MASSIVE COMPUTING

DATA MINING FOR DESIGN AND MANUFACTURING

Methods and Applications

Dan Braha (Ed.)

KLUWER ACADEMIC PUBLISHERS

Data Mining for Design and Manufacturing

Methods and Applications

Edited by

Dan Braha
Ben-Gurion University

KLUWER ACADEMIC PUBLISHERS
DORDRECHT / BOSTON / LONDON

1. Data mining

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 1-4020-0034-0

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Sold and distributed in North, Central and South America
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

Printed on acid-free paper

Printed in the Netherlands.

DATA MINING FOR DESIGN AND MANUFACTURING

Data Mining for Design and Manufacturing

The productivity of individual companies as well as the efficiency of the global economy can be dramatically affected by Engineering Design and Manufacturing decisions and processes. Powerful data acquisition systems (such as minicomputers, microprocessors, transducers, and analog-to-digital converters) that collect, analyze, and transfer data are in use in virtually all mid-range and large companies. Over time, more and more current, detailed, and accurate data are accumulated and stored in databases at various stages of design and production. This data may be related to designs, products, machines, materials, processes, inventories, sales, marketing, and performance data and may include patterns, trends, associations, and dependencies. There is valuable information in the data. For instance, understanding the data and the quantitative relationships among product design, product geometry and materials, manufacturing process, equipment capabilities, and related activities could be considered strategic information. Extracting, organizing, and analyzing such useful information could be utilized to improve and optimize company planning and operations.

The large amount of data, which was generated and collected during daily operations and which contain hundreds of attributes, needs to be simultaneously considered in order to accurately model the system's behavior. It is the abundance of data, however, that has impeded the ability to extract useful knowledge. Moreover, the large amount of data in many design and manufacturing databases make it impractical to manually analyze for valuable decision-making information. This complexity calls for new techniques and tools that can intelligently and (semi)automatically turn low-level data into high-level and useful knowledge.

The need for automated analysis and discovery tools for extracting useful knowledge from huge amounts of raw data suggests that Knowledge Discovery in Databases (KDD) and Data Mining methodologies may become extremely important tools in realizing the above objectives. Data mining is primarily used in business retail. Applications to design and manufacturing are still under utilized and infrequently used on a large scale. Data Mining is often defined as the process of extracting valid, previously unknown, comprehensible information from large databases in order to improve and optimize business decisions[1]. Some researchers use the term KDD to denote the entire process of turning low-level data into high-level knowledge.

---

[1] Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (Eds.), Advances in Knowledge Discovery and Data Mining. Cambridge, MA: AAAI Press/MIT Press, 1996.

Although data mining algorithms are at the core of the data mining process, they constitute just one step that usually takes about 15% to 25% of the overall effort in the overall data mining process. A collaborative effort of domain expert(s) (e.g., designer, production manager), data expert(s) (e.g., IT professionals) and data mining expert(s) is essential to the success of the data mining integration within design and manufacturing environments. A successful implementation of the data mining process often includes the following important stages[1]. The first step involves understanding the application domain to which the data mining is applied and the goals and tasks of the data mining process; e.g., understanding the factors that might affect the yield of a Silicon wafer in the semiconductor industry. The second step includes selecting, integrating, and checking the target data set that may be stored in various databases and computer-aided systems (such as CAD, CAM, MRP or ERP). The target data set may be defined in terms of the records as well as the attributes of interest to the decision-maker. The third step is data preprocessing. This includes data transformation, handling missing or unknown values, and data cleaning. In the fourth step, data mining takes place for extracting patterns from data. This involves model and hypothesis development, selection of appropriate data mining algorithms, and extraction of desired data. In the fifth step, the extracted patterns are interpreted and presented in a user-readable manner; e.g., using visualization techniques, and the results are tested and verified. Finally, the discovered knowledge may be used and a knowledge maintenance mechanism can be set up. The data mining process may be refined and some of its steps may be iterated several times before the extracted knowledge can be used for productive decision making.

Data Mining techniques are at the core of the data mining process, and can have different goals depending on the intended outcome of the overall data mining process. Most data mining goals fall under the following main categories[1]:

- **Data Processing** is concerned with selection, integration, filtration, sampling, cleaning and/or transformation of data.
- **Verification** focuses mainly on testing preconceived hypotheses (generated by the decision-maker) and on fitting models to data.
- **Regression** is concerned with the analysis of the relationships between attribute values within the same record, and the automatic production of a model that can predict attribute values for future records. For example, multivariate linear regression analysis may be used to identify the most significant factors affecting process capability.
- **Classification** involves assigning a record to predetermined classes. For example, wafer-cleaning processes (that include parameters set at various levels) may be classified according to the quality of the cleaning process outcome; thus, the outcome of new cleaning processes can be identified.
- **Clustering** focuses on partitioning a set of data items with similar characteristics together. For example, identifying subgroups of silicon

wafers that have a similar yield.

