A. Kusiak, Data Mining in Design of Products and Production Systems, *Proceedings of INCOM'2006: 12th IFAC/IFIP/IFORS/IEEE Symposium on Control Problems in Manufacturing*, May 2006, Saint-Etienne, France, Vol. 1, pp. 49-53.

DATA MINING IN DESIGN OF PRODUCTS AND PRODUCTION SYSTEMS

Andrew Kusiak

Intelligent Systems Laboratory
Industrial Engineering
3131 Seamans Center
The University of Iowa
Iowa City, Iowa 52242 - 1527
Tel: 319 - 335 5934   Fax: 319-335 5669
andrew-kusiak@uiowa.edu
http://www.icaen.uiowa.edu/~ankusiak

Abstract: As a new science, data mining is acquiring its own identity by refining the concepts from other disciplines as well as by entering new application areas. Designing products and manufacturing systems is increasingly becoming affected by the data mining pursuit. The paper outlines areas of product and manufacturing system design with a potential for data mining applications. The key challenges of data mining in the application domain discussed in the paper are summarized.

Keywords: Data mining, data analysis, product design, manufacturing, production systems.

A. Kusiak, Data Mining in Design of Products and Production Systems, *Proceedings of INCOM'2006: 12th IFAC/IFIP/IFORS/IEEE Symposium on Control Problems in Manufacturing*, May 2006, Saint-Etienne, France, Vol. 1, pp. 49-53.

1. INTRODUCTION

We are witnessing significant changes taking place in the way corporations conduct business. Many of these changes are direct results of greater use of data in design and manufacturing.

The flow of data and information in a traditional design and manufacturing system is essentially unidirectional (see Fig. 1).

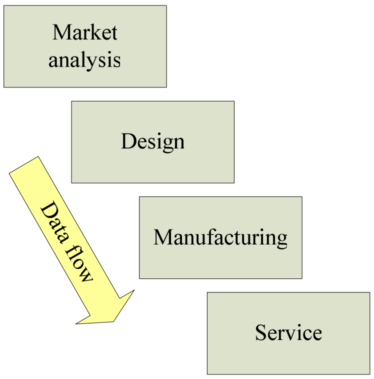

Fig. 1. Data flow in traditional design and manufacturing systems.

Any local bidirectional flow (loops) of information has been often attributed to imperfections of the system, e.g., design negotiation, manufacturing errors.

The developments in networking, data storage, and data mining tools have contributed to the emergence the closed loop system illustrated in Fig. 2.

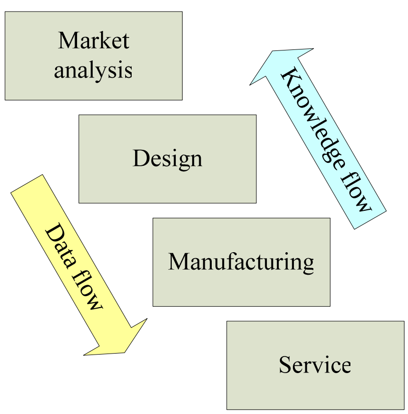

Fig. 2. Data and knowledge flow in a modern design and manufacturing system.

The activities around material, parts, and products across the market analysis, design engineering, manufacturing, and service cycle create a data trail. This data is of growing importance in modern systems. Data mining algorithms extract knowledge from this large volume of data leading to significant improvements in the next generation of products and services. In fact, the knowledge discovery activity could become the key factor to the innovation and business success.

2. KNOWLEDGE DISCOVERY

Data-mining algorithms extract knowledge from data that may be used for different applications. There are two general classes of data mining, descriptive and predictive. The goal of descriptive data mining is to discover patters, e.g., product configurations formed in mass customization applications. The predictive data mining aims at building models to determine (predict) outcomes, e.g., stock level. Since the width of data analyzed by the data-mining algorithms is essentially unlimited, the discovered patterns could be unanticipated and interesting. The value delivered by these patterns is related to the quality of data and textual databases. Besides the comprehensiveness of data processing, data mining brings yet another advantage – it supports the needs of an individual object, e.g., a part or a customer.

