The Application of Data Mining Technology for Intelligent Enterprise Resource Planning System

Ruey-Shun Chen, Yung-Shun Tsai
Institute of Information Management, National Chiao Tung University, Taiwan
rschen@iim.nctu.edu.tw

Abstract

This paper proposed an intelligent ERP system by integrating enterprise resource planning, data warehouse, online analytical processing, data mining and artificial intelligence. The data warehouse for this system is provided by the massive amounts of data gathered from an ERP system. Through a three process of integrating ERP systems with data warehouses, data warehouses with decision analysis and decision analysis with data mining systems, a three-tiered web-based systematic framework has been established. The result of this study is the integration of the ERP system and data mining system. According to experimental analysis, the defect rate of short-circuiting for parts, the defect rate of the part & components' empties (solder), the rate of solder cracking and brittleness defect all have been improved for manufacturing industry.

1. Introduction

In the 1960's, enterprises began using computers to manage information relating to daily business transactions, thus, saving valuable manpower and increasing the accuracy of information. This process represents the automation of basic and routine tasks. In this decade, with advancements in information technology, many new tools have brought forth revolutionary developments and applications in the business world such as object-oriented techniques, soft computing techniques, artificial intelligence, internet, data warehouse, and data mining technology. One of these tools, data mining technology, can dig up information buried within raw data, allowing high level personnel to make quick decisions, which helps to increase an enterprise's competitive advantage.

According to Nonaka et al. [15], to gain a competitive advantage, knowledge must first be grasped. In an era of knowledge economy, knowledge will replace the traditional factors of production and become the most important of resources. To achieve success, enterprises must constantly create new forms of knowledge, infuse new types of knowledge into existing organizational systems, rapidly absorb new types of technology and replace outdated technology with new products. As more enterprises implement ERP systems, they accumulate massive amounts of data. This research will explore some fundamental problems to be faced by enterprises in the future. These problems are discussed below:

- *ERP system's effectiveness in increasing competitive advantage is gradually disappearing:* ERP systems stress the rapid integrative processes of online transaction processing, focusing upon software to increase enterprise's work efficiency. As many enterprises have implemented ERP systems, the effectiveness of increasing competitive advantage for enterprises is gradually disappearing. Thus, bringing the valuable assets held in store by ERP systems back to life, by transforming raw data into organized information, knowledge into intelligence, and by broadening the range of applications for enterprise's information systems, is precisely the key factor for enterprises in creating a diverse competitive advantage.
- *The inadequacy of traditional decision support systems:* When responsible personnel are in the decision-making process, the information provided by traditional decision support systems may be inadequate. This information may not have reached optimal levels of effectiveness. Therefore, it is necessary to make up for the inadequacy of traditional decision-making support systems in the integration and analysis of information.

The goal of this research is to implement an intelligent ERP system. Through the integration of five major subject areas (enterprise resource planning, data warehouse, online analytical processing, data mining and artificial intelligence) the massive amounts of data assets accumulated by ERP systems may be used to obtain information valuable to enterprises in making important

decisions. As part of this research, the framework for a data mining system has been designed. This framework allows for the integration of ERP system and data mining tasks. Moreover, in accordance, an appropriate information system is practically planned out and established.

2. Literature review

2.1. Intelligent management decision system

Intelligent management decision systems are the combination of information technology and artificial intelligence. Such systems allow organizational networks to work toward the computerization of management systems, which are made up of particular regulations and tasks. Moreover, by putting to use such computerized systems, assisting systems may carry out data management, information analysis and anticipatory system monitoring in a convenient manner. Consequently, by performing all types of responses and management and decision-making, the computerized systems may carry out their assigned tasks.