- ♦ **Association** (link analysis) involves the discovery of rules that associate one set of attributes in a data set to other attribute sets. An example for a relation between attributes of the same item is "if the Lot Size > 100 $\wedge$ Tool Change = Manual $\rightarrow$ (Inventory = High)."
- ♦ **Sequential Pattern Analysis** is concerned with the discovery of causal relationships among temporally oriented events. For example, the event of environmental attack and high service stresses can lead to a stress-corrosion breaking within the next 2 hours.
- ♦ **Model Visualization** focuses on the decision makers' attempt to convey the discovered knowledge in an understandable and interpretable manner. Examples include histograms, scatter plots, and outliers identification.
- ♦ **Deviation Analysis** is used to detect deviation over time, deviation from the mean, and deviation between an observed value and a reference value as applied, for instance, in quality control.

A variety of techniques are available to enable the above goals[2]. Different data mining techniques serve different purposes, each offering its own advantages and disadvantages. The most commonly used techniques can be categorized in the following groups: Statistical methods, Artificial Neural Networks, Decision Trees, Rule Induction, Case-Based Reasoning, Bayesian Belief Networks, and Genetic Algorithms and Evolutionary Programming. An introductory overview of data mining is provided in Chapters 1 and 2.

Several techniques with different goals can be applied successively to achieve a desired result. For example, in order to identify the attributes that are significant in a photolithography process, clustering can be used first to segment the wafer-test database into a given predefined number of categorical classes, then classification can be used to determine to which group a new data item belongs.

Over the last decade, data mining mechanisms have been applied in various organizations, and have led to a wide range of research and development activities. It is primarily used today by companies with a strong customer focus such as in retail, insurance, finance, banking, communication, and direct marketing. Although data mining is widely used in many such organizations, the interest in data mining reveals an astute awareness among manufacturing companies across many industry sectors regarding the potential of data mining for changing business performance. For example, data mining techniques have been used by Texas Instruments (fault diagnosis), Caterpillar (effluent quality control and warranty claims analysis), Ford (harshness, noise, and vibration analysis), Motorola (CDMA Base Station Placement), Boeing (Post-flight Diagnostics), and Kodak (data visualization). Still, the application

[2] Berry, M. J., and Linoff, G., Data Mining Techniques. New York: John Wiley & Sons, 1997.

of data mining to design and manufacturing is not broadly integrated within companies. Decision makers are hampered from fully exploiting data mining techniques by the complexity of broad-based integration. The objective of this book is to help clarify the potential integration and to present a wide range of possibilities that are available by bringing together the latest research and application of data mining within design and manufacturing environments. In addition, the book demonstrates the essential need for the symbiotic collaboration of expertise in design and manufacturing, data mining and information technology.

Data Mining in Product Design and Development

The integration of data mining to design and manufacturing should be based on goals and capabilities of data mining as well as goals and weaknesses of current design and manufacturing environments. To broaden our understanding of how data mining can overcome a variety of problems in design and manufacturing we consider a wide range of activities within manufacturing companies. The first important activity is the product design and development process. A product development process is the sequence of activities that a manufacturing company employs in order to turn opportunities and ideas into successful products. As can be seen in Figure 1, product development goes through several stages, starting with identifying customer needs and ending with production and then delivery to market[3].

The nature of the product development process can be viewed as a sequential process. The design process evolves from concept through realization. For instance, a part cannot be assembled until the components are machined; the components cannot be machined until the NC code is created; the NC code cannot be created without a dimensioned part model; the part model cannot be dimensioned without a set of requirements and a general notion of what the part looks like; and presumably the last two items come from a need that must first be identified. All this points to the seemingly undeniable truth that there is an inherent, sequential order to most design processes. One can reason equally effectively, however, that product development is an iterative process. First, product development teams are only human and have a bounded rationality. They cannot simultaneously consider every relevant aspect of any given product design. As the product development process progresses, new information, ideas, and technologies become available that require modifying the product design. Second, design systems are limited; there is no known system that can directly input a set of requirements and yield the optimum design. Rather, the designer must

[3] Braha, D, and Maimon, O., A Mathematical Theory of Design: Foundations, Algorithms and Applications. Boston, MA: Kluwer Academic Publishers, 1998.

iteratively break down the set of requirements into dimensions, constraints, and features and then test the resulting design to see if the remaining requirements were satisfied. Finally, the real world often responds differently than is imagined. The real world is full of chaotic reactions that are only superficially modeled in any design system. All this points to the seemingly undeniable truth that there is an inherent, iterative nature to the product development process[3].

