

Using Data Mining Technology to Design a Quality Control System for Manufacturing Industry

R. S. Chen, Y. C. Chen and C.C. Chen

Department of Information Management, China University of Technology, Taiwan
rschen@cute.edu.tw

Department of Information Management, Chung Cheng University, Taiwan
chen1868@gmail.com

Department of Information Management, Tung-hsi University, Taiwan
Emily@thu.edu.tw

*Corresponding author rschen@cute.edu.tw

Abstract: - This paper is a quality control system involved in using data mining to discover the main inconsistency reasons in the manufacturing process of semiconductor plants and compare the correctness of classification analysis of the two methods, so as to set up a quality control system providing an efficiency tool for analyzing problems, with a view to identifying the causes of problems, making decision immediately, and eventually reducing the cycle time taken to solve quality-related problems.

The contributions of this paper are as follows. Predictions made by decision tree analysis, indicating that decision tree analysis is an effective mean of classification analysis in semiconductor quality problems, whereas evaluation of feasible methods by data mining followed by establishment of the basis for a quality analysis system environment, that is characteristic of knowledge sharing may be applied to analysis of the quality problems in all corporation.

KeyWords: Data mining, Quality Control, Manufacturing Industry, OLAP.

I. Introduction

With the keen competition among enterprises, use of the Internet popularly, and competition in international trade, it is vital to catch up rivals' competition strategies and fulfill customers' needs to take proper problem-solving steps timely. However, owing to a huge amount of information, finding unearthed information and identifying clients' responses to any problems with products are usually time-consuming [1] [13]. Customers will be satisfied to a greater extent, only if the staff acquires intended or required information more efficiently and swiftly. Some key issues are urgent to be built in a useful quality control system, i.e., providing useful information, constructing know-how that integrating data warehouse and data mining, analyzing the data about any previous problems with product quality through the creation of a database, discovering the reasons of problems with the products timely, and solving the problems quickly. Enterprises engaged in keen competition attach great importance to improve products quality; whoever obtains accurate information faster, and make decisions immediately, than rivals do, will have a chance of being successful [19]. Therefore, to discover the reasons of product quality problems and cope with the problems, one has to use those information systems with quality control functions to explore and solve the problems.

Another key issue that a corporate faced is information system nowadays is to grasp correct information at the right time and deliver it to the correct executives. Hence, the creation of an integrated, fast-reacting information

platform intended to timely provide situational analysis and information is a matter that demands immediate attention. In a word, an intelligent quality analysis control system based on data warehouse and data mining is to be constructed with a view to achieving the following goals.

1. Establishing data warehouse of problems with semiconductor product quality.
2. Discovering the main reasons that cause the problems of products quality and developing applicable methods.
3. Decreasing the product cycle time and increasing the total improvement rate of products.

II. Literature Review

A. Data mining

According to Cabena, Hadjinian, Stadler, Verhees and Zamasi [2], data mining is a process extracting effective information, which is unknown previously, from a large database for executives to make critical decisions. Besides, Frawley, Paitetsky-Shapiro, and Matheus [4] define data mining as a process exploring in database unobvious, implicit, unprecedented information which may be useful. Thus data mining is to use specific techniques, generalize and organize data from database and then excavate unknown, hidden information for executives to make decisions.

According to the report released by SAS Institute [10] in 1999, the process of data mining involves:

1. Definition of problem: To define the definition the problems which enterprises will encounter and to describe the goal of an enterprise.

2. Data collection and integration: There is a close relation between model creation and data; incomplete or deviated data always leads to varied models. Data collection and relevant procedures are shown in Fig. 1.
3. Establishment of learning strategies: There are two pattern of learning strategies: supervised learning and unsupervised learning. With supervised learning, the learner generalizes a interpretable data from a group of positive and negative samples. With unsupervised learning, the learner does not have a gold standard training corpus with which accuracy can be measured. Instead, they try to use information from the distribution of unambiguous words to find reliable disambiguating contexts.
4. Model training, verification and test.
5. Outcome analysis: Fayyad] argues that problem definitions and data collection, integration are preparatory work essential for data mining, therefore eighty percent of the time in the course of data mining would be spent on data processing. But, practically, data mining employs data to create some genuine models which can describe patterns and relations. These models are used to (1) understand the patterns and relations to provide decision-related information; and (2) find these data patterns for prediction.

