Application Data Mining Technique to improve Manufacturing Quality –A Case Study of LCD Driver IC Packaging Industry

Abstract

In recent year, because of the professional teamwork, to improve the qualification percentage of products, to accelerate the acknowledgement of product defects and to find out the solution, the LCD driver IC packaging factories have to establish an analysis mode for quality problems of product for more effective and quicker acquisition of needed information and to improve the customer's satisfaction for information system.

This research employs the star schema of data warehousing as the base of line analysis, and uses decision tree and neural network in data mining to establish a quality analysis system for the defects found in the production processes of packaging factories in order to provide an interface for problem analysis, enabling quick judgment and control over the cause of problem to shorten the time solving the quality problem. The result of research shows that the use of decision tree and neural network in reducing the numbers of defected inner leads, resin and chips has been improved, and the decision tree is more suitable than neural network in quality problem classification and analysis of the LCD driver IC packaging industry. The result can also be used in quality problem analysis for computer assembling, electrical component fabrication and semi-conductor packaging and fabrication industries.

Keywords: customer relation management, data mining, data warehousing, decision tree.

1. Introduction

Professional teamwork is one of the emphases in semi-conductor manufacturing industry, as well as the obvious supply chain of up and downstream industries. How to compete among the industries, improve customer's satisfaction and maintain the loyalty of customers has become the key factor of corporate competition. From the research result of Burgess [9] we have learned that the priority of the competition policy for manufacturing is Quality, Cost, Time and Flexibility. Therefore, how to increase the qualification percentage of product, improve the quality of product, lower the costs of production match the shipping deadlines given by the customers and the flexible production needs have become urgent issues for industries to think about.

In a time of knowledge economy, corporations have realized the importance of customer relation management and the uses of information technology to amplify their own advantages in competition. Recently, more and more new technologies and concepts, such as artificial intelligence, data warehousing, and data mining, have become available. The corporations are aided in data processing and information analysis by using computerized systems and the information of sources and customers is integrated for rapid response and problem solving, and eventually progress to active marketing and services, and help the corporations make reasonable reactions or decisions [5, 10, 11], i.e. the thought that computerization is for the most basic computer processing elevates and develops the management for information and further becomes the purpose of aids for decision making, and uses of knowledge rules can reach the purpose of optimal intelligent decision making [2].

Currently there are several research projects dealing with quality problems in production, including mining fabrication data by classification for combinations of machines in semi-conductor fabrication processes to avoid the combination of machines that compromises the quality of products and to improve the qualification percentage of semi-conductor production [14], the uses of decision tree to develop data mining in process incident diagnosis for semi-conductor wafer deviation tests and related process data [15], and the uses of SOM and decision tree to identify the characteristics and rules of production achievement index for the production and fabrication data of semi-conductor as the basis of the decision for the fabrication in the semi-conductor factories [16].

However, there is no current research dedicated in this area for the semi-conductor driver IC packing industry. Therefore, with a comprehensive integration and use of information and system, the basis for decision making can be developed using the technologies of data warehousing and data mining in the analysis for the occurrence of abnormality in process. Via the acquisition and establishment of basic and historical data, the analysis on the data of the product quality in t previous incidents may quickly identify the cause of product problem and reduce the time for responding to a problem in order to improve the customer's satisfaction.

2. Data mining

2.1. Knowledge discovery

Data mining may be considered as part of Knowledge Discovery in Database, KDD. KDD is a series of procedures that create knowledge and data mining is one of them [4, 13]. Brachman et al [6] think that "all the activities and processes in the knowledge discovery are to find useful models in the data using the algorithm of data mining and the post-processing and re-processing of knowledge in order to find out the key factor of solving problems. Data mining is in fact the most important core procedure in the entire process knowledge discovery, and the entire data mining process is merely repeating the 4 steps of confirming problem, analyzing data, taking action and evaluating result [7].

Pyle [8] pointed out that using data mining to discover knowledge is somewhat similar. For example, first the problem to be solved and the need have to be clearly defined, and secondly, the target database has to be carefully selected to eliminate the noise and the data will be properly coded to transfer the format of data display to improve the effects processed.