In the most industries, especially like as Taiwanese PC manufacturers, it is thought that integrated information systems will permit fast cost-effective responses to unpredictable and ever-changing product demand, and support rapid product launches for previously unplanned products tailored to meet changing customer desires. Many Taiwanese PC manufacturers have been compelled to alter their production strategy and enhance their product quality [18]. In general computerized information systems implemented in the PC manufacturers are typically used to support the following management:

- *Manufacture:* OEM service, design support, product quality, and customer engineering.
- *Marketing and sales:* Customer information, market analysis, and sales information.
- *Technology:* Process technology and featured technology.
- *Supporting:* Quality assurance, administration, procurement, and finance.

2.2. Data warehouse and data mining

Inmon [19] believes that a data warehouse is an integrated, subject oriented, time variant and nonvolatile collection of data in support of management's decisions. Berson et al. [1] believes that data mart is a data store that is a subsidiary of a data warehouse of integrated data. A data mart might, in fact, be a set of denormalized, summarized, or aggregated data. Online analytical processing gathers together, sorts through and analyzes the data stored in data warehouses, creating substantial data. Various modes of data are presented for users to access. This allows users to view data using various perspectives and subjects [7].

Groth [17] points out that data mining is the process of finding trends and patterns in data. The objective of this process is to sort through large quantities of data and discover new information. Organizational data-mining is defined as leveraging data-mining tools and technologies to enhance the decision-making process by transforming data into valuable and actionable knowledge to gain a competitive advantage [5]. Zhengxin Chen [20] points out that the work process of data mining is composed of eight primary tasks. The eight primary tasks are task relevant data, background knowledge, problem statement, kinds of knowledge to be mined, data mining algorithm, models or knowledge patterns mined, interestingness, and user.

3. Problems with ERP systems

ERP systems integrate key business and management processes within and beyond a firm's boundary [6]. ERP systems are configurable information systems packages that integrate information and information-based processes within and across functional areas in an organization [11]. O'Leary [3] believes that ERP systems are computer-based systems designed to process an organization's transactions and facilitate integrated and real-time planning, production, and customer response. Chorafas [2] believes that ERP systems are systems of application that go beyond departmental differences and share in common all enterprise resources to integrate all of the functions that constitute enterprise organization.

When considering the role and function of ERP Systems, it is apparent that such systems are only able to solve problems concerned with mid-level management. ERP systems are unable to carry out a complete process of analysis, allowing massive amounts of data stored within the system to become more meaningful. Moreover, ERP systems are unable to satisfy the needs of high-level management personnel in policy-making decisions, to speed up the creation of new strategy.

As for the role played by ERP systems in the area of E-commerce, they act as foundational information systems, responsible to integrate data related to basic processes of E-commerce. ERP systems, however, are unable to withdraw valuable knowledge from massive accumulations of data stored with the B2C and B2B. Moreover, such systems are unable to rapidly and

accurately analyze data, which may be useful to business operations, in an objective manner.

When considering the extended functions of ERP systems, it is clear that the integration of data warehouses with ERP systems has ameliorated some problems with traditional ERP systems. For example: the ability to integrate, acquire and analyze information. Nevertheless, how to allow enterprises to improve the automation of transaction processes, henceforth, increasing the automation of decision-making analysis and the intelligence of application decision, is the major problem that ERP systems must overcome.

4. Design an intelligent ERP system

4.1. The overall description of proposed framework for intelligent ERP system

Based upon the data mining system, the overall description of proposed framework for intelligent ERP systems is illustrated in Figure 1. The design primarily focuses upon the following problems: the integration of ERP systems and data warehouses, the integration of data warehouses and decision-making analysis, the integration of decision-making analysis and data mining systems, and the set up of a data mining engine.