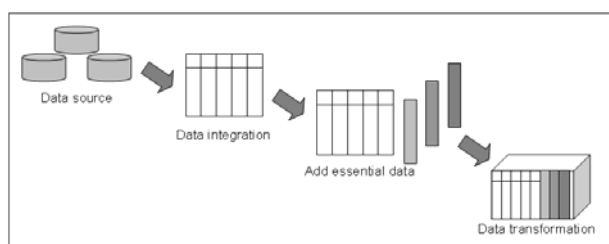


Fig 1. Data collection and relevant procedures in the course of data mining.

B. Evolution of IC packaging

A semiconductor manufacturing process consists of IC (Integrated Circuit) design, mask production, wafer fabrication, wafer packaging and testing. The whole manufacturing process can also be divided into front-end process and back-end process; wafer packaging and wafer testing are included in back-end process. Wafer packaging involves the operation connecting ICs to external circuits, enabling the signal transmission between electronic components. With heat conduction features of package materials, packaging can also eliminate the heat resulted from transmissions between the circuits to prevent ICs from being damaged due to excessive heat. Besides, IC packaging offers sufficient mechanical strength and proper protection for vulnerable chips, which also prevent sophisticated ICs from being contaminated [15].

The evolution of IC packaging is generally subdivided into four generations: The earliest generation, in which a plate-through hole used for interconnection of conductors on different sides or layers of a circuit board; and then followed by the generation in which the surface mount technology is applied to decrease package volume and increase the number of I/O pins. However, in both generations the lead frame principally serves as a carrier

using an old plated wire to connect chip electrodes with pins on the lead frame, which is classified as a peripheral packaging method and has limitations on the decrease of package volume and the increase of pins. Until the third generation, the packaging technique in evolved into area array that dramatically increases the number of I/O (Input/Output) pins and fulfills the demands for higher rate, higher power, and ultra-thinness, due to the application of area array and the introduction of organic substrate carrier modules. The fourth generation, bare chip packaging, shows prospects, especially area-arrayed flip chip that would become the mainstream packaging after mature development of peripheral technologies [3] [5] [16].

III. Problem

A. Semiconductor process

A semiconductor packaging process consists of four stages: the stage of wafer cleaning/mounting/saw, the stage of die bonding/wire bonding/curing, the stage of molding/marking, and the stage of forming/packing/storage, shown as Fig. 2. Different products undergo different processes and packaging patterns, which will then influence the processing methods. A processing method may be designated based on customer special requests [7].

1. Stage of wafer cleaning/mounting/saw: The major procedures at this stage are: Wafer income inspection, Wafer mounting, which means to glue blue tape on wafer to prevent die delamination after wafer saw; Wafer Sawing/Cleaning, Post Saw Inspection, and sampling inspection by the QC (Quality Control) division. A production batch will not be transferred to the next stage until the inspection is approved.
2. Stage of die bonding/wire bonding/curing: The major procedures at this stage are: Die Bonding, i.e. to glue dies on the lead frame one after one; Epoxy Curing, which places the semi-finished goods in an oven for curing; Wire Bonding that connects the gold plated wire from a chip to the lead frame; and Post bond inspection, and QC inspection (a production batch will not be transferred to the next stage until the inspection is approved).
3. Stage of molding/marking: The major procedures at this stage are: Molding, referring to package dies as common ICs with black colloid to prevent from external damage; Backside Marking, which means to have a mark on the bottom of an IC; Post mold cure, that sends an IC semi-finished goods for curing once more; Trimming/Dejunking, trimming the pins on leads; Solder Plating, soldering external pins on ICs; and QC inspection (a production batch will not be transferred to the next stage until the inspection is approved).
4. Stage of forming/packing/storage: The major procedures at this stage are: Top Marking, which marks on the front side of ICs; Post Mold Cure, which sends IC semi-finished goods to oven for curing again; Forming/Singulation; Final Visual Inspection; and QC inspection (a production batch will not be transferred to the next stage for packing until the inspection is

approved, and then should be inspected by the QC division for packing compliance and conformity with customer requirements prior to storage.

