from the design and the design process. Earlier computer aided designs or old design drafts can be reused and be very helpful in the redesign of the product providing they can be analysed in a sensible way. A potential difficulty exists when attempting to re-use previous design knowledge that exists in the form of archival design documents, testing and analysis reports [7]. The difficulties lie in the forms in which the data has been archived, as some paper-based or hard forms are difficult and slow (and therefore expensive) to search and reuse. Therefore when designers or managers want to consult (and learn) from these designs or previous results the costs may be substantial in terms of time and overheads. Designers spend 20-30% of their time looking for information and same amount of time in handling information. Therefore any design system should incorporate all the designers' own files, to make information easy to find and easy to use [7].

Data searching and retrieval become less of a problem when information is computer based. However substantially more efforts than are commonly used at present are needed to reuse such knowledge effectively. Data Mining could provide a solution to these problems by finding relationships between design problems and production problems or other aspects relating to the product life cycle. Manufacturing system design requires considerable amounts of information to be collected, and processed in order to improve the design and performance of the processes. Several information modelling techniques, and processing methodologies are already in use as described in detail within [2][3].

The popularity of computer aided design (CAD) and computer integrated manufacturing (CIM) have increased the amount of available digital data. This is easy to store and recover or search, helping the modern manufacturing enterprise to make better use of computing power in the analysis of its valuable data. Many researchers believe that computational approaches to design should enhance, not replace, human practice [8] [9]. With the help of knowledge discovery in databases, the causes and contributory factors to faults may be more clearly identified, enabling human designers to focus and concentrate their efforts on important problem areas and thus the time to design or redesign a product can be reduced.

It is generally important to consider how a product will be produced in parallel with the design of the product. Improvements to the manufacturing system can also result in improvements to the product. The manufacturing system, modelling and design, will therefore now be considered in the context of data mining.

Advanced Manufacturing Systems:

Achieving absolute flexibility requires substantial quantities of information about the current processes and related activities, which directly or indirectly, link the whole manufacturing system. Information about the product life cycle also plays an important role in designing the flexible manufacturing process. The details of the product life cycle and employment of data mining on it are discussed at the end of this section.

Flexible manufacturing can be achieved by having tools, machining programmes and parts all quickly changeable and being able to respond rapidly by minimizing the part lead times. Machine utilization should be maximised together with an almost instantaneous response both to customers and to any problems that may occur. The philosophy behind this arrangement simultaneously aims for minimisation of lead times for the parts to be processed along with the maximisation of the utilisation of the machines doing the processing [14]. But how can this be achieved successfully?

Since it is difficult to measure manufacturing flexibility, it is often hard to financially justify investments aimed at increasing the flexibility of a manufacturing system [15]. Manufacturing enterprises are struggling to make all the steps in their operation as flexible as possible, but often without proper knowledge and information sharing between the various stages of the manufacturing processes. The hurdles in making the flexible manufacturing systems can be analysed using data that has been recorded during individual processes within the whole system. Keeping in mind the problems data mining will give the solutions to the problems that are actually indirectly involve in the brittleness of the system. Once the information about the system's brittleness is discovered the system can be made more elastic by removing the obstructions in achieving the goal.

The first step towards applying data mining techniques for achieving the maximum flexibility in the manufacturing system is to make a system to archive the data recorded during different operational stages of the manufacturing organizations. Most advanced manufacturing enterprises record such data but it is often in the form of simple database files that are not well organized. A well designed database or model is required, otherwise substantial pre-processing of the data may be necessary before it can be mined and this will consume considerable resources. The first step towards achieving this goal is to develop a data warehouse, and this will be discussed further in section three.

If a data warehouse is utilized properly it can not only help in developing the new strategies but it can also be used for repairing the strategies of an enterprise in crisis. For ongoing fault detection, it will some times be necessary to have an on-line data mining system, which does not have to rely on the data warehouse. The on-line systems will work on fresh data direct from the manufacturing system to find out any developing trends towards bad quality or to help better scheduling of resources. In an advanced manufacturing system there are advantages to automating the data mining system to help directly in controlling the process.

Manufacturing Strategies and Data Models:

An enterprise needs to redesign if its aims are not being achieved or if its aims or strategies change. Good strategy and a well structured enterprise result in profit whereas bad strategy or a business that does not meet the competitors' challenges, damages the company in the marketplace. Therefore management decisions should be based on accurate and reliable information that is structured within a data warehouse and a factory data model [2]. A factory model focuses on operation and infrastructure in contrast to a data warehouse provides information about the behaviour of the existing enterprise. Both these source of information are vital for the design and redesign of an enterprise and for performance evaluation.