2.2. Data mining

In the process of data mining, different definition of problem generates different types of mining result to provide for the uses of different decisions. Currently, the most common types of problems are:

1. Prediction: Berry and Linoff [7, 12] think that prediction will classify or estimate the future possible value and trend of some variable according to certain future behaviors.
2. Association Rule:
   Berson et al [12] think that association rule is the relation of association which is mined from the database and satisfies certain conditions.
3. Clustering/Segmentation:
   Berson et al [12] think that clustering can separate the database into several sub-sets based on a group of characteristics.
4. Classification: What classification does is to find a reasonable description or model for each individual type, and then to use these descriptions or models to classify new unknown data [7].

3. Quality problems in IC packaging

3.1. LCD driver IC packaging process

As the IC process is rapidly refined, the traditional packaging technique of wire bonding no longer catch up with the demands of high electrical performance, and the flip chip technology which directly connects the bumps and substrates can answer to the trend of making the electronic product smaller and chip performance faster, placing the Chip Size Packaging (CSP) satisfying the demand of light, thin, short and small at the center of stage. Since the packaging of LCD driver IC is the center of attention, LCD driver IC is one of the major components in making LCD, and is mainly used in large size TFT-LCD panels for laptop computers and LCD monitors, as well as the smaller STN panels for mobile phones, PDAs and electronic translators.

The packaging process of LCD driver IC can be divided into the following steps and the structure of flows is shown in Fig. 2:

From the quality point of view, to establish the analysis system for the quality problems in the LCD driver IC packaging process, the factors that may affect the quality of production in each of the work stations have to be found, such as the process operations mainly contained in the fabrication station and possible defects, as shown in Table 1.

3.2. Quality problem analysis

Quality analysis system mainly consists of database, line analysis and data mining. As shown in Fig. 3, they are divided into several levels:

In the data preparation stage, 3 major activities, data collecting and screening, data scrubbing and new data generating, will be carried out Therefore, the primary fields for establishing problem data format must include the number, attributes and problem of the initial production order. For the data attributes, 4 major attributes are defined, which are man, machine, material and method, also known as 4M. They are defined as follows:

1. Man: such as incompliance with the process, skipping inspection, and failure to set up parameters, etc.
2. Machine: including any issue regarding operation of equipments, such excessive pressure, tool head too high and sensor not responding, etc.
3. Material: problems related to material, such as glue too thick or inner leads too thin, etc.
4. Method: for example, poor programming, instruction not clear, not using dust-free wipers, etc.

First, number the attribute variables and quality problems from 0, 1, 2, and so on. The attribute variable is 0 if "no error occurred," and then align them in their numbers. In the same fashion, align the quality problems by their numbers. The comparison table for fabrication attribute variables and quality problems is shown in Table 2.

The quality problem data of the case study companies were collected from Jan. 2002 to Dec. 2003, and there were 9,752 pieces of data in total. There were 15 types of problems and the numbers of individual variables under the 4 attributes of data mining were: 17 for man, 13 for machine, 10 for material and 12 for method.

4. Design of quality problem mining system

(1). Decision tree algorithm

Decision is a classification analysis method based on tree data structure. The detailed algorithm is listed as follows:

Step1: Set the current node C as the root node. Every object is included in the object set of C at this moment.

Step2: If all the objects in C belong to the same type, then set node C as this specific type and stop. Or go to Step 3.

Step3: Calculate the entropy E(C). If set C belongs different j types, then $E(C)=-\Sigma p_j \log_2(p_j)$, where p_j= (the number of objects belonging to type j)/(the number of objects in set C)。

Step4: For all the attributes A_i (known as candidate attributes) that are not shown in the path from root node to current node, divide the object set C with A_i individually and calculate the entropy $E(A_i)$ and the information gain $G(A_i)=E(C)- E(A_i)$ of the resulting partial decision tree. Select one of the attributes A_i as the decision tree node, which means m leaves will be established under this node. Therefore, the entropy of sub-decision tree generated from the decision tree node, which is attribute A_i, is determined as follows:

$E(A_i)=\Sigma(n_k/n)*E(C_k)$

Where C_k is the object sub-set k in the object set C where the attribute values of A_i are equal.