Data mining techniques have been successfully deployed in engineering, medical, and business applications. Design and manufacturing is a natural candidate for data mining applications by the virtue of large volumes of data. Besides enhancing innovation, data mining methods can reduce the risks associated with conducting business and shorten the response time of businesses.

Data mining methods provide an integrated environment for decision support utilizing classification, knowledge extraction, reporting, and visualization. Inria (2000) discussed the following intelligent system aspects:

- Technological intelligence, to know about an existing or an emerging technology
- Competitive intelligence, to know about activities, products, or services of competitors or other actors in the enterprise market
- Commercial intelligence, to know about the enterprise commercial environment such as distributors, suppliers, and customers
- Strategic intelligence, to support the enterprise mangers' strategic decisions.

Some of the most commonly used data mining techniques are (Kantardzic 2003):

- Clustering
- Decision trees
- Fuzzy sets
- Neural networks
- Bayesian methods
- Regression

The goal of data mining may range from obtaining a general understanding of the nature of data to very accurate modelling and prediction. For example:

- Data description and summarization: Description of data characteristics, typically in elementary and aggregated form.

- Segmentation: Separation of data into interesting and meaningful subgroups or classes.
- Concept description: Description of concepts or classes in an understandable form.
- Dependency analysis: Finding a model that describes significant dependencies between objects or events.
- Classification: Building classification models that assign the correct class to previously unseen and unlabeled objects.

For data mining to be effective, several technologies have to work together. Data mining algorithms extract patterns from data to create a meaning from data that would be otherwise useless. Visualization techniques provide visual understanding of data, rules, patterns and trends. Data warehousing is a critical technology for organizing, cleaning and preparing data for mining. The computer network infrastructure is important, especially for distributed data mining. These technologies need to be integrated for effective data mining.
Some of the application areas of data mining (DM) in design and manufacturing are discussed next.

3. DM APPLICATIONS DOMAINS

3.1 Knowledge-derived innovation

The recent years has brought renewed interest in innovation, especially after the Innovate America Report (NIIR 2004), has been published. Though innovation has been a subject of intensive studies by diverse research communities, many will agree that the produced results have not translated into meaningful innovation gains in the industry. Rather industry is awaiting methodologies, processes, and tools leading to innovation breakthroughs.
It appears that the product-life cycle data is of importance to innovation. Based on the data source, two types of innovation can be defined:

- Customer driven innovation
- Domain expert driven innovation

The data provided by the customers and domain experts will ultimately be transformed into product design requirements. Data mining algorithms will likely sift through the vast content of data warehouses to produce innovation related requirements. Evolutionary computation techniques, in particular genetic programming (De Jong 2006), may further enhance the value of the data and knowledge.

3.2 Mass customization

Mass customization is defined as permitting "customized manufacture on a mass basis" (Davis 1989). According to Da Silveira *et al.* (2001), there are three main ideas justifying the use of mass customization. The first is the advent of flexible manufacturing and information technologies that enable production systems to deliver higher variety of products at lower costs. The second idea is the fact that consumers are constantly increasing their need for product variety and customization. Finally, the shortening of product life cycles and expanding industrial competition has led to the shift away from mass production, increasing the need for production strategies focused on individual customers.
To realize benefits of mass customization companies seek new production strategies (Agard and Kusiak 2004). The two traditional production strategies are make-to-stock and make-to-order. While the former strategy results in excessive inventory and implies a low pressure on the process set-up reduction, the latter leads to low inventory levels and calls for the process set-up reduction.

The assemble-to-order strategy offers a compromise between the two traditional production strategies and is supports the mass customization concept. To illustrate this strategy consider five sale records of a simple tractor (see Fig. 3).

	Frame	Engine	Cabin	Wheels	Backhoe
Customer 1	Large	Medium	Green	Large	No
Customer 2	X Large	X Large	Pink	Large	No
Customer 3	Small	Meduim	Yellow	Medium	Yes
Customer 4	Large	Large	Red	Large	Yes
Customer 5	Small	Small	Yellow	Medium	No

Fig. 3. Sales records of a simple tractor.

Clustering the data in Fig. 3 produces the matrix in Fig. 4. The four subassemblies formed S1 through S4 are used to realize the assembly-to-order strategy.