However, the existence of this useful information is only part of the solution. How it can be utilized effectively to produce the required results is equally important. Simple statistics normally work well to give a very good picture of the current overall manufacturing process but there may be much more hidden knowledge waiting to be discovered. Machine learning and artificial intelligence tools can be used to gain insight into the data and to discover hidden patterns and trends.

The above analysis regarding the ways of extracting information may not appear to provide adequate solutions as the results of the mining process cannot be predicted. Hence, we cannot be sure that the value of the knowledge that may be discovered in the data will be greater than the time and resources that need to be spent in the mining process. However implementations of the same technology in other areas of human sciences like banking, finance, marketing, insurance, telecommunication, health care etc. have given very good results [16] [17] and people are now benefiting from the knowledge they have gained in their respective fields – so why should manufacturing enterprise not also benefit?

Product Life Cycle and Data Mining:

The product's life cycle is based on the design of the product and is important for a data collection point of view and for the analysis for the mining perspective. Information collected during a product's life cycle provides feedback on a product's performance that can be used to assess the quality of the design [18]. Nahmias [19] divided the product life cycle into the following four phases: start-up, rapid growth, maturation and decline. Hazelrigg [4] made a further in depth division and split the whole process into seven stages: engineering and design; test and evaluation; manufacturing; distribution and sales; operations, maintenance & repair and disposal. All these stages produce data, which should be more or less easily available for analysis to gain insight into the whole life cycle of the product. Knowledge gained through this analysis can be used for redesign or for the introduction of new products.

Once the production process is completed there are many reasons why it is difficult to collect information relating to the costs. This information is very important regarding the design or redesign of the products. Prasad [20] concluded from different studies that indirect costs could be as much as 4-5 times the amount of direct labour and material costs. In this situation if all the relevant data from the product life cycle is available then it is far easier to analyse the data and to target the areas for improvement in the design and the manufacturing process. This is the point where data mining comes into play. Figure1 shows a very high level architecture of collecting the data during the product life cycle and storing it in a data warehouse, which is actually coupled with a data-mining engine. The engine would use either one or a combination of the many available data mining techniques depending on the type of the data and the type of problem to be solved. A brief introduction and relevant details of a few of these techniques and their applicability can be found in the fourth section of this paper. A design improvement that takes into consideration all the aspects of the product life cycle gives a much more controlled and efficient design compared with a design based on market demand and production capabilities. The advantages of mining data relating to the product life cycle are that it can improve the design or redesign of the product and also that it provides more information that is helpful for design of the production process, marketing strategy, environment effects, quality and reliability of the product. This information collected during a product's life cycle provides feedback on product's performance that can be used to assess the quality of the design [18].

Data for mining can be structured so that the overall data from all the major stages of the life cycle can be examined together. In addition, analysis should also be carried out on sections of the data relating to individual stages in the life cycle. This is necessary to detect any hidden relations and effects on subsequent life cycle stages so that corrective actions can be taken to improve the design of the product or the manufacturing process.

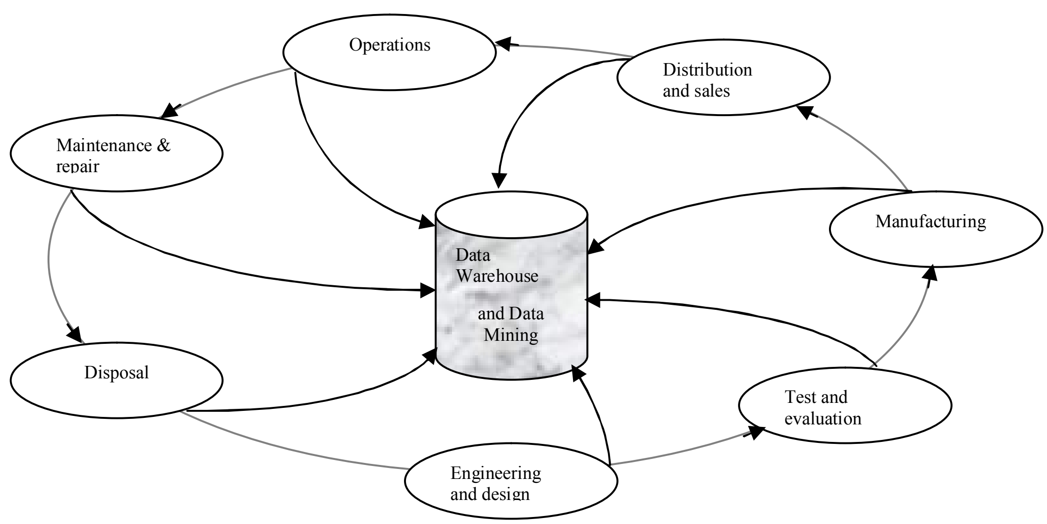

Figure1: Product Life Cycle's Data Collection Architecture.