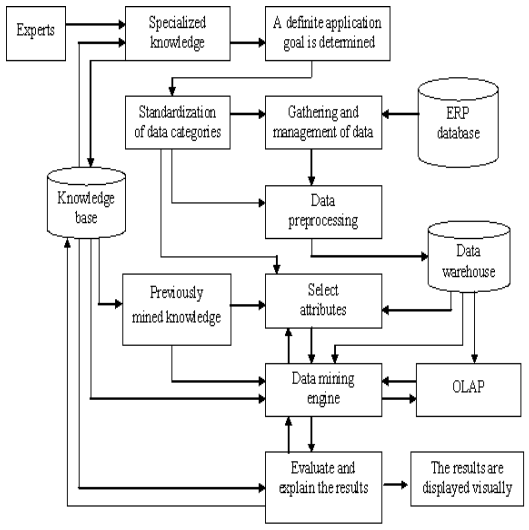

Figure 1. The overall description of proposed framework for intelligent ERP system

4.2. The framework for integrating ERP systems and data warehouses

When integrating an ERP system and a data warehouse, it is common to encounter the problem of inconsistent, incomplete and duplicate data. Therefore, the integration of ERP systems and data warehouses involves the collection of different types of data from their original sources. This data is placed in a data staging area where it undergoes such processes as the cleaning, pruning, combination and removal of duplicates. Next, the data is stored within a presentation server. At this point, users can carry out search tasks.

The procedures for integrating ERP systems and data warehouses include eight distinct steps. These steps are explained below:

- *Collection:* After gathering primary data, needed data is copied into a data staging area for further processing.
- *Transformation:* This includes revising of data accuracy and the removal from storage of unneeded data.
- *Loading an indexing:* Transformed data is saved in a data mart and indexed.
- *Quality control check:* Assuring the quality of data.
- *Announcement or publication:* The preparatory work for the official on-line installation of the system.
- *Renovation:* The continuous revision of out of date and inaccurate data.
- *Search:* Data search services are provided.
- *Checks and preparations:* Assuring the safety of the data warehouse to avoid potential damages.

4.3. The framework for integrating data warehouses and decision-making analysis

This research utilizes the star schema in designing the schema for data warehouse. This schema is based upon the manufacturing fact table, a time dimension table, an area dimension table, a product dimension table, and a quality dimension table.

When designing a fact table, several factors must be taken into consideration:

- Decide the data warehouse period of for all functions. The time periods designed for the data warehouse in this research include a two year period to measure trends in quality data, a six month period to analyze quality data, and a one year period to analyze the positioning of quality data.
- Determine a principle to be used in statistical sampling for all functions.
- Determine which fields are included in the fact table and eliminate unneeded data occupying these fields.
- To effectively save space for significant data, the size of fields included in the fact table should be minimized.
- Determine whether or not to use an intelligent key to speed up the data search process.

The dimension table is designed through a process of denormalization. For example: when the quality of factory product is analyzed, a new dimension table can be designed and used within a table to address related quality considerations. Such considerations may include: increasing slope of heating up for temperature, decreasing slope of cooling down for temperature, transmission speed of conveyor, drying time of solder, etc.

When analyzing data, multiple dimensions are brought together as one point of consideration. This process is called a multidimensional data model [14]. Data warehouse systems may include many data cubes. Each data cube may be the product of different dimension and fact tables. A data cube may be an N-dimensional data model [4]. In order to provide an even wider range of search capabilities, this research uses the four dimensions of time, area, product and quality to construct a four-dimensional data cube model.

4.4. The framework for integrating decision-making analysis and data mining system

After completing the construction of a data cube, it is possible to integrate decision-making analysis and the data mining system [20]. The goals of integration are to allow OLAP analysis results to supply the knowledge base within the data mining system, thus providing analysis information to the data mining system and creating a point of reference for data mining tasks. OLAP technology is able to blend together people's observations and intelligence within the data mining system, thus improving the speed and depth at which data is excavated. Furthermore, the intelligence discovered by the data mining system acts as a guide in OLAP analysis tasks, increasing the depth of analysis. As a result, information left unearthed by the OLAP, is extremely complex and delicate in nature.

4.5. Designing a data mining system

This paper's algorithm used decision tree method for data classification and prediction. The algorithm acts as the nucleus of the data mining engine in classifying data hidden within the database and in anticipating information. A step-by-step description of this algorithm is given below [16,8,9]:

Step 1. Prepare previously classified training data.