$E(C_k)$ is the entropy of the sub-set C_k

n is the total number of objects in C

n_k is the number of object in sub-set C_k

Information Gain is the change in entropy caused by selecting attribute A_i as the decision tree node; that is, the difference between the entropies of the original object set and the sub-decision tree with A_i as the decision node. The calculation is:

$G(A_i)=E(C)-E(A_i)$

Step5: Calculate Gain-Ratio as the way to select and measure the attributes, and the calculation is:

$GR(A_i)=G(A_i)/IV(A_i)$, where $G(A_i)$ is the information gain obtained from A_i dividing object set C, $IV(A_i)$ is the information value of attribute A_i $=-\Sigma p_m \log_2 p_m$ (m=1..k), where p_m = (the number of objects whose attribute value of Ai is a_m in set C)/(the number of total objects in C)

Step6: Select the candidate attribute with the largest gain ratio as the category attribute of node C.

Step7: Establish leavesC_1, C_2, ….C_m under C (assuming that there are m attribute values in the selected category attribute). Distribute all the objects in C to the appropriate leaves according to the category attribute values.

Step8: Set every leaf C_i as current node C and repeat Step2.

The standard for C4.5 decision tree trimming is determined with Predicted Error Rate. The error rate is used for evaluating the error rate of other non-learning data. Assume that the number of all learning data categorized in a certain sub-tree is N, where there are E category errors of learning data, and the predicted error ratio is the use of E/N to evaluate the possibility that the result is not normal when there is new data being tested [1]. Table 3 is the data of quality problems in LCD driver IC packaging industry with machine, material, method and man as the primary predicting variables.

Generally, information gain is used on every node of decision tree to select the test attribute key, assuming S is the set containing s data samples. Assume that there are m different values in category label (flag) attributes, and it is defined that there are m different categories. $C_i(I=1,..,m)$. Assume that S_i is the number of samples in category C_i. For an expected information needed for a given sample category, it is calculated with the following formula:

(a). Calculation of the expected information and the information gain for a predicted variable:

$I(s_1,s_2,\ldots,s_m)=-\Sigma p_i \log_2(p_i) \quad I=1,..,m$

(1). Calculating the expected information and the information gain for a material variable:

For material ="Glue itself is too thick"

$S_{11}=2 \quad S_{21}=3 \quad I(S_{11},S_{21})=0.971$

For material ="Glue itself contains bubbles"

$S_{12}=4 \quad S_{22}=0 \quad I(S_{12},S_{22})=0$

For material ="Tape itself contracts"

$S_{13}=3 \quad S_{23}=2 \quad I(S_{13},S_{23})=0.971$

$I(S_1,S_2)=I(9,5)=0.940$

Expected information

$E(material)=(5/14)*I(S_{11},S_{21})+(4/14)*I(S_{12},S_{22})+(5/14)*I(S_{13},S_{23})=0.694$

Information gain

$Gain(material)=I(S_1,S_2)-E(material)=0.246$

(2). Calculating the expected information and the information gain for a machine variable:

For machine ="None"

$S_{11}=3 \quad S_{21}=1 \quad I(S_{11},S_{21})=0.811$

For machine ="Abnormal marking pressure "

$S_{12}=2 \quad S_{22}=2 \quad I(S_{12},S_{22})=1$

For machine =" Abnormal curing temperature"

$S_{13}=4 \quad S_{23}=2 \quad I(S_{13},S_{23})=0.918$

Expected information

$E(machine)=(4/14)*I(S_{11},S_{21})+(4/14)*I(S_{12},S_{22})+(6/14)*I(S_{13},S_{23})=0.911$

Information gain

$Gain(machine)=I(S_1,S_2)-E(machine)=0.940-0.911=0.029$

(3). Calculating the expected information and the information gain for a method variable:

For method ="Abnormal temperature"

$S_{11}=3 \quad S_{21}=4 \quad I(S_{11},S_{21})=0.985$

For method ="None"