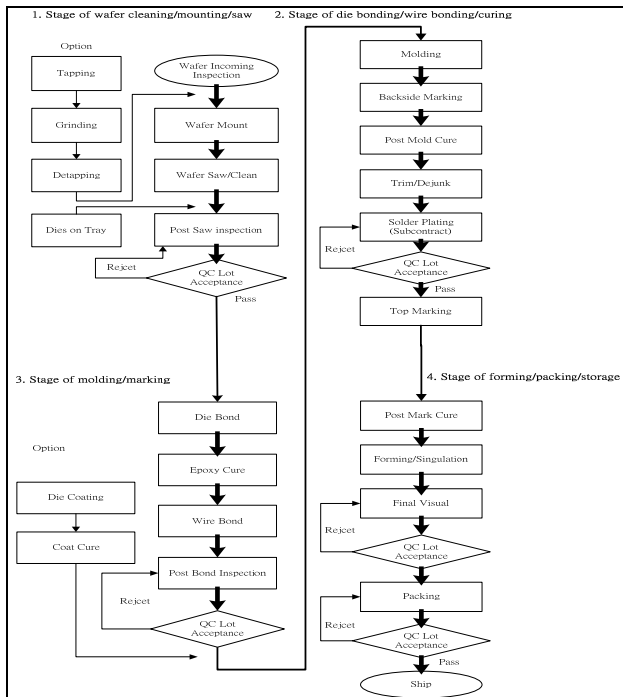


Fig. 2. Semiconductor process

B. Problem definition

The study is designed to discuss data mining in the semiconductor packaging industry with an aim to identify unknown but useful knowledge. As the manufacturing industry always uses CAR (corrective action request) to record quality-related problems raised by customers, lacking an effective way to make the most of information, it results in wasteful time and unnecessary cost on investigations and analysis when the problem reoccurs. Furthermore, data is often large in size and complicated, the personnel in charge of quality-related problems can hardly identify the discrepancy factor or generalize the characteristics or types of the problems rapidly or correctly. For these two main reasons, the challenge required to be settled in the study lies in the conclusion of the problems arising in connection with the semiconductor manufacturing plant.

There are cases applying data mining in a number of literature reviews, including manufacturing, financing and telecommunications. Among the available data mining tools, the common classification methods are Decision Tree, Naïve Bayesian, and Neural Network.. In applied telecommunications cases, Naïve Bayesian is better than Decision Tree in prediction effect [6]. Thus the study is intended to use the data mining methods, Decision Tree and Naïve Bayesian, and acquire the previous CAR data content for analysis, and identify which algorithm data mining result is superior in application to the semiconductor packaging industry, based on implementation outcomes. In this way we may make a reasonable conclusion for the patterns of incidents and build a problem diagnosis analysis system and, through

discussion with the experts in the domain concerned, assist the personnel in charge of quality-related problems to reduce the diagnosis time and scope of the incident.

IV. 4 System Design

A. System design framework

Based upon the data mining system, the complete design framework for intelligent quality analysis control system is illustrated in Fig. 3 [11]. The function of each design element is shown in the following steps.

1. Experts, domain knowledge: At first, determine which goals to achieve with data mining for relevant data collection, data pre-processing, selection of data attributes and data mining methods.
2. Data collection: The conversion of historical data from the existing system, i.e. WIP, ERP, MES, into the processing area should be considered.
3. Data standardization: After determination of data acquisition source, standardize the data type to ensure the consistency between subsequently collected data and pre-processed data.
4. Data preprocess: When data is collected in the processing area but not stored in the data warehouse, there might be lost or inaccurate data in some fields. To enhance processing efficiency and accuracy, it's essential to proceed with data integration, conversion, extraction, and cleaning.
5. Data warehouse, OLAP (Online Analytical Process): As the data to be processed is distributed in different databases and is always large in size, reduced data search time is the key to the whole process of data mining. Hence data warehousing is applied to address these challenges. Besides, OLAP operations in data cubes include rollup, drilldown, slice, dice, and pivot.
6. Select attributes: Typically, data analysis is proceeded after a proper attribute is selected, which is based on the proper attribute for specific analysis target determined by the expert in the domain concerned, because either insufficient or excessive attributes cannot achieve correct analysis results.
7. Decision tree data mining engine: A data mining engine is the core, and also the critical part, in the system framework. The most commonly used methods are classification, clustering, and sequential patterns, etc.
8. Bayesian data mining engine: The design of the data mining engine is vital, as the data mining engine serves as the nucleus of the entire framework.
9. Results evaluation: Tremendous mined data and patterns may exist; the mined result can be more available and interpretable only through parameter setup.
10. Results display: The mined result may be presented by user preference.
11. Knowledge base: The knowledge base, which stores expert expertise and the rules available after data mining, can be updated from time to time to be the basis for various decision-making supports.