Step 2. Establishing a decision tree node.

Step 3. The expected information of the classified data samples selected for calculation:

The expected information needed to classify a given sample is given by

$$I(S_1,S_2,...S_m) = -\sum_{i=1}^{n} P_i \log_2(P_i) \quad (1)$$

Step 4. The expected information of the test attribute selected for calculation:

The entropy, or expected information based on the partitioning into subsets by attribute A, is given by

$$E(A) = \sum_{j=1}^{v} \frac{s_{1j}+...s_{mj}}{S} I(s_{1j},s_{2j},...s_{mj}) \quad (2)$$

$$I(s_{1j},s_{2j},...s_{mj}) = -\sum_{i=1}^{m} P_{ij} \log_2(P_{ij}) \quad (3)$$

$$P_{ij} = \frac{s_{ij}}{s_j} \quad (4)$$

Step 5. The information gain of the test attribute selected for calculation:

The encoding information that would be gained by branching on A is

$$\text{Gain(A)} = \text{I(s1,s2,…,sm)} - \text{E(A)} \quad (5)$$

Step 6. Repeat steps 2-5 until the information gain of the test attributes is completely calculated.

Step 7. Select the test attribute with the highest information gain to act as the node of partition for the decision tree.

Step 8. To complete set up of the decision tree, follow this sequence of steps to find test attribute nodes at each level.

5. Practical implementation and operation of intelligent ERP system

5.1. Operational procedures of the intelligent ERP system

With the systematic information infrastructure provided by the worldwide web, users may use the Internet to interact in more convenient manner. This three-tier information infrastructure of the intelligent ERP system is used by the company.

The operational procedures of the intelligent ERP system are described as follows:

- *User entry system:* Users operate related procedures of the data mining system through the Intranet or Internet.
- *Loading the data of the ERP system into the data warehouse system:* Based on the requirements of the manufacturing subject of the data warehouse, the manufacturing data tables, quality data tables, product data tables of the ERP system will be considered data source of the entries. Afterwards, through the clearing, collation, and transformation of the data, they are then entered into the data warehouse system.
- *Establish schema of data warehouse:* The data warehouse system is built according to the manufacturing subject, and the star schema is established. The manufacturing fact table is at its

center, with the related quality dimension table and the product dimension table.

- *Establish OLAP decision-making analysis:* According to the manufacturing fact table, quality dimension table, and product dimension table, the operation of multidimensional data cube is simulated using ROLAP method.
- *Select source of data and attribute for data mining:* The data in the data warehouse system and the results of OLAP operations can be sources of data for data mining. The mission model editor of the data mining engine can be used to help user select the source of data and attribute.
- *Select algorithm and functions of data mining:* The data mining algorithm provided by the calculation database include decision tree, neural network, genetic algorithm, and market basket analysis; data mining functions include classification, clustering, prediction, and affinity grouping. This system uses decision tree as the algorithm, and it uses classification and prediction as the function.
- *Executing data mining system:* Mission processor is the core of the data mining. It uses a target-oriented processing system to execute data mining and acquire the wanted results.
- *Interpretation, evaluation, and exhibition of results:* The results acquired from data mining system are usually some abstract data. Consequently, the system uses the rule based knowledge presentation method of the expert system and complements it with a web-based framework to express and interpret data mining results to help user understand the results gained.

5.2. Practical operation of data mining system

High quality data from the electronics industry database is used by this research to analyze categories of quality factors in the manufacturing process. The training data are shown in Table 1. Training data will be partitioned in accordance with the data mining system. When the goal of data mining is to classify or anticipate the results of data and create easy to comprehend rules, the decision tree algorithm is best suited to act as the nucleus of data mining [12,13]. The completed decision tree is illustrated in Figure 2.