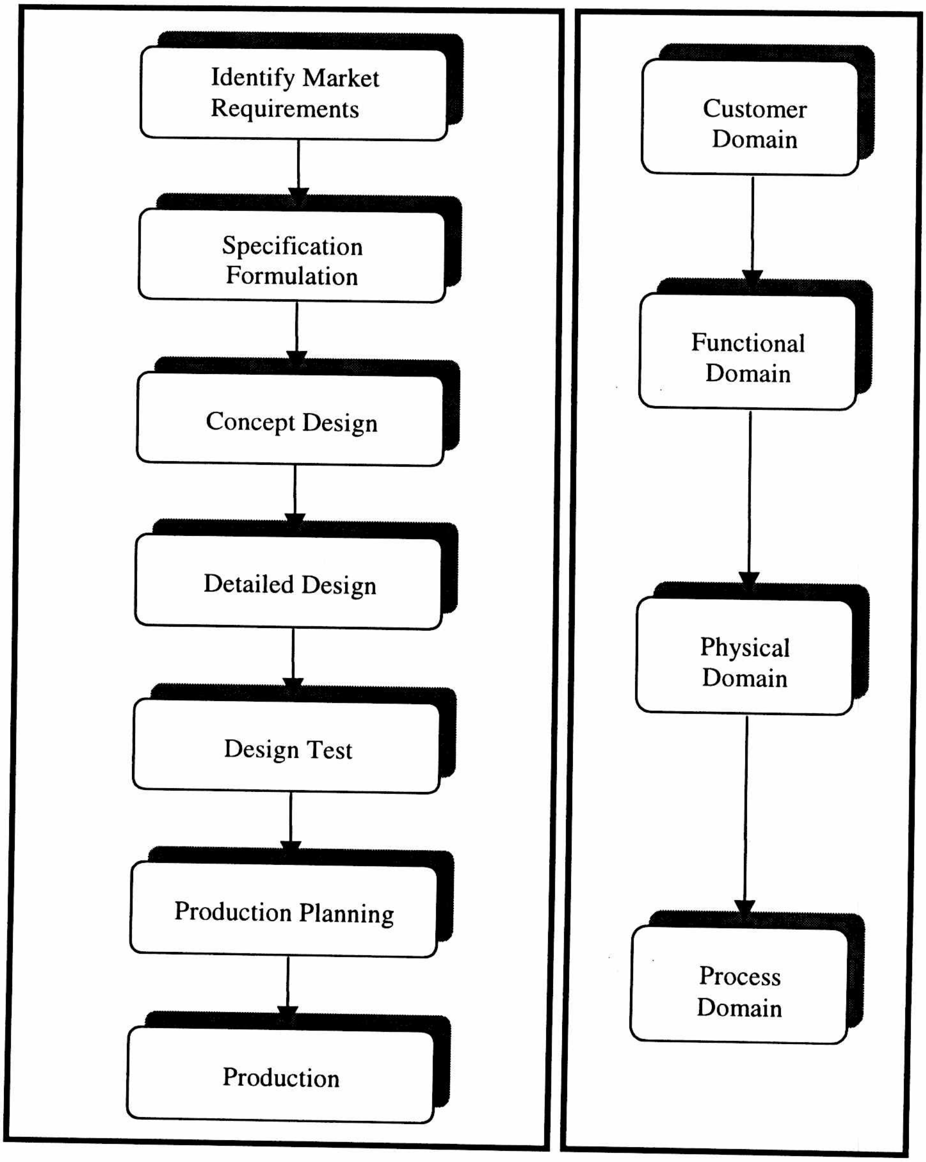

Figure 1 Traditional View of Product Design and Development

In order to reconcile these two disparate visions of the product development process, we categorize product development iteration is categorized as occurring either *between* stages (*inter-stage* iteration) or *within* a stage (*intra-stage* iteration)[3]. In this model (see Figure 2), product development still flows sequentially from initial concept through realization, each process stage providing the data and requirements for the subsequent stage. Within each process stage, however, the designer iteratively creates a design that meets the given requirements.