3. Data Warehousing

Data warehousing has developed over the last 20 to 30 years, as they started emerging between 1984-88 [21]. The trends of today's enterprises towards data warehousing can be judged by a survey conducted by the META Group (Stanford, USA) that shows almost 95% of the corporations have plans to build data warehouses [22].

A database is a generalised and integrated collection of stored and operational data together with their descriptions, and is managed in such a way that it can fulfil the differing needs of its different users [14]. A data warehouse can be defined as a single, complete, and consistent store of data obtained from a variety of sources and made available to end users in a way they can understand and use in a business context [21]. Mattison[23] explained data warehouse as a collection and organization of data to serve as a neutral data storage area that is non volatile, can readily be used for data mining and/or other applications, and meets the specific business requirements.

The data warehouse must be structured to respond to queries related to different aspects of the process or business [29]. Users normally access operational databases using transactions that are also called On-Line Transactional Processes (OLTP), and often use executive information system (EIS). The disadvantage of OLTP is that the results of the queries of two users may be different if their queries are made at different times because the data is changing continuously and the second user may get updated results. However, with the data warehouse all the users will get the same results for similar queries, providing the queries are made during the intervals between updating the warehouse.

4. Knowledge Discovery in Database

The term knowledge discovery in databases (KDD) was formalized in 1989 in reference to the general, 'high level' concept of seeking knowledge from data. This term (data mining) has been used by statisticians, data analysts and the management information systems (MIS) community. KDD has been mostly used by artificial intelligence and machine learning researchers.

Data mining is just one of several terms, used by the people in the data mining field, including knowledge extraction, data archaeology, information harvesting, and even data dredging, that actually describe the concept of knowledge discovery in databases.

Data Mining includes all methods and techniques, which allow practitioners to analyse very large data sets to extract and discover previously unknown structures and relations out of huge quantities of details. Information is filtered, prepared and classified so that it will be a valuable aid for decisions and strategies [30]. The most authentic definition of knowledge discovery in database or data mining is "non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [31]. With the explosive growth of data in databases the desire to extract useful information is also increasing. The manufacturing enterprises' databases contain treasures of information that tempt analysts to detect trends or patterns in them and react flexibly to them. However useful information may be hidden within the mountains of other data, and cannot be discovered using conventional database management systems. Data mining is becoming an increasingly important research area [32] [33] [34], since knowledge, e.g. extracted knowledge trends and patterns, can be used to help and improve business decision making.

Data Mining:

Data Mining normally works together with a data warehousing as this is necessary to organize historical information gathered from large-scale client/server bases applications as discussed in the previous section. Due to the explosive growth of data in the companies in the recent years and the non-availability of any proper technology to exploit this data in the past made data mining a very important research topic [31] [39] [40].

In order to understand how KDD can be implemented in manufacturing enterprises, it is important to understand the whole process, especially the discovery stage and its tools. For simplicity the whole knowledge discovery process is divided into six different stages: data cleaning, integration, selection, transformation, mining and evaluation/visualization [41] [42]. Fayyad [31] proposed five steps including, retrieval of data, selecting of data, sampling and cleaning, applying the appropriate transformations and fitting models to the proposed data. IBM [43] defined the four major operations for data mining as predictive modelling, database segmentation, link analysis and deviation detection. The above divisions show the actual discovery or mining process comes at the end and takes a very small fraction of the total time involved in the discovery process. Gonzalez and Kamrani [44] conclude that as much as 80% of KDD is about preparing data, and only the remaining 20% is about mining. Pre-processing is therefore very important and the details of how to implement the pre-processing steps can be found in lots of available data mining books in the market.

Some standard methodologies like SEMMA (Sample, Explore, Modify, Model, Assess) [45] and CRISP-DM (CRoss Industry Standard Process for Data Mining) [46] have also been developed for data mining process and to simplify its implementation in the industry.

Data mining in its traditional forms has been used to find patterns in the historical databases like banking, insurance, fraud detection, telecommunication data etc keep their old data for future strategy and planning. In the manufacturing process both the old and current trends of the process; policies & strategies and quality are important. The traditional way of data mining can help in finding the faulty processes and bad strategies in the manufacturing process and suggest remedies for them but finding the online or current trends is very important in any kind of manufacturing processes to control the whole process for better scheduling and quality. Data mining can also be used to solve this problem by embedding it in the process to find out the run time errors of the process if they occur [47]. In this kind of mining process, all the pre-processing steps are eliminated by automating the whole process for collecting the data, analyzing it and making corrective actions.