5.3. Analysis of results

The knowledge rules possessed by the decision tree described above allow for the convenient gathering of information, by tracing this information along the path from root node to leaf node. In this research, knowledge rules are expressed through the rule based knowledge presentation method of the expert system, thus, revealing the knowledge rules possessed by the decision tree. These rules are described below:

- IF *heat_up* = "<2" AND *transmission_speed* = "90cm" THEN *quality_information* = "*NG*"
- IF *heat_up* = "<2" AND *transmission_speed* = "70cm" THEN *quality_information* = "*OK*"
- IF *heat_up* = "2~3" THEN *quality_information* = "*OK*"
- IF *heat_up* = ">3" AND *cool_down* = "<=-3" THEN *quality_information* = "*NG*"
- IF *heat_up* = ">3" AND *cool_down* = ">-3" THEN *quality_information* = "*OK*"

Table 1. Training data from the quality database

item	heat_up	cool_down	transmission_speed	drying_time	Class (quality_information)
1	<2℃/sec	>-3℃/sec	90cm/min	120sec	NG
2	2℃~3℃/sec	>-3℃/sec	90cm/min	120sec	OK
3	>3℃/sec	>-3℃/sec	70cm/min	80sec	OK
4	<2℃/sec	>-3℃/sec	90cm/min	100sec	NG
5	>3℃/sec	>-3℃/sec	70cm/min	100sec	OK
6	2℃~3℃/sec	<=-3℃/sec	90cm/min	100sec	OK
7	>3℃/sec	<=-3℃/sec	90cm/min	100sec	NG
8	<2℃/sec	<=-3℃/sec	90cm/min	120sec	NG
9	>3℃/sec	>-3℃/sec	90cm/min	100sec	OK
10	>3℃/sec	<=-3℃/sec	70cm/min	80sec	NG
11	2℃~3℃/sec	<=-3℃/sec	70cm/min	80sec	OK
12	<2℃/sec	>-3℃/sec	70cm/min	80sec	OK
13	<2℃/sec	<=-3℃/sec	70cm/min	100sec	OK
14	2℃~3℃/sec	>-3℃/sec	70cm/min	120sec	OK
15	<2℃/sec	<=-3℃/sec	90cm/min	100sec	NG
16	<2℃/sec	>-3℃/sec	70cm/min	100sec	OK
17	2℃~3℃/sec	>-3℃/sec	70cm/min	100sec	OK
18	>3℃/sec	<=-3℃/sec	70cm/min	120sec	NG

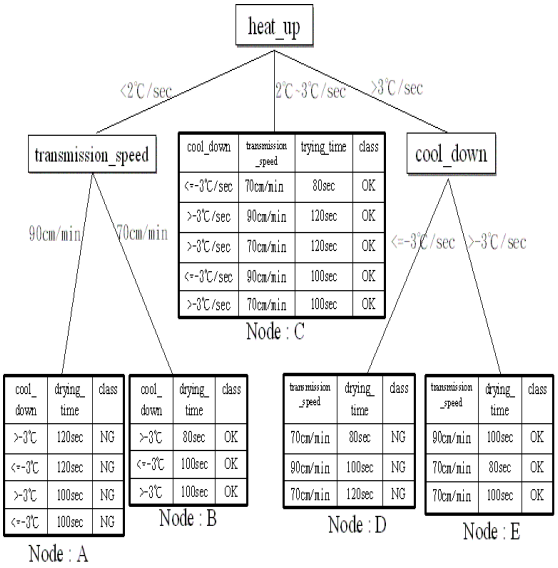

Figure 2. The decision tree for training data

This research uses the data mining system, with the vast amount of quality data from the manufacturing process as basis, to produce accurate

quality information and knowledge rules. Moreover, based on the acquired knowledge rules, is able to understand the characteristics and attributes of the important causes of quality, allowing it to stay on top of the quality issue. By strengthening the training of quality assurance personnel, adding extra heaters for the increase of temperature, installing adjustable fans for the decrease of temperature, and controlling the speed of conveyors, these methods improve product quality, lower manufacturing costs, and turn raise the overall business performance.