	Frame	Wheels		Engine	Cabin	Backhoe
Customer 4	Large	Large		Large	Red	Yes
Customer 1	Large	Large		Medium	Green	No
Customer 3	Small	Medium		Medium	Yellow	Yes
Customer 5	Small	Medium		Small	Yellow	No
Customer 2	X Large	Large		X Large	Pink	No

S1 S2 S3 S4

Fig. 4. Clustered rows and columns of Fig. 3.

An approach followed by some companies, short of the assembly-to-order strategy, aims at developing preassembled configurations at attractive prices. Grouping the rows (customers) of the matrix in Fig. 3 has resulted in the matrix of Fig. 5.

		Frame	Engine	Cabin	Wheels	Backhoe
P1	Customer 4	Large	Large	Red	Large	Yes
	Customer 1	Large	Medium	Green	Large	No
P2	Customer 3	Small	Medium	Yellow	Medium	Yes
	Customer 5	Small	Small	Yellow	Medium	No
	Customer 2	X Large	X Large	Pink	Large	No

Fig. 5. Clustered rows of Fig. 3.

The first two rows in Fig. 5 are labelled P1, the next two P2, and the last row remains unlabeled. The configurations P1 and P2 could be further transformed by offering 2 engine upgrades, 1 cabin downgrade, and 2 backhoe upgrades to the configurations shown in Fig. 6.

		Frame	Engine	Cabin	Wheels	Backhoe
P1	Customer 4	Large	Large	Green-	Large	Yes
	Customer 1	Large	Large+	Green	Large	Yes+
P2	Customer 3	Small	Medium	Yellow	Medium	Yes
	Customer 5	Small	Medium+	Yellow	Medium	Yes+
	Customer 2	X Large	X Large	Pink	Large	No

Fig. 6. Transformed data of Fig. 5.

3.3 Supply chain management

A supply chain is a contractual linkage among various parties, ideally to achieve a "just-in-time" flow of supplied goods. The purpose of the supply chain designer is to quickly generate the electronic trade scenarios. Supply chain management involves the adoption of electronic linkages between two businesses that are related as supplier/customer within a single industry channel or supply chain (Westland and Clark 1999).

Data mining is a powerful tool for supply chain management, especially in the e-commerce environment. It can be used in the following ways:

- Reduce the level of risk to the business. Most of the payments in e-commerce are through credit cards. Checking the customer's credit history is a very important measure to reduce business risk. In the traditional supply chain, it is very time consuming and error-prone work. Imagine companies that receive thousands of orders per-day. Each processing clerk can only spend a very short amount of time on each case to make decisions. The quality of the decision depends on his/her previous experience and intuition because there is not enough time to analyze all relevant data. Using data mining techniques, we can find very useful patterns to support decisions. The rules are much easier for humans to understand than the rough data since the rules are extracted from large, otherwise incomprehensible data sets. Decisions based upon the extracted rules will be more reliable.
- Control the inventory. For example, in retail business, inventory is very expensive and represents a large liability. Knowledge of data mining output can analyze past business, monitor present transactions, and predict future sales. With better control of inventory, the retailer can achieve higher profitability. The same concept applies to distributors and manufacturers.
- Prediction of the customer's behaviour to control the inventory. When customers buy products online, they always want to receive the goods as soon as possible. In most cases, the manufacturing process takes time to produce goods. To meet the customer's requirements, certain inventory levels have to be kept. If we have better predictions of customers' behaviour patters, we can have better balance between inventory levels and customers' needs and increase profitability.
- Customer Relationship Management (CRM) is another hot research issue in recent supply chain research. The CRM should integrate customer data with different resources. It should also provide a deep understanding of customer behaviour and needs. Data mining approaches are very attractive for CRM. Data mining methods can generate out decision tree and cluster rules for CRM.

3.4 Pattern discovery

Important patterns might be hidden in the volumes of industrial data. For example, data mining applied inn the customer domain may reveal such patterns that may provide answer to questions such as:

- What characterizes frequent buyers?
- What characterizes customers that are keen on promotions?
- What characterizes customers making quick purchase decisions?
- What characterizes customers that do not purchase?