Data Mining Techniques for Manufacturing Enterprises:

The analysis or mining of the manufacturing enterprise data can be done using all the popular data mining techniques. Some of the effective techniques for data mining like association rule, rule induction etc. are mostly used for retail market or basket analysis [48] [49] but are helpful in any kind of manufacturing databases too. The data mining techniques can be divided into three main categories, statistical techniques which uses simple to complex statistics to analyses the databases, the second category is artificial intelligence tools which become popular with the increase in the computing power over the last two decades and the third is machine learning tools which are actually a combination of statistics and artificial intelligence tools.

A very interesting survey has been done by [50] analyzing the efficiency and productivity of different data mining techniques with different kinds of problems. For example the survey shows that for data that has many attributes, like manufacturing data, and that is numeric in nature, the most suitable algorithms are decision trees, nearest neighbors and neural networks. But it really depends on what kind of problems are being examined in the manufacturing enterprises. The above techniques can be used to search for any kind of trends in the past data but in any specific problem the choice of the technique really also depends on other factors of the problem and algorithms.

A few of the most common data mining tools are listed here in the context of manufacturing databases. Only a brief introduction is given, as details can be traced back the references provided.

Statistics:

Statistics can be counted as a data mining tool since statistics is actually the origin of data mining. There are lots of statistical techniques including regression, discriminant analysis, classification, clustering and time series which are very popular in the data mining community and are extensively used for the large database analysis. With the passage of time these techniques are now mixed with the artificial intelligence tools to give even better, more reliable and faster results than the current simple statistical methods.

Statistics has always been a very popular tool in the manufacturing enterprises for process and quality controls. In any kind of data mining process the preprocessing stage commonly uses different statistical techniques and the initial analysis of the data is also done using statistics and Structured Query Language (SQL). [42] [51] and [52] have stated that for the most part, about 80% of the interesting information can be abstracted from a database using SQL commands. However, as [30] stress, extraction of the remaining 20% of hidden information requires advanced techniques like expert systems, fuzzy expert systems, case-based learning, decision trees, neural networks, genetic algorithms etc. Hence statistics are a good starting point for the analysis of the data, to try to identify some trends for further detailed analysis.

Decision Trees:

Decision trees are normally used for classification purposes. These are tree shape structures resulted by the decision taken at each node. The database is divided into different fields that enable the analysts to look at the behavior of the database at different stages or to distinguish among different patterns present in the data. Different decision tree methods used as a data mining technique are Classification and regression Trees (CART) [53] and Chi Square Automatic Interaction Detection (CHAID) [54]. The first efficient decision tree model called ID3 [55] was based on the concept of entropy means the choice of the next feature used for branching should increase knowledge [56]. Decision trees are simple enough to understand and explain, and are easy to build, have relatively short training time and need very low memory [7].

Neural Networks:

Neural networks, is a very popular AI technique that mimics the working of neurons of human brain. Neural networks are not new as they trace back their history about 50 years ago when McCulloch and Pitts started working on them. [57], [58]. Neural networks are complex to interpret but very good in terms of accuracy. Carol [7] tabulated different data mining techniques such as neural networks, rule induction, decision trees, nearest neighbor etc and tabulated their important characteristics. It is a good idea to keep in mind the different characteristics of the data before choosing a specific data mining technique.

Artificial neural networks are simple computer programmes, which can automatically find non-linear relationships in data without any predefined model. According to [59], neural network-based database approach consists of three major-phases:

1 - Network construction and training: in this phase a layered neural network based on the number of

attributes, number of classes and chosen input coding method are trained and constructed.

2 – Network pruning: in this phase, redundant links and units are removed without increasing the classification error rate of the network.

3 – Rule extraction: rules are extracted in this phase.

Genetic Algorithms:

This is one of the most recent methodologies used as a data-mining tool. Their basis is on the evolutionary computing which become very popular within the machine learning methodologies [60]. The basic concept of Genetic Algorithms comes from Darwin's theory of evolution. A genetic algorithm is reminiscent of sexual reproduction in which the genes of two parents combine to form those of their children and only the fittest will survive. The next generation improves and is better than the previous generation only if the strongest members of the population mate together to produce the next generation. The same principle can be applied to problem solving if the population consists of possible solutions to the problem. Each of these generated solutions have some characteristics that enable them to be categorized as a more or less fit as member of the next generation of offspring. The best members of a generation are given more chance for mating and producing the subsequent generation. In this way, each successive generation consists of better solutions, until an optimal solution is generated