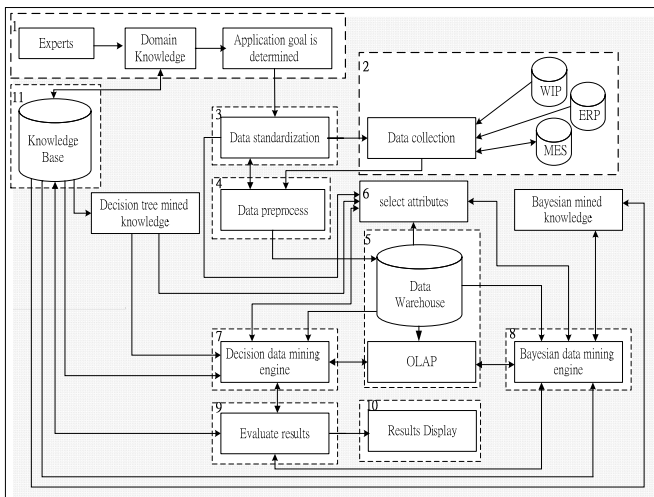


Fig. 3. Complete design framework for quality control system.

B. Establishment of Data Mining System

B.1 Classification and predictions procedures

By collecting tuples within the database, defining them systematically on the basis of the specific target of analysis, searching out common characteristics and establishing a class process, classification is accomplished. Furthermore, previously classified historical data may be utilized to anticipate which class each item belongs to. Data classification is basically comprised of the following two-step process [8] [9]:

1. Training model: Through the collecting of items within the database, a training data set is determined. This set is analyzed in accordance with the algorithm used to classify data, for example, decision tree and Bayesian. The learning model or classifier is represented in the form of classification rules.
2. Classification: Through the collecting of items within the database, a test data set is established. This set is entered into a classifier. After deviations within the classification model have been rectified, unknown data is entered into the revised classifier, thus, predicting subsequent results.

B.2 Design of decision tree data mining engine

When building a decision tree engine, we adopt the decision tree algorithm of SQL server 2005 for calculation [6]. Due to the enormous implementation case in the study, we proceed with calculation by selecting the sample data from the *AllElectronics* customer data base, where the training samples are 14 database tuples as shown in Table 2. The data sample is described by the attributes of age, income, student and credit_rating, and the class label attribute is buys_computer with two different values, which are {yes, no}.

Training data will be partitioned in accordance with the selected test attribute. The completed decision tree is illustrated in Fig.49.

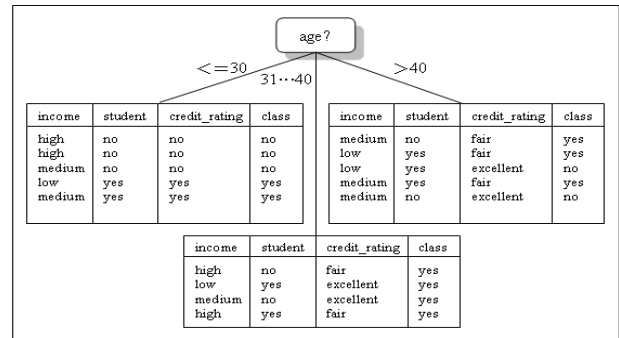


Fig. 4. The decision tree for training data.