Most database systems, such as MS-Access and Oracle provide capabilities providing answers to some of these higher-level questions. However, for in-depth analysis, data mining algorithms are needed (Kusiak 2005).

3.5 Trends detection

Industrial companies are increasingly developing data warehouses to collect business data. Data mining algorithms not only can extract the static patterns in data, but can also discover dynamic trends. Mining time series is an active research area (Kusiak *et al.* 2005). The trends reflect customer interest shifts, technology development, and the response to marketing strategies.

3.6 Data dimensionality reduction

Modern databases may contain very large numbers of rows (transactions) and columns (features). The large size of the database makes it very difficult to retrieve useful information from it. One important research interest in the pre-processing of data for mining is the concept of dimensionality reduction. Unrelated data items and features can be eliminated from the dataset to reduce the data mining effort.

3.7 Visualization

Visualization tools, such as format, shape, colour, font, and 2D/3D drawing, enhance humans' understanding of the data and relationships. For example, graphs, charts, and tables make information easier to understand than the original data. The relationships between different data items become obvious when they are displayed. To make full use of the data mining potential, a user-friendly information interface needs to be developed. The interface needs to contain a technical component capable of easily receiving information and a component that makes it easily understood by the user. There is a great demand for high quality visualization techniques in industrial applications.

- Legal concerns. Knowledge extracted by data mining algorithms may involve sensitive information that opens door to legal concerns. This problem needs to be solved from technological, legal, and social aspects.

4. CONCLUSION

Although numerous successful applications of data mining applications in design and manufacturing have been reported, many challenges are ahead. Some challenges come from data mining itself and others come from the application domain. The author of reference [URL] discussed some of challenges coming from data mining. The key challenges are summarized as follows:

- Make data mining models comprehensible to users. Users are not data mining experts. A good presentation format is crucial for full utilization of data mining results. Due to the variety of formats, the data mining result can be expressed in different ways. For example, the decision tree is a good method of presenting classification results.
- Scalability of data mining algorithms. Most of the data mining algorithms assume that the data can fit into available memory. It is quite common to encounter multi-gigabyte databases. Developments in data mining algorithms are needed to deal with large databases.
- Support for unstructured data. Most existing data mining algorithms handle only numeric or textual data. Design and manufacturing systems could provide other data types, such as image and video. Dealing with the unstructured data is a challenging problem for data mining research.
- Integrating data mining algorithms with other applications to improve their respective performance.
- Distributed data mining algorithms. Many data bases are distributed. However, most of the data mining algorithms are not designed for a distributed environment.

REFERENCES

Agard, B. and Kusiak, A. (2004), A data-mining based methodology for the design of product families, *International Journal of Production Research*, Vol. **42**, No. 15, pp. 2955-2969.

Davis, S. (1989). From "Future Perfect": Mass customizing, *Planning Review*, Vol. **17**, No. 2, pp. 16-21.

Da Silveira, G., Borenstein, D., and Fogliatto, F. (2001). Mass customization: Literature review and research directions, *International Journal of Production Economics*, Vol. **72**, No. 1, pp. 1-13.

De Jong, K.A. (20036), Evolutionary Computation: A Unified Approach, MIT Press, Cambridge, MA.

Inria, R.D. (2000), Knowledge management and the Internet, *Intelligent Systems*, May/June.

Kantardzic, M. (2003), *Data Mining: Concepts Models, and Algorithms*, IEEE Press and John Wiley, New York.

Kusiak, A. (2005), Selection of invariant objects with a data mining approach, *IEEE Transactions on Electronics Packaging Manufacturing*, Vol. **28**, No. 2, pp. 187-196.

Kusiak, A., Burns, A, Shah, S., and Novotny, N. (2005), Detection of events causing pluggage of a coal-fired boiler: A data mining approach, *Combustion Science and Technology*, Vol. **177**, No. 12, pp. 2327-2348

NIIR (2004), Innovate America, Council for Competitiveness, National Innovation Initiative Report, Washington, DC.

Westland, J.C. and Clark, T.H.K. (1999), *Global Electronic Commerce: Theory and Case Studies*, MIT Press, Cambridge, MA.

URL: http://www.ncdm.uic.edu/m3d-finalreport.htm.