Genetic Algorithms can be very helpful in finding solutions that are very difficult to optimise. Another advantage of using genetic algorithms is that they can propose many possible solutions of a problem. The main advantage of using genetic algorithms (GAs) is, they can be synthesized without making use of the detailed, explicit knowledge of the underlying process. This means they will find a pattern if any exists even if the problem is new and no previous solutions are known. However, limited or noisy training data may result in inconsistent, meaningless output. This has been known to be a severe problem of genetic algorithms [61]. Another problem for genetic algorithms is they require lot of computing power to achieve a significant solution. In data mining problems specially to find out the relationships between the different entities genetic algorithms prove to be very effective. [62] shows a successful implementation of a search technique in big databases using genetic algorithms in solving a data mining problem.

The generic genetic algorithm consists of the following steps [56]

1- Each offspring (generation) is evaluated for fitness
2- The population is increased through mating and mutation of fit members to generate the new set of rules (generation)
3- Weak members from the generations are eliminated (reducing the size of the population)
4- A terminating condition is checked and if the optimal solution is not achieved then mating and mutation are done to produce the new generation.

There are many other popular data mining techniques including fuzzy logic, rule induction, association rule, k-nearest neighbor, intelligent agents etc, and this list continues to increase.

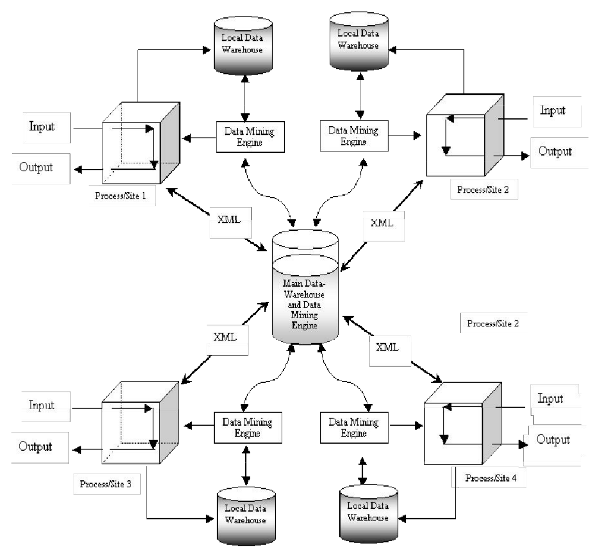

Figure 3. An Integrated Data Mining System Model

5. INTEGRATED DATA MINING

As explained in the first section of this paper a manufacturing enterprise can benefit from data mining in solving problems, but there is an important need for an integrated data mining system for diagnosing and solving online manufacturing problems. Here we introduce a relatively complex integrated data-mining architecture for a manufacturing enterprise. This architecture can be used to suggest solutions to particular problems, to learn from this, and subsequently to also help solve problems at other processes/sites with similar kind of parameters.

The KDD process is generally driven by individual user skills and experience and is not efficient to use in manufacturing applications where fast intervention may be required if things are going out of control. Since the production process in a factory is usually automated and manufacturing data is obtained continuously from the process, the data must be monitored without interruption so that any anomaly in the process may be detected and eradicated immediately, therefore, the KDD process must run concurrently with the production process [47]. Enterprises aim for increasing levels of accuracy and improvement as manufacturing processes become more advanced and sophisticated and data is continuously recorded.

Manufacturing databases are dynamic with very regular updates made to the records. In the proposed integrated data mining model, data mining techniques can be applied at the micro level within sections of the manufacturing processes, whilst a Main Data Mining Engine may coordinate, share and exchange knowledge between the individual data mining engines. The main Data Mining Engine when connected to individual data mining engines establishes a data-mining network (see Fig. 3), which helps to mine the whole process. The whole manufacturing process may be divided into small steps and the data relating to each step and its adjacent steps are mined independently. The same principle can be applied to the extended manufacturing enterprises where the sites are not in one premises but are located at different places, cities, countries or even continents. The whole process is therefore supported and explored by the remainder of the network. Its activities and results are consistently communicated through the Main Data Mining Engine. Thus individual areas of data mining activity are kept to a manageable size whilst still supporting related areas of activity.

In an integrated manufacturing environment where the product is developed at separate locations or in discrete steps data mining can best be used to control any individual process or step through identification of hidden information in its associated data. The data-mining algorithm say rule induction can discover the relationships within the individual process, or dependent processes. It cannot however be utilized for the other discrete steps unless a common data warehouse collects the data. The concept of the proposed integrated data mining model works on the basis that rules; principles and concepts applicable to one manufacturing stage/site may also be utilized (tested and applied) for other similar stages, (by the exchange of activity and rules information via the Main Data Mining Engine). This integrated data mining model will work where a product goes into different stages and data for each and every step is collected and stored in a pre-designed data warehouse or in the pattern warehouse as knowledge is much more compact than data [63]. Data mining activities and the main data warehouse will work in parallel during the whole activity with the production process data and company data warehouses.