$S_{12}=6 \quad S_{22}=1 \quad I(S_{12},S_{22})=0.592$

Expected information

$E(method)=(7/14)*I(S_{11},S_{21})+(7/14)*I(S_{12},S_{22})=0.789$

Information gain $Gain(method)=I(S_1,S_2)-E(method)=0.151$

(4). Calculating the expected information and the information gain for a man variable:

For man ="None"

$S_{11}=6 \quad S_{21}=2 \quad I(S_{11},S_{21})=0.811$

For man =" Abnormal marking pressure"

$S_{12}=3 \quad S_{22}=3 \quad I(S_{12},S_{22})=1$

Expected information

$E(man)=(8/14)*I(S_{11},S_{21})+(6/14)*I(S_{12},S_{22})=0.892$

Information gain $Gain(man)=I(S_1,S_2)-E(man)=0.048$

(b). Calculating Gain-Rate(GR)：

$GR(material)=0.246/(-(5/14)*\log_2(5/14)-(4/14)*(\log_2(4/14)-5/14*\log_2(5/14))=0.156$

$GR(machine)=0.029/(-(4/14)*\log_2(4/14)-(4/14)*(\log_2(4/14)-6/14*\log_2(6/14))=0.019$

$GR(method)=0.151/(-(7/14)*\log_2(7/14)-(7/14)*(\log_2(7/14))=0.151$

$GR(man)=0.048/(-(8/14)*\log_2(8/14)-(6/14)*(\log_2(6/14))=0.048/0.986=0.049$

Since material has the highest information gain ratio among all attributes, it is chosen for the tests of attributes and nodes are established accordingly. Others will be manipulated with trimming calculation or depth control for the purpose of decision tree categorization.

(2). Neural network algorithm

Neural network is a network consisting series of neuron nodes and weighed links. The multiple layers of neural network are categorized as input layer, hidden layer and output layer, and the neurons between layers are fully and forward connected. The algorithm is shown as follows [5]:

Assume that there is only one hidden layer in the neural network, and there are i neurons in input layer and j neurons in hidden layer. The weights are shown as W_{ij} and allowable errors as θ_j.

Step1: Set the network parameters, including the number of neurons in every layer, learning rate l, number of training and allowable errors θ_j.

Step2: Set the initial weight of network, W_{ij}, which is randomly generated.

Step3: Normalize the input and output vectors based on the domain of neuron transformation function. The neuron transformation function is $f(x)=1/(1+e^{-x})$.

Step4: Calculate the output of hidden layer and output layer.
The input vector of neuron j $I_j=\sum W_{ij}*O_j+\theta_j$
The output vector of neuron j $O_j=1/(1+e^{-I_j})$

Step5: Calculate the errors of hidden layer and output layer.
The output layer error is
$Err_j=(O_j(1-O_j)(T_j-O_j))$
The hidden layer error is
$Err_j=(O_j(1-O_j)\sum Err_k*W_{kj})$
Where W_{kj} is the weight between neurons k and j in the next higher layer, and Err_k is the error of neuron k, and is the actual output of j based on the given known category label (flag).

Step6: Calculate the correction for network weights and correct.
Correction for network weights ($W_{ij}=W_{ij}+1*Err_j*O_i$)
Correction of error ($\theta j=\theta j+1*Err_j$)

Step7: Go back to Step 3 and repeat the calculation until the correction is less than allowable error or the given number of training.

Fig. 4 is an example of multi-layer forward-feeding neural network, assuming the learning rate is 0.9.

(a). The initial weight and deviation of that network and the 1st training sample X=(1,0,1), as shown in Table 4, where learning rate is l, allowable error θ_j(j=4,5,6), and the initial weight W_{ij}(i=1,2,3,4,5;j=4,5,6) is randomly generated.

