The Data Mining Technology Based on CIMS and Its Application on Automotive Remanufacturing

Ke Chen
School of Mechanical Engineering
University of Tianjin
Tianjin, China
chenke@tju.edu.cn

Jia Wang
Auto Industry Policy Research section
China Automotive Technology &Research Center
Tianjin, China
wangjia.tju@gmail.com

Abstract

Nowadays, with the development of computer technology, data mining has been widely used in various fields. This paper describes a CIMSMINER that combines the data mining with CIMS (Computer Integrated Manufacturing System) and instructs its objectives, model, physical architecture and methods. Considering the characteristics of remanufacturing of automotive products in China, the CIMSMINER is used to get the information concourse together and obtain the data mining results to help the improvement of products. The application in automotive remanufacturing is a reform to make the automotive product information chain be not only an information carrier, but also an information miner. Currently, the government strongly emphasizes energy saving and emission reducing, which is closely related to the sustainable development of China. Obviously, CIMSMINER is an effective tool to support the implementation of this policy.

1. Introduction

As an emerging multi-interdisciplinary field and along with the constant upgrading of computer technology, data mining has developed into a powerful tool. The Intelligent Miner developed by IBM Company has abilities of automatic generation, association analysis, sequence discovery and visible view of dataset, etc. Enterprise Miner put forward by SAS Company is a general platform for data mining. This system adopts the "sampling-transforming-modeling-evaluating-strategy" method. Stanford University has brought forward MLC++ model, while Washington University has brought forward Brute model. Moreover, Korean National Intelligent Institute provides many algorithms, and Israel has already applied data mining techniques to the clinic medicine studies.

Besides the above data mining platform, the application of data mining still has other forms. The data mining based on CIMS is one of them. Though the application of CIMS in China is gradually wider, a large amount of data is only stored without effective utilization, resulting in a great waste of enterprise information resources. CIMSMINER that uses data mining techniques based on CIMS can resolves this problem. It can explore the implied information of data from CIMS databases and plays a great significant role in the process of product manufacturing, product recycling and new product development. CIMSMINER improves the ability of remanufacturing for enterprises, which helps the implementation of relevant laws and regulations about automotive products remanufacturing in China.

2. CIMSMINER

As a synthetical technology, CIMS (Computer Integrated Manufacturing System) plays an important role in enhancing the competitiveness of China's enterprises. It provides a new modern enterprise management model and advanced manufacturing technology. CIMS is integrated by MIS (Management Information Systems), MRPII (Manufacturing Resource Planning), CAD （Computer-Aided Design）, CAPP （Computer-Aided Process Design System) and many other subsystems. CIMSMINER is a model based on CIMS database, data warehousing and a variety of data mining algorithm [1].

2.1. Objectives of System

e-Forensics 2008, January 21-23, 2008, Adelaide, Australia.

The design objectives of CIMSMINER are to research on database, data warehousing and data analysis, and provide a reference schema. Its objectives are described as follows:

1) The ability of extracting data from the heterogeneous data sources.

2) The ability of maintaining data warehouse, such as supporting storage optimization and increment.

3) The ability of integrating OLAP with data mining.

4) The ability of managing various data mining algorithms.

5) The ability of algorithms-override, namely when the data sources change, system can intelligently choose the proper algorithms.

6) Provide algorithm primitive to keep the algorithms easy to be transplanted.

7) Provide some open interfaces to the users.

8) Have some visual ability to show the mining results.

9) With the meta-database and meta-knowledge base, the system has some self-adaptability for data analysis under the instructor.

Among these objectives, the trait of open is very important. Because of large differences among database systems in enterprises that have implemented CIMS, a well-defined system should not only provide API interfaces (such as JDBC, ODBC and so on) to access various databases, but also has the ability to manage the extensible algorithms base. By this way, users can operate algorithms by the open interfaces according to the practical needs in order to make the system adapted to different mining work.

2.2. Model of CIMSMINER

The logical architecture of CIMSMINER is divided into three layers described as fig. 1[2].