Each stage will have its own local data mining engine and data warehouse where the data will be stored after cleaning. The data will also be transferred to the main data warehouse, which has a direct link with a pattern warehouse for the analysis or mining of the whole system's data.

If the same methodology is applied to an integrated system where production is being done at different sites then the central data warehouse will be built on a standard format. The data from the individual sites will be transferred to the main data warehouse using XML format where the data will be mined for the whole process and rules/knowledge extracted will be returned back in the same format.

The integrated data mining will be productive in the sense that if different rules are identified for two individual, but similar small manufacturing/production steps, the rules can be shared, and each data-mining engine can use its knowledge to refine the "best" one for its particular application. In this way, knowledge can be fed into the main data warehouse, so results can be reused in the future, as a pattern warehouse will be developed. Future applications can then make use of the stored patterns and rules instead of always having to return to the original manufacturing process databases.

The figure shows fours different manufacturing processes or sites. The input and output of the manufacturing process is shown and the data from the manufacturing process is collected at the local data warehouse and is also transferred to the main central data warehouse in a neutral format say XML. The outcome of the data mining process (if any) is implemented to the manufacturing process and same information is also reported to the central system which analyze the kind of problem tackled by the local system and stores an index and the parameters for the problem occurred. If any of the other processes indicate similar kinds of problem then the central or Main Data Mining Engine first tries the same solution to that problem to see if it works. This methodology will help in future to minimize the time spent on understanding the problems and finding the solution. The data-mining engine can also suggest alternative solutions, but using an old solution and refining it to suit the present requirements should help to tackle the problem in more efficient, cost effective ways.

The integrated model should also be aided with the online visualization model so that the worker who is working on the machines/products can get a clear idea about the process and products. Visualisation should allow a user to discuss and explain the logic behind the model with colleagues, internal/external customers, and other users [64]. For example if there is a recurring problem with a particular step or process on a product then visual checks for previous trends/rules or current results at other sites with the same or similar steps will help in getting a workable idea to fix the problem. Therefore visualisation of the whole mining process will help and enable the output of the data mining system to be understood qualitatively.

Conclusion:

It has been showed that along with other areas, manufacturing enterprises can be benefited with the data mining techniques. There are lots of areas within manufacturing enterprises, few of them are explained in this paper, where data mining can find its ways of implementation and can give results comparable to any other corrective measures based on the engineering methodologies. The data mining architecture proposed in this paper can be refined from a small factory to an extended enterprise. Such kind of data mining approach will be an essential part during the designing of an advanced manufacturing process in future which can learn from its own mistakes and will do the corrective actions not only for its own processes but will help the other processes with its experiences.

References

[1] Amir Netz et al, "Integration of Data Mining and Relational Databases"

Proceedings of the 26th International Conference on Very Large Databases, Cairo, Egypt.

[2] J.A. Harding, B. Yu, "Information-centred enterprise design supported by a factory data model and data warehousing". Computers in Industry, Elsevier, 40 (1999) 23-36

[3] Harding, J. A., Yu, b., Popplewell, J., "Information modelling: an integration of views of a manufacturing enterprise", International Journal of Production Research, 1999, Vol. 37, No. 12, 2777-2792

[4] Hazelrigg, G. A., "Systems Engineering: An approach to Information-Based Design". Upper Saddle River, NJ: Prentice-Hall, Inc., 1996

[5] Pahl, G., and Beitz, W., Engineering Design London: Springer-Verlag, 1999

[6] Dym, C. L., and Levitt, R. E., "Knowledge-Based Systems in Engineering" , New [7] Carol J. r., Rakesh N., "A data Mining-Bases Engineering Design Support System: A Research Agenda", Data Mining for Design and Manufacturing: Methods and Applications Ed. Dan Braha, Kluwer Academic Publishers 2001

York: McGraw-Hill, 1991.

[8] Duffy, S. M., and Duffy, A. H. M., "Sharing the learning activity using intelligent CAD", Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 10, pp. 83-100, 1996

[9] Reich, Y., "Layered models of research methodologies", Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 8, pp. 263-274, 1993

[10] ESPRIT Project 688/5288, AMICE, Vol 1, CIMOSA: Open system architecture for CIM, ESPRIT Consortium AMICE (Eds.), 2nd revised and extended edn., Springer, 1993.