(b). Give the 1st training sample X. First, provide the sample to network and calculate the net input and output of each neuron. These values are given in Table 5. Calculate the output of hidden layer and output layer with the transformation function $f(x)=1/(1+e^{-x})$.
The input vector of neuron j$I_j=\sum W_{ij}*O_j+\theta_j$
The output vector of neuron j $O_j=1/(1+e^{-I_j})$
$I_4=\sum W_{ij}*O_i+\theta_j==-0.7$
$O_4=1/(1+e^{0.7})=0.332$
$I_5=\sum W_{ij}*O_i+\theta_j=0.1$
$O_5=1/(1+e^{-0.1})=0.525$
$I_6=-0.105$
$O_6=1/(1+e^{0.105})=0.474$

(c). Now, calculate the error of each neuron and pass on backward. These errors are given in Table 6.
The error of output layer is $Err_j=O_j(1-O_j)(T_j-O_j)$ and the error of hidden layer is $Err_j=(O_j(1-O_j)\sum Err_k*W_{kj})$, where W_{kj} is the weight between neurons k and j in the next higher layer, and Err_k is the error of neuron k, and is the actual output of j based on the given known category label (flag).
$Err_6=(O_6(1-O_6)(T_6-O_6))=0.1311$
$Err_5=(O_5(1-O_5)\sum Err_6*W_{65})=-0.0065$
$Err_4=(O_j(1-O_j)\sum Err_6*W_{64})=-0.0087$

(d). The updated weights and deviations are given in Table 7. The correction of error is ($\theta_j=\theta_j+L*Err_j$) and the correction of network weight is ($W_{ij}=W_{ij}+L*Err_j*O_i$)
$\theta_6=\theta_6+L*Err_6=0.1+0.9*(0.1311)=0.218$
$\theta_5=\theta_5+L*Err_5=0.2+0.9*(-0.0065)=0.194$
$\theta_4=\theta_4+L*Err_4=-0.408$
$W_{46}=W_{46}+L*Err_6*O_4=-0.261$
$W_{56}=-0.2+0.9*0.1311*0.525=-0.138$
$W_{14}=0.2+0.9*(-0.0087)*1=0.192$
$W_{15}=-0.3+0.9*(-0.0065)*1=-0.306$
$W_{24}=0.4+0.9*(-0.0087)*0=0.4$
$W_{25}=0.1+0.9*(-0.0065)*0=0.1$
$W_{34}=-0.5+0.9*(-0.0087)*1=-0.508$
$W_{35}=0.2+0.9*(-0.0065)*1=0.194$

(e). Repeat these steps until the learning is completed.

5. Practice

5.1. System environment

The environment for the practices of product quality problem analysis system can be divided into the following parts:

1. MES Server: recording daily production transactions.
2. ERP server: recording daily transactions in the operation processes of the corporation.
3. DataMart Server: integrating the data guidelines and data of the ERP and MES related to analysis system.
4. Data Mining Server: the operation environment for the data mining system.
5. OLAP Server: capturing data from the DataMart Server for OLAP analysis.
6. Client PC: displaying in visualization for the users to operate the analysis.

The entire environmental structure of quality problem analysis system is shown in Fig. 5.

5.2. Practice of data mining mode

5.2.1. Decision tree

1. Select the data of quality problem from the established data warehouse and define the attribute variable field and category variable filed. The attribute variables are man, machine, material and method, and the category variables are the quality problems.
2. Set the variables of decision tree algorithm, such as whether to trim the decision tree, the greatest depth of decision tree, the ratio between training group and test group, etc. The data of case study companies from Jan. 2002 to Dec. 2003 have been acquired, 24 months and 9,752 pieces o data in total. The category analysis was carried out in random selection with the training group to test group ratio of 3:1.
3. For the previously set parameters, the actual and estimated matrices of decision tree are as follows. The total number of training group is 7,314 and the number of correct categorization in the test group is 5,851, which is therefore 80% of correctness, as shown in Fig. 6.
4. Finally, for the previously set parameters, the predicted estimate and verification were carried out in decision tree algorithm based on the random selection of test group, and the total number of test group is 2,438. The actual and estimated matrices are listed as follows. The number of correct categorization in the test group is 1,682, yielding a percentage of correctness of 69%, as shown in Fig. 7.