The obtaining layer of CIMSMIINER is made up of data obtaining module. It can samples proper data from the CIMS database, completes data cleaning and mode transforming, and then integrates data into various subjects of data warehouse. Data obtaining layer requires much background knowledge from meta-database and knowledge base to instruct data pre-processing.

The storage layer is consisting of one or more databases and data are stored and integrated to a certain subjects. Every subject is stored in one database which is made of many tables including fact table, dimension table and temporary table of data mining.

The mining layer includes multidimensional analysis tools, data mining tools and visual tools. Many algorithms are integrated in this layer and the users implement data analysis mainly through this layer. It is the focus of the data mining research.

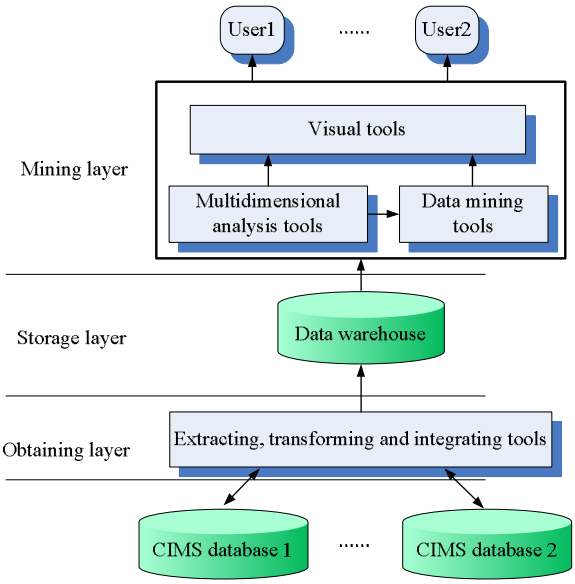

Figure 1. Logical architecture of CIMSMINER

2.3. Physical architecture of CIMSMINER

Considering the database size, the execution speed and system-operability, we adopt the physical architecture as fig. 2 shows.

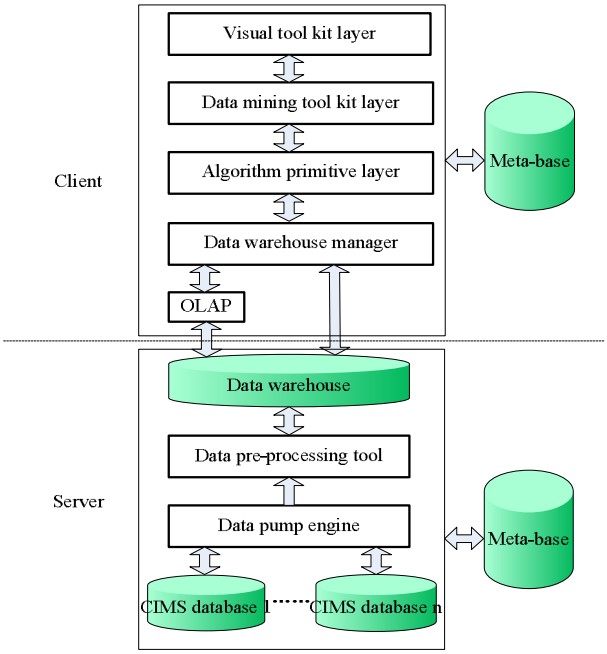

Figure 2. Physical architecture of CIMSMINER

The server mainly include three components: data pump engine, data warehouse and mete-base. The main function of the data pump engine is to pump data from the heterogeneous data sources into the warehouse. The data warehouse stores the preprocessed data and the marshaled data that can be used for OLAP operations. Data stored in the meta-base can instruct various operations for the pump and warehouse. In order to access heterogeneous data sources, CIMSINER introduces the data pump, which is the tool of extracting data from the operational system and periodically updates the warehouse to make the data synchronized with the data source. Data pump provides a three-layer data sources navigation mechanism. The first layer is the data source layer and the entire data source including its location is listed. The second layer is the table layer, which lists the user-chosen tables in data source from the navigation. The third layer is the attribute layer and provides the entire attribute for users to choose. The three-layer navigation mechanism ensures users to know the detailed information before loading data [2].