[11] A.W. Scheer, Architecture of integrated information systems: Foundation of Enterprise Modelling, Springer, Berline, 1992

[12] G.Doumeingts, D.Chen, B. Vallespir, P.Fenie, GIM (GRAI Integrated Methodology) and its evalutions – a methodology to design andd specify advanced manufacturing systems, in: H.Yoshikawa, J.Goossenaerts (Eds.), Proceedings of the JSPE/IFIP TC5/WG5.3 Workshop on the Design of Information Infrastructure Systems for Manufacturing, DIISM, 93, Tokyo, Japan, 8-10 November 1993, North-Holland, pp. 101-117

[13] Bartezzaghi, E., and Turco, F., "The impact of just-in-time on Production system performance: an analytical framework". International Journal of Operations and Production Management, 9(8), 40-62, 1989

[14] Roger Hannam, "Computer Integrated Manufacturing: from concept to realization", Addison-Wesley Longman Limited, 1996

[15] Hill, T., and Chambers, S., Flexibility: a manufacturing condundrum. International Journal of Operations and Production Management, 11(2), 5-13, 1991.

[16] Hashimot, K, Matsumoto, K, Terabe, M. "Applying Data Mining Techniques to Telecommunication Services" Japanese Society for Artificial Intelligence, vol 17, part 3, p. 320-325 ISSN: 09128085, 2002

[17] Omer F. A., Ertan K., Piero M., "Data Mining for Database Marketing at Garanti Bank", Data Mining II, Edecken, NFF et al (Editors), WIT Press, 2000.

[18] Ertas, A., and Jones, J., "The Engineering Design Process." New York: John Wiley & Sons 1996

[19] Nahmias, S., Production and Operations Analysis. Burr Ridge IL: Irwin, 1993.

[20] Prasad, B., "Survey of life-cycle measures and metrics for concurrent product and process design." Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 14, pp. 163-176, 2000

[21] Devlin, B. "Data Warehouse: From Architecture to Implementation", Addison-Wesley, Reading, MA. (1997).

[22] Brown, A. J., "What is a data warehouse?" UNIX Review, Vol. 13 No. 9, p. 39 (1995)

[23] Mattison, R., "Data Warehousing: Strategies, Technologies and Techniques", McGraw-Hill, New York. (1996)

[24] Poe, V., Klauer, P., Brobst, S. Building a data warehouse for decision support, 2nd edition, Prentice Hall PTR, 1998, p. 18

[25] Westerman, P. "Data Warehousing": Morgan Kaufmann Publishing, Inc. San Francisco, CA 2001

[26] W. H. Inmon, "Building the Data Warehouse", 2nd edn., ISBN 0-471-14161-5, Wiley computer Publishing, wiley, 1996

[27] A. Subramanian, L.D.Smith, A.C. Nelson, J.F. Campbell, D.A.Bird, Strategic Planning for data warehousing, information and Management 33(2) (1997) 99-113

[28] M. Ferelli, "Why data warehousing?", Storage Management Solutions 3 (2) (1998) 4.