5.2.2. Neural network

1. Establish single format database with the data of quality problems and define the attribute variable field and category variable filed.
2. Set the parameters of neural network algorithm, such as learning rate, number of hidden layer and the ratio between training group and test group, etc. The data of case study companies from Jan. 2002 to Dec. 2003 have been acquired, 24 months and 9,752 pieces o data in total. The neural network category analysis was carried out in random selection with the training group to test group ratio of 3:1.
3. For the previously set parameters, the actual and estimated category matrices and the percentage of correctness for the learning simulation in the training group of neural network are shown in Fig. 8. The total number of data is 9,752. The total number of training group is 7,314 and the number of correct categorization in the test group is 4,535, which is 62% of correctness.
4. Finally, for the previously set parameters, the predicted estimate and verification were carried out in neural network algorithm based on the random selection of test group, and the actual and estimated matrices are listed as follows. The total number of test group is 2,438 and the number of correct categorization in the test group is 1,292, yielding a percentage of correctness of 53%. The actual and estimated matrices of neural network are shown in Fig. 7.

5.3. Effect analysis

According to the established database, the effects that decision tree and neural network have accomplished are:

1. Reduction in the number of defected inner leads:

Referring to the number of defected inner leads before using the system from Jan. to Jun. 2004 and that after using the decision tree and neural network from Jul. to Dec. 2004, as shown in Table 8, the total average improvement percentage of using decision tree is 10.4%, while that of using neural network is 2.95%, and the statistics curves of the number of defected inner leads before and after the improvement are shown in Fig. 10.

2. Reduction in the number of defects of resin problem:

Referring to the number of defected resins before using the system from Jan. to Jun. 2004 and that after using the decision tree and neural network from Jul. to Dec. 2004, as shown in Table 9, the total average improvement percentage of using decision tree is 9.24%, while that of using neural network is 4.4%, and the statistics curves of the number of defected resins before and after the improvement are shown in Fig. 11.

3. Reduction in the number of defected chips:

Referring to the number of defected chips before using the system from Jan. to Jun. 2004 and that after using the decision tree and neural network from Jul. to Dec. 2004, as shown in Table 10, the total average improvement percentage of using decision tree is 11.6%, while that of using neural network is 4.6%, and the statistics curves of the number of defected chips before and after the improvement are shown in Fig. 12.

The categorization analysis is carried out for the quality problems of case study companies with data mining decision tree and neural network. The total number of data is 9,752 and the data are divided into training group and test group by random selection with the ratio of 3:1. The number of training group is 7,314 and that of test group is 2,438. For the categorization analysis carried out in decision tree, the number of correct data in the training group is 5,851 with a percentage of 805, and the number of correct data in the test group is 1,682 with a percentage of 69%; For the categorization analysis carried out in neural network, the number of correct data in the training group is 4,535 with a percentage of 62%, and the number of correct data in the test group is 1,292 with a percentage of 53%. Also, after the use of decision tree and neural network in Jul. to Dec. 2004, the total average improvement percentage of the number of defected inner leads is 10.4% using decision tree and 2.95% using neural network; the total average improvement percentage of the number of defected resins is 9.2% using decision tree and 4.4% using neural network; and the total average improvement percentage of the number of defected chips is 11.6% using decision tree and 4.6% using neural network. Therefore, it is safe to say that decision tree is more suitable for the categorization analysis for quality problems, as shown in Table 11.

6. Conclusion

The conclusion of this research provides the quality problem analysis modes for LCD driver IC packaging factories with the studies of case companies. For the LCD driver IC packaging factories, to increase the qualification percentage of products, to rapidly understand the cause of product defects and to find out the solution, the establishment of a product quality problem analysis mode enables more effective operations, faster acquisition of information really needed and the improvement in the satisfaction of customer in information system.

For the effects of quality problem analysis, the total average improvement rate for the number of defected inner leads is 10.4%, 9.2% for the number of defected resins and 11.6% for the number of defected chips, using decision tree; 在 the total average improvement rate for the number of defected inner leads is 2.95%, 4.4% for the number of defected resins and 4.6% for the number of defected chips, using neural network. The decision tree is obviously more suitable than neural network in quality problem categorization analysis.

This method can be used in quality problem analysis for computer assembling, electrical component fabrication and semi-conductor packaging and fabrication industries.