The client mainly includes data mining algorithms tool kit, algorithms primitive and data warehouse manager. Algorithms tool kit that implements knowledge discovery from database is the main component. Its main function can be described as follows:

1) Association. It discovers relations between data items, and is often denoted as association rules.

2) Clustering. Similar data are clustered according to some attributes.

3) Classification. By analyzing the training set, we can construct a model based on the characters on the data for each class. So users can easily understand the character of data.

4) Predication. Missing values and developing trends can be forecasted by some data distribution and strong association rules.

2.4. Data mining methods

Many data mining methods can be integrated in CIMSMINER, such as decision tree, rough sets, fuzzy sets, neural networks, clustering and so on. Here, we only take the clustering as an example to introduce the data mining methods.

Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way". A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data.

Fuzzy c-means (FCM), developed by Dunn in 1973 and improved by Bezdek in 1981, is a method of clustering which allows one piece of data to belong to two or more clusters. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2, \quad 1 \le m < \infty \tag{1}$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center.

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|}\right)^{\frac{2}{m-1}}} \text{ and } c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \tag{2}$$

This iteration will stop when

$$\max_{ij}\left\{\left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right|\right\} < \delta \tag{3}$$

where δ is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

3. Application

The definition of automotive products remanufacturing is "the process using advanced surface technology, composite surface technology and other new technologies to make the wasted automotive products renewable and have full application".

In the process of automotive products remanufacturing, we generally concerns how to employ new technology to improve the quality of remanufacturing products, while we rarely pay attention to the analysis and application of the information that accompanies products in their life time circle. Generally, a whole life circle of product mainly includes design, manufacturing, disassembling and remanufacturing, in which so much product data is created. It is easy to understand we should make great efforts to explore some useful information that hides deeply in these data resources to help the

remanufacturing of automotive products. But there are still some problems to be resolved. In China, besides some automotive manufacturing enterprises taking charge of remanufacturing, more automotive products are recycled (mainly through disassembling and remanufacturing) by the professional enterprises. These professional enterprises generally don't have any or detailed original information (such as designing and manufacturing information) about the recycling products, so they can't choose the most proper technology to do remanufacturing according to the original information of product. The situation is similar to the manufacturing enterprises. They even don't know how to disassemble and remanufacture the wasted products efficiently. When manufacturing engineers want to improve the product, what they refer to is only the theory and experience. They are not sure which structure is convenient for disassembling and where the problem of the products for reusing is. Due to the incomplete information transformation, many automotive products in China can't be remanufactured and be reused fully. In this case, just as the fig.3 shows, products information circle chain is broken and so much precious information data is wasted.

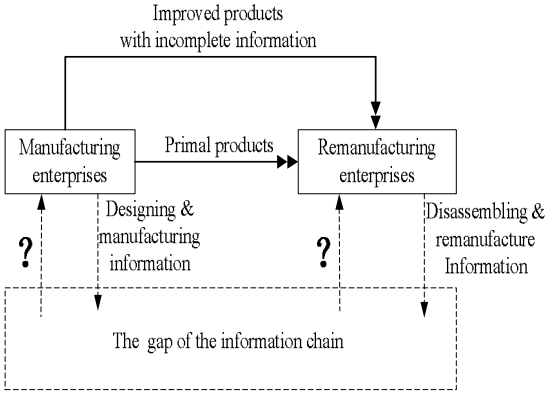

Figure 3. Incomplete information chain

Fig.3 describes the automotive products information circulating situation vividly. There is a gap between manufacturing enterprises and remanufacturing enterprises. For manufacturing enterprises, on one hand, they can't get the detailed corresponding remanufacturing information for the improvement of products; on the other hand, they don't know who need the designing and manufacturing information. The truth is that so much precious information data is stored in their data warehouse but little shared. The remanufacturing enterprises face the embarrassment too. In such a case, the products that are improved with incomplete information can hardly meet the requirements of markets.

If we want to improve the situation and make the information used fully, we must resolve two problems: the information break and the low utilization rate of information.