6. Conclusions

Through the actual establishment of data mining and intelligent ERP system, and the use of these systems on the manufacturing processes of the electronics industry for quality improvement, the experience has verified the following results:

- *The basic framework of establishing intelligent ERP system:* This allows corporate computerized applications, which will not be limited to the level of data processing, but it can also aggressively work towards an information-based, knowledge-based, and intelligent form of management information system.
- *Intensely and rapidly integrating the ERP system and data mining system, to allow for an intelligent ERP system:* Integrating ERP system and data warehouse system will allow to effectively integration the vast amount of transactions from the company's day-to-day business operations into the data warehouse system, which will become an important source of information to help decision-making.
- *Using the intelligent ERP system to improve the product quality:* The company uses the data mining of the intelligent ERP system to improve product quality, lower manufacturing costs, and turn raise the overall business performance.

References

[1] Alex Berson, Stephen Smith and Kurt Thearling, *Building Data Mining Applications for CRM*, McGraw-Hill, 2000.

[2] Dimitris N. Chorafas, *Integrating ERP, CRM, Supply Chain Management, and Smart Materials*, Auerbach, 2001.

[3] Daniel E. O'Leary, *Enterprise Resource Planning Systems: Systems, Life Cycle, Electronic Commerce, and Risk*, Cambridge University Press, 2000.

[4] Edward Hung, David W. Cheung and Ben Kao, "Optimization in Data Cube System Design", *Journal of Intelligent Information Systems*, 2004, Vol. 23 No. 1, pp. 17-45.

[5] Hamid R Nemati and Christopher D Barko, "Key Factors for Achieving Organizational Data Mining Success", *Industrial Management and Data Systems*, 2003, Vol. 103 No. 4, pp. 282-292.

[6] Hitt, L. M., Wu, D. J. and Zhou, X., "Investment in Enterprise Resource Planning: Business Impact and Productivity Measures", *Journal of Management Information Systems*, 2002, Vol. 19 No.1, pp. 71-98.

[7] Jens Lechtenbörger and Gottfried Vossen, "Multidimensional Normal Forms for Data Warehouse Design", *Information Systems*, 2003, Vol. 28 No. 5, pp. 415-434.

[8] Jiawei Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.

[9] J. Ross Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.

[10] Karl M. Kapp, *Integrated Learning for ERP Success: A Learning Requirements Planning Approach*, The St. Lucie Press, 2001.

[11] Kumar, K. and Hillegersberg, J. V., "ERP Experiences and Evolution", *Communication of The ACM*, 2000, Vol. 43 No. 4, pp. 22-26.

[12] Michael J.A.Berry and Gordon S.Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons, 1997.

[13] Michael J.A.Berry and Gordon S.Linoff, *Mastering Data Mining: The Art & Science of Customer Relationship Management*, John Wiley & Sons, 1999.

[14] N. Gorla, "Features to Consider in a Data Warehousing System", *Communications of The ACM*, 2003, Vol. 46 No. 11, pp. 111-115.

[15] Peter F. Drucker, Ikujiro Nonaka and David A. Garvin, *Harvard Business Review on Knowledge Management*, Harvard Business School Press, 1998.

[16] Rajeev Rastogi and Kyuseok Shim, "Public: A Decision Tree Classifier that Integrates Building and Pruning", *Data Mining and Knowledge Discovery*, 2000, Vol. 4 No. 4, pp. 315-344.

[17] Robert Groth, *Data Mining: Building Competitive Advantage*, Prentice-Hall, 2000.

[18] Ruey-Shun Chen and Kun-Yung Lu, "A Case Study in the Design of BTO/CTO Shop Floor Control System", *Information & Management*, 2003, Vol. 41 No. 1, pp. 25-37.

[19] W. H. Inmon, *Building the Data Warehouse*, John Wiley & Son, 2002.

[20] Zhengxin Chen, *Data Mining and Uncertain Reasoning: An Integrated Approach*, John Wiley & Sons, 2001.