B.3 Establishment of data mining procedures

The operational process and knowledge rules excavated by data mining can be displayed through visual figure. This allows users to operate data mining more easily and to understand the meaning represented by excavated data. A graphical user interface for a data mining engine is composed of the functional elements listed below [9]:

1. Data collection and data mining query composition: This component allows the user to specify task-relevant data sets and to compose data mining queries.
 2. Presentation of discovered patterns: This component allows the display of the discovered patterns in various forms, including tables, graphs, charts, curves, and other visualization techniques.
 3. Manipulation of data mining primitives: This component allows the dynamic adjustment of data mining thresholds, as well as the selection, display, and modification of concept hierarchies. It also allows the modification of previous data mining queries or conditions.
 4. Interactive multilevel mining: This component allows roll-up or drill-down operations on discovered patterns.
- Other miscellaneous information: This component includes on-line help manuals, indexed search, debugging, and other interactive graphical facilities.

V. System Implement

A. Data aggregation

In this stage, we diagnose and analyze quality data. After discussing with experts in the domain concerned, understanding the problem definition and purpose, having insight into the industry's characteristics, we proceed to collect data and select data items from the desired analysis scope for follow-up analysis.

Data attributes are defined as man, machine, material and method, which are described below:

1. Man factors: refer to associated man operation factors, such as failure to follow procedures and methods, failure to inspect, and failure to set parameters, etc.
2. Machine factors: refer to associated machine operation problems, such as excessive pressure, extremely high tool head position, and no sensor reaction, etc.
3. Material factors: refer to associated material problems, such as excessively strong glue, excessively thin internal pins, etc.
4. Method factors: such as poor programming, indefinite instructions, and failure to use a dust-free cloth, etc.

Table 1. Independent coding table for all attribute variables and judgment items

No	Man	Machine	Material	Method	Problem
IIM010234-01	MARKING Maladjustment of parameters (2)	ILB Pressure Error (1)	N/A (0)	ILB No positioning (1)	Defective chip (16)
IIM030046-02	POTTING Maladjustment of parameters (4)	POTTING Pressure Error (3)	N/A (0)	Excessive glue (7)	Resin overflow (3)
IIM030052-04	MARKING Maladjustment of parameters (2)	MARKING Sensor error (5)	N/A (0)	N/A (0)	Overlap marking (8)
.....
IIM310113-06	POTTING Maladjustment of Parameters (4)	ILB Pressure Error (1)	Excessively thin internal pin (3)	N/A (0)	Broken internal pin (5)
IIM310196-06	Marking Maladjustment of parameters (2)	N/A (0)	Excessively strong glue (1)	Abnormal Temp. (12)	Bubble (9)
IIM310201-08	N/A (0)	N/A (0)	Glue with bubble (2)	Abnormal Temp.(12)	Bubble (9)

B. Data cleaning

The four major parts in this stage:

1. Disposition of data loss: such as null value, inexistent value, and incomplete data adjustment and settlement.
2. Disposition of deviated value inspection: delete or retain the scattered data beyond normal distribution.
3. Clarification of fuzzy definition: In case of any value in different fields represents the same meaning, it is necessary to make the data consistent to clarify the definition.
4. Disposition of error values: If a field value does not conform to the effective value, there may be a problem with input errors or program errors. Determine the disposition method according its effectiveness.

C. New data generation and conversion

The original data of the whole quality system is extracted from WIP, ERP and MES systems. The ETL is also known as the combination of extracting, transformation, and loading. The processing procedures are: extracting data from the system, loading data to the operational data store (ODS), and purifying and transforming the data, and transferring.

D. Results comparisons and benefit analysis

To establish the database, we have collected the data form January to December in 2005. With the decision tree and Bayesian provide by the intelligent quality analysis control system for reference of man operation analysis and problem settlement form January to December in 2004, the four major factors that influence quality are: bubble, resin overflow, overlap marking, broken internal pin. After improvement with the decision tree, we have the improvement rate 12.4%, 12.0%, 13.5%, 12.7%, 13.0%, 12.0%, 11.9%, 12.8%, 13.6%, 12.1%, 12.2%, 12.7%, 12.5%, 12.3%, and 12.1% for each quality problem, and overall average improvement rate is 12.5%. By using Bayesian, we have the improvement rate 8.1%, 7.8%, 8.2%, 7.5%, 8.3%, 7.1%, 7.7%, 8.8%, 8.9%, 8.0%, 7.2%, 7.9%, 7.4%, 7.1%, and 7.3% for each quality problem, and overall average improvement rate is 7.8%. The total average improvement rate shown in Table 8 and the statistical curves before improvement vs. after improvement is shown in Fig. 16. Hence, decision tree method is more effective and accurate than Bayesian

method to apply to the quality problems in the semiconductor packaging industry.