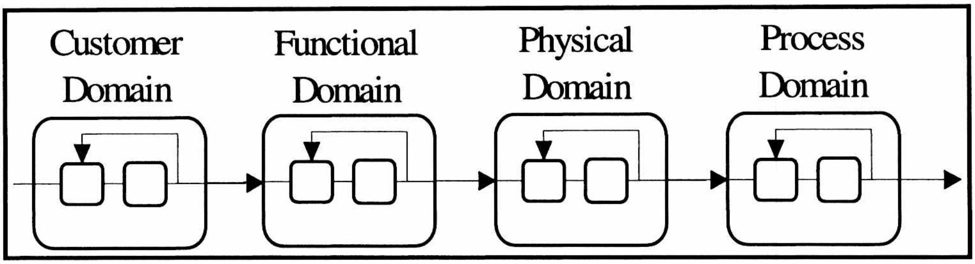

Figure 2 *Combining Sequential and Iterative Design*

This model largely represents the current state-of-the-art in CAD/CAM/CAE systems. While there are numerous software modules to assist the designer during intra-stage design iteration (e.g., QFD software to help identify customer needs and CAE software to analyze a current design), the tools are generally not well integrated at the inter-stage level. It has been recognized that minimizing the inter-stage and intra-stage iterations tends to reduce product development lead times and cost. Data mining has the potential of becoming one of the key components in achieving these goals.

During the product design and development process data mining can be used in order to determine relationships among "internal" factors at each stage and "external" factors at consecutive and previous stages. Following are some examples of how data mining can be utilized.

- ♦ Data mining can be used to extract patterns from customer needs, to learn interrelationships between customer needs and design specifications, and to group products based on functional similarity for the purpose of benchmarking, modular design, and mass customization.
- ♦ At the concept design stage, data mining can support concept selection by dynamic indexing and retrieval of design information in knowledge bases (e.g., patents and benchmarking products), clustering of design cases for design reuse, extracting design knowledge for supporting knowledge-based systems, extracting guidelines and rules for design-for-X (manufacturability, assembly, economics, environment), and exploring interactively conceptual designs by visualizing relationships in large product development databases.

- During system-level design, data mining can aid in extracting the relationships between product architecture, product portfolio, and customer needs data.
- At the detailed design stage, data mining can support material selection and cost evaluation systems.
- In industrial design, information about the complex relationships between tangible product features (such as color and ergonomics) and intangible aspects of a product that relate to the user (such as aesthetics, comfort, enthusiasm, and feelings) can be extracted with data mining and used in redesign.
- When testing the design, product characteristics can be extracted from prototypes. This may be used for determining the best design practice (e.g., design for reuse).
- During product development planning, data mining can be beneficial to activities such as the prediction of product development span time and cost, effectiveness of cross-functional teams, and exploration of tradeoffs between overlapping activities and coordination costs. Data mining may also be used for identifying dependencies among design tasks, which can be used to develop an effective product development plan.
- On an organizational level, data mining can be seen as a supportive vehicle for organizational learning, e.g., based on past projects, the factors (and their interdependencies) that affect a project's success/failure may be identified. This may include the intricate interaction between the project and the company, market, and macro environment.

In summary, the utilization of data mining in understanding the relationships among internal and external factors facilitates the inter-stage and intra-stage iterations. This will lead to improved product design and development.

Data Mining in Manufacturing

At the end of the product design and development process, after a stage of production ramp-up, the ongoing manufacturing system operation begins. There are a wide range of domains within manufacturing environments to which data mining techniques can be applied. On an abstract level, data mining can be seen as a supportive vehicle for determining causal relationships among "internal" factors associated with the manufacturing processes and "external" elements related to the competitiveness of the manufacturing company (e.g., production indices, performance parameters, yield, company goals). Since manufacturing environments have an inherently temporal or spatial context, time and/or space factors may be taken into account in the mining process in order to correctly interpret the collected data (e.g., from a certain production date the number of defects is much higher than normal). For instance, the process of wafer fabrication is a series of 16-24 loops, each adding a layer to the device. Each loop comprises some or all of the major steps of photolithography, etching, stripping, diffusion, ion implantation, deposition, and chemical mechanical planarization. At each stage, there are various inspections and measurements performed to monitor the equipment and process. The large number of parameters that are measured after each operation could identify causal interrelationships between processing steps and various test data. This information could then be utilized to eliminate faults during the manufacturing process, and thus enhance yield levels.