[29] L.Handen, P.Boyle, “Data warehousing: managing knowledge”, Telecommunications (International Edition) 32 (1) (1998) 61, 67-68
[30] Johannes, G., Andras R., “Techniques of Cluster Algorithms in Data Mining”, Data Mining and Knowledge Discovery, 6, 303–360, 2002, Kluwer Academic Publishers.
[31] Fayyad, U.M., Piatetsky-Shapiro, G and Smyth, P (1996), "From data mining to knowledge discovery" in ‘Advances in knowledge discovery and data mining’, (eds) Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P and Uthurusamy, R, AAAI Press/MIT Press, CA., p 1-34, ISBN: 0-262-56097-6.
[32] W. Kl.osgen and J. Zytkow. Knowledge discovery in database terminology. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, Advances in Knowledge Discovery and Data Mining, pages 573{592. AAAI/MIT Press, 1996.
[33] D. E. O'Leary. “Knowledge discovery as a threat to database security”, In G. Piatetsky-Shapiro and W. J. Frawley, editors, Knowledge Discovery in Databases, pages 507{516. AAAI/MIT Press, 1991.
[34] A. Silberschatz, M. Stonebraker, and J. D. Ullman. “Database research: Achievements and opportunities into the 21st century”, ACM SIGMOD Record, 25:52-63, March 1996.
[35] Boyer, C.B. “A history of Mathemactics”. : John Wiley & Sons Inc. New York 1968.
[36]Vapnik, V.N. The nature of statistical learning theory New York: Springer-Verlag, 2000
[37] Langely, P. “Elements of machine learning”: Morgan Kaufmann Publishing, Inc. San Francisco, CA 1996
[38] Hosking, J.R.M., Pednault, E.P.D., & Sudan, M. 2000. A statistical perspective on data mining. Further generation computer system. 13: 117-134, 2000
[39]Piatetsky-Shapiro, G. and Frawley, W.J. “Knowledge Discovery in Database”, AAAI/MIT Press. (1991)
[40] Weiss, S.H. and Indurkhya, N. “Predictive Data Mining: A Practical Guide”, Morgan Kaufmann Publishers, San Francisco, CA. (1998)
[41] Jiawei Han, Micheline Kamber, Simon Fraser University “Data Mining: Concepts and Techniques”, Morgan Kaufmann publishers 2001
[ISBN 1-55860-489-8]
[42] Adriaans Pieter, Zantinge Dolf, “Data Mining”, Addison-Wesley 1996 [ISBN 0-201-40380-3]
[43] Technology Forecast (Report): 1997, Price Waterhouse World Technolgy Center, Menlo Park, CA, 1997
[44] Ricardo Gonzalez, Ali Kamrani, “A survey of Methodologies and Techniques for Data Mining and Intelligent Data Discovery” in “Data Mining for Design and Manufacturing: Methods and Application”, Kluwer Academic Publishers, 2001, [ISBN: 1-4020-0034-0]
[45] Enterprise Miner™
http://www.sas.com/technologies/analytics/datamining/miner/semma.html
[46] CRISM-DM 1.0, “Step by step data mining guide”, © SPSS (2000)
[47] Hideyuki M, and Yuko T., “ Development of Automated Data Mining System for Quality Control in Manufacturing”, Lecture Notes in Computer Science, Online publication: August 28, 2001 © Springer-Verlag Berlin Heidelberg 2001
[48] Maria de Fatima Rodrigues, Carlos Ramos & Pedro R. Henriques, “Knowledge Discovery with Retail Basket Analysis”, Data Mining Editor: Nelson F. F. Ebecken, WIT Press [ISBN: 1853126772] pp. 193-204
[49] R. Agrawal, R. Srikant: "Fast Algorithms for Mining Association Rules", Proc. of the 20th Int'l Conference on Very Large Databases, Santiago, Chile, Sept. 1994
[50] Moustakis, V.S., Letho, M., Salvendy, G., “Survey of expert opinion: which machine learning method may be used for which task?” Special issue on machine learning of International Journal of HCI, 8(3), pp. 221-236. 1996
[51] Yevich, R., “Data Mining”, in Data Warehouse: Practical Advice from the Experts, pp. 309-321, Prentice Hall, 1997.
[52] Scott, P.D. and E. Wilkins, “Evaluating data mining procedures: techniques fr generation artificial data sets”, Information and software technology, 41: 579-587, 1999.
[53] Breiman *et al*, “Classification and Regression Trees”, Belmont, CA: Wadsworth Int’l Group, 1984.
[54] Kass, G. V., “An Exploratory Technique for Investigating Large Quantities of Categorical Data”, Applied Statistics, Vol. 29, p. 119-127, 1980.
[55] Quinlan J. R., “Induction of Decision Trees”, Machine Learning, 1, pp. 81-106, 1986.
[56] Michael L, Gargano and Bel G, Raggad, “ Data Mining – a powerful information creating tool” OCLC Systems & Services Volume 15, Number 2 1999, pp 81-90 © MCB University Press, [ISSN 1065-075X]
[57] Zurada J. M., “Artificial Neural Systems”. St. Paul, MN: West Publishing, 1992
[58] Maureen Caudill and Charles Butler, “Understanding Neural Networks” Vol 1, Basic Networks. The MIT Press, 1992, ©Massachusetts Institute of Technology, [ISBN: 0262530996]
[59] Lu, H., Setiono. R, and Liu, H (1996), “Effective Data Mining Using Neural Networks”, IEEE Transactions on Knowledge and Data Engineering, Vol. 8 No. 6 pp 957-961.
[60] Goldberg D.E., “Genetic Algorithms in Search, Optimization and Machine Learning”. Reading, MA: Addison-Wesley, 1989.
[61] Gene Miner, “GenoFuzzy Computing”, Technical Report TR-2001-408 Copyright ©1993-2002, KDiscovery.com
[62] Darrin J. Marshall, “Data Mining using Genetic Algorithms”, Industrial Applications of Genetic Algorithms, Edited by Charles L. Karr, L. Michael Freeman, CRC Press LLC, [ISBN: 0-8493-9801-0] pp.159-195
[63] http://www.datamining.com/pwhs.htm
[64] http://www.thearling.com/text/dmviz/modelviz.htm

2/12/2012