Since the CIMS technology is widely used in automotive industry of China, CIMSMINER is our proper solution. We integrated all the information data together by CIMS and then let CIMSMINER to be the public platform to get useful information from the CIMS databases. All the enterprises can obtain the data they need and require more information they concern through CIMSMINER.

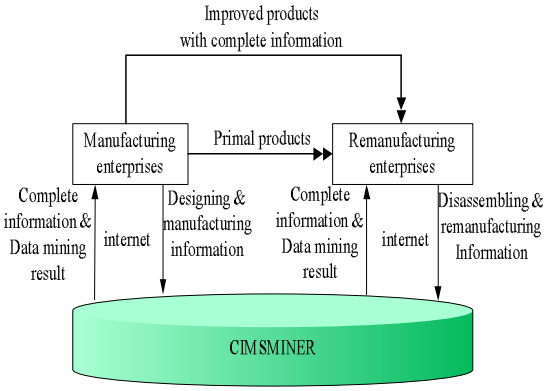

Figure 4. Complete information chain

As fig. 4 shows, CIMSMINER connects all the enterprises together to make the automotive products information chain complete. Every enterprise in the information chain transfers the product information data into CIMSMINER and get data mining results from CIMSMINER via internet. Thus, a complete information chain is constructed for automotive products improvement and remanufacturing.

Take a manufacturing enterprise for instance. In the past, if the manufacturing engineer wanted to improve a product, he modified its structure and replaced the material according to theory and experience, considering little about disassembling and remanufacturing. Obviously, this improvement is not reasonable and efficient. After the application of CIMSMINER, the manufacturing engineers not only gets the status information of wasted products, but also obtains the data mining results such as the relationship between the remanufacturing rate and the structure, the main factor resulting in wasting, the degenerative trend of material and so on. With the all-around information, the improved product will be easier to be disassembled and the quality after remanufacturing will be higher. Thus, we can do better to reduce wastes and save resources, which relates to the long-term development of our country.

4. Conclusion

Nowadays, environment protecting and resource saving has been a universal concerning issue and all the countries pay attention to the reuse, recycling and recovery of automotive products. Till now, so many laws and regulations have been issued. In China, the government also clearly defines the limits of recycling rate and already taken some measures to encourage the remanufacturing of automotive products. These measures have prompted both the manufacturers and remanufactures to seek for methods to enhance the ability of remanufacturing. CIMSMINER integrates the automotive products information depending on CIMS platform and network, and provides valuable information to different enterprises by data mining technology. With the help of CIMSMINER, the quality of China's automotive remanufacturing products will be greatly improved.

With the development of society and the progress of industry, the demand on information sharing and data mining is increasing faster. Undoubtedly, CIMSMINER will be used in more and more fields and get greater success.

5. References

[1] Dechang Pi, "Research on Key Technologies about Data Mining Database", *Dissertation,* Nanjing University of Aeronautics and Astronautics, Nanjing, 2002.

[2] Dechang Pi, Fenglin Zhang, Ningsheng Wang, Xiaolin Qin, "A Data Mining Framework Oriented CIM for Cooperative Manufacturing", *Proceeding of the 25th Chinese control conference*, Harbin, Heilongjiang, 7-11, August, 2006, pp. 1357-1362.

[3] P.S. Georgilakis, A.T. Gioulekas, A.T. Souflaris, "A Decision Tree Method for the Selection of Winding Material in Power Transformers", *Journal of Materials Processing Technology*, vol. 181, 2007, pp. 281-285.

[4] L. Wehenkel, *Automatic Learning Techniques in Power Systems,* Kluwer Academic, Boston, 1998.

[5] R.T.Lund, "The Remanufacturing Industry—Hidden Giant", *Research Report*, Manufacturing Engineering Department, Boston University, Boston, 1996.

[6] Binshi Xu, Zhenxue Zhang, "Surface Engineering and Remanufacturing Technology", *International Conference on Advanced Manufacturing Technology*, New York Press, 1999, pp. 1129-1132.

[7] G. Ferrer, R.U. Ayres, "The Impact of Remanufacturing in the Economy", *Ecological Economic*, vol. 32, no. 3, March, 2000, pp. 413-429.