Additional examples where data mining can be applied successfully include:

- fault diagnosis such as predicting assembly errors and defects, which may be used to improve the performance of the manufacturing quality control activity;
- preventive machine maintenance, which is concerned with deciding the point in time and type of maintenance of tools and instruments. For instance, cutting tool-state may be classified and used for tool condition monitoring;
- manufacturing knowledge acquisition by examining relevant data, which implicitly contains most of the required expert knowledge. The extracted knowledge rules can then be incorporated by expert systems for decision support such as fuzzy controllers, diagnosis systems, and intelligent scheduling systems;
- operational manufacturing control such as intelligent scheduling systems that learn the effect of local dynamic behavior on global outcomes, and use the extracted knowledge to generate control policies. These operational systems are inherently adaptive, since data that is accumulated in real-time can change the baseline policies generated by the data mining algorithms;
- learning in the context of robotics (e.g., navigation and exploration, mapping, feature recognition, and extracting knowledge from numerical and graphical sensor data;
- quality and process control, which is concerned with monitoring standards; taking measurements; and taking corrective actions in case deviation from the norm is detected and/or discernible patterns of data over time are present. Extracted knowledge may include classification to predetermined types of deviation from the norm, and causal relationships among temporally oriented events;
- adaptive human-machine interface for machine operation;
- summarization and abstraction of large and high-dimensional manufacturing data;

- enabling supply and delivery forecasting, e.g., by classifying types of suppliers involved in transportation and distribution of the product.

Enabling Technology for Data Mining

In order too successfully implement data mining to design and manufacturing environments, several key issues such as the selection, integration, cleansing, and preparation of data should be addressed. Thus, enabling or supportive technologies that help carry out these processes are valuable.

- One of the most important supportive technologies is data warehousing, which is defined as the process of integrating legacy operational systems (storing data related to product, process, assembly, inventory, purchasing, etc.) within a company to provide centralized data management and retrieval for decision support purposes. Thus, the preprocessing, including cleaning and transforming data with the intent of analysis and discovery, can be facilitated by a data warehouse.
- Report generators, which are used to present the extracted patterns in a user-readable way, are another type of supportive technology. If discovered knowledge is further used by various computers (e.g., CNC systems, industrial robots, etc.), it is imperative that computers be able to interpret the output.
- Computationally massive data mining operations can be enabled through parallel-computing platforms and distributed computation. The parallelism aspect is especially important when data mining is deployed proactively and systematically throughout the manufacturing environment, and is used for continuous tasks such as preventive machine maintenance and real-time monitoring of the overall manufacturing processes.

Data Mining for Design and Manufacturing: Methods and Applications brings together for the first time the research and application of data mining within design and manufacturing environments. The contributors include researchers and practitioners from academia and industry. The book provides an explanation of how data mining technology can be exploited beyond prediction and modeling, and how to overcome several central problems in design and manufacturing environments. The book also presents the formal tools required to extract valuable information from design and manufacturing data (e.g., patterns, trends, associations, and dependencies), and thus facilitates interdisciplinary problem solving and optimizes design and manufacturing decisions.

The book includes aspects of topics such as: data warehouses and marts, data mining process, data mining tasks (e.g., association, clustering, classification), data mining methods (e.g., decision trees and rules, neural networks, self-organizing feature maps, wavelets, fuzzy learning, and case-based reasoning), machine learning in design (e.g., knowledge acquisition, learning in analogical design, conceptual design, and learning for design reuse), data mining for product development and concurrent engineering, design and manufacturing warehousing, Computer-aided Design (CAD) and data mining, data mining for Computer-aided Process Planning (CAPP), data mining for Material Requirements Planning (MRP), manufacturing data management, process and quality control, process analysis, data representation/visualization, fault diagnosis, adaptive schedulers, and learning in robotics.

Chapters are arranged in four sections: Overview of Data Mining; Data Mining in Product Design; Data Mining in Manufacturing; and Enabling Technologies for Data Mining in design and manufacturing.

Dan Braha
2001

Acknowledgements

I wish to acknowledge the authors for their fine contributions and cooperation during the book preparation. Chapters underwent a rigorous selection process. Many thanks to the reviewers. I appreciate the friendly assistance of Armin Shmilovici and Arieh Gavious. The technical support and suggestions of Sageet Braha were invaluable. I am grateful to John Martindale of Kluwer Academic Publishers for his forthcoming support of this important endeavor. Finally, to little Neoreet who provided entertainment along the way.

PREFACE

PART I: OVERVIEW OF DATA MINING

PART II: DATA MINING IN PRODUCT DESIGN

PART III: DATA MINING IN MANUFACTURING