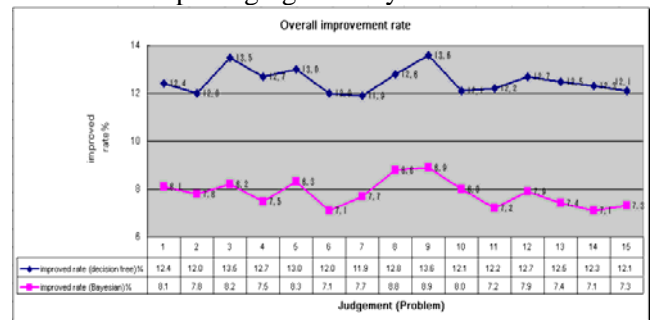


Fig. 5. Diagram of overall improvement rate curves for each quality problem yearly.

VI. Conclusions

The results and contributions of this research are listed as follows.

- A. Decision tree algorithm is more effective and appropriate than Bayesian algorithm to analyze the quality problems in the semiconductor packaging industry.
- B. In the semiconductor industry, it is found in the study that, among the four attributes, man, machine, material and method. To achieve the optimal effects with the least time, it is recommended to take machine as the first priority for data mining.

References:

[1] Alex Berson, Stephen Smith, Kurt Thearling, "Building Data Mining Applications for CRM", McGraw-Hill 2000.
 [2] Cabena, P., P. O. Hadjinian, R. Stadler, DR. J. Verhees, and A. Zanasi, "Discovering Data Mining from Concept to Implementation", Prentice Hall, p.12, 1997.
 [3] Chen C. C., "High Density IC Packaging—CSP Packaging", Materials Magazine, Vol. 151, pp.92-99, 1999.
 [4] Frawley, W. J., G. Piatetsky-Shapiro, and C. J. Matheus, "Knowledge Discovery in Database: An Overview", Knowledge Discovery in Database, AAAI/MIT Press, p.1-30, 1991.
 [5] Fan Y. M., "Development Trend of IC Packaging Materials", Materials Magazine, Vol. 151, pp.69-77, 1999.
 [6] Hsieh C.B., "Data Mining and Business Intelligence: SQL Server 2005" Ting Mao Publish Company, 2005.
 [7] Hsu G. H., "An Advanced Packaging Technology—Wafer Packaging Technology", Materials Magazine, Vol. 151, pp.86-91, 1999.
 [8] J. Ross Quinlan, "C4.5 : Programs for Machine Learning", Morgan Kaufmann Publishers 1993.
 [9] Jiawei Han, Micheline Kamber, "Data Mining: Concepts And Techniques", Morgan Kaufmann Publishers 2001.
 [10] Kristin, R. N., and I. P. Matkovsky, "Using Data Mining Techniques for Fraud Detection", SAS Institute Inc. and Federal Data Corporation, 1999.
 [11] Lee J. F., "Research and Exploration of Data Mining", Information and Education Magazine, August 2001.
 [12] Lin J. B., Liu M. D., Chen H., "Data Mining and OLAP: Theory and Practice", Kings Information Co., Ltd., 2002.
 [13] Michael J. A. Berry, Gordon S. Linoff., "Data Mining Techniques: For Marketing, Sales, and Customer Support", John Wiley & Sons 1997.
 [14] Michael J. A. Berry, Gordon S. Linoff, "Mastering Data Mining: The Art & Science of Customer Relationship Management", John Wiley & Sons 1999.
 [15] "Middle- and Long-Range Development Strategy of Taiwan's Electronics Packaging Industry; Electronic Packing Technology Development Program, MOEA", published by Industrial Technology Research Institute, 1997.
 [16] Pan J. P., "Substrate Semiconductor Packaging Industry and Technology Trend", Materials Magazine, Vol. 151, pp.78-85, 1999.