Applying data mining methodology to establish an intelligent decision system for PCBA process

Chien-Yi Huang
Department of Industrial Engineering and Management, National Taipei University of Technology, Taipei, Taiwan
Marvin Ruano
Department of Management, National Taipei University of Technology, Taipei, Taiwan
Ching-Hsiang Chen
A Test Lab Techno Corp. Taipei, Taiwan, and
Christopher Greene
Binghamton University, Binghamton, New York, USA

Abstract

Purpose – This paper aims to consider the practical production environment of electronics manufacturing industry firms, and the large quantities of information collected on machine processes, testing data and production reports, while simultaneously taking into account the properties of the processing environment, in conducting analysis to obtain valuable information.

Design/methodology/approach – This research constructs a prediction model of the circuit board assembly process yield. A decision tree is used to extract the key attributes. The authors also integrate association rules to determine the relevance of key attributes of undesirable phenomena.

Findings – The results assure the successful application of the methodology by reconfirming the rules for solder skip and short circuit occurrence and their causes.

Originality/value – Measures for improvement are recommended, production parameters determined and debugging suggestions made to improve the process yield when the new process is implemented.

Keywords Data mining, *A priori*, Printed circuit board assembly, Process defective

Paper type Research paper

1. Introduction

In recent years, the development of intelligent factories has become mainstream in global manufacturing. Industry 4.0 has changed the distribution of work in production and product integration models and lead to integration of the full range of the industrial value chain (Lin *et al.*, 2016; Kune *et al.*, 2016). The global electronics industry is booming. Presently, market demand for electronic goods is trending toward lighter, thinner and smaller. Surface mount technology (SMT) has become the key process of printed circuit board assembly (Huang *et al.*, 2011). Important steps in the manufacturing process include: stencil printing, coating the solder paste on the bonding pad on the PCB, and using automated equipment to precisely place the parts on the PCB (Lau *et al.*, 2016; Borecki and Serzysko, 2016; Huang and Huang, 2014). Next, reflow soldering is carried out. The solder paste is melted and then cooled to form the ideal solder joint to join the component pin or solder ball to the PCB (Tsai, 2011; He *et al.*, 2014; Merkle and Götzen, 2015; Huang, 2015). The final step is using automated optical inspection (AOI) to visually inspect the final assembly of PCB components for defects. An in-circuit test is then conducted to determine whether the electronic components exhibit failures after high temperature soldering.

At present, large data analysis for electronics manufacturing industry applications includes: production scheduling, fault diagnosis systems, quality control, process improvement, defect analysis, and yield improvement. Srimani and Prathiba (2016) developed a PCB defect detection and classification model, through a hybrid approach for feature reduction and classification. A genetic algorithm and MATLAB 2013 b were used for data preprocessing to achieve the feature reduction and confidence measurement. With rising environmental awareness in recent years, green packaging has become popular, along with lead-free processing. Thus, parameter setting and materials selection are more complicated than in the past (Bath *et al.*, 2007; Huang *et al.*, 2012). Failure to comply with quality control standards for circuit board assembly must be addressed by additional inspection and work, which can increase the total production cost by roughly 50 per cent (Tsai and Tsai, 2014; Soto, 1998). Thus, manufacturers must focus on enhancing the manufacturing process. However, data generated at the process and inspection machines in the current production environment are only used for auditing or shipment inspection. Subsequent analysis of this data to turn it

The current issue and full text archive of this journal is available on Emerald Insight at: **www.emeraldinsight.com/0954-0911.htm**

Soldering & Surface Mount Technology
31/4 (2019) 271–278

[DOI 10.1108/SSMT-10-2018-0036]

Received 4 October 2018
Revised 11 February 2019
17 March 2019
Accepted 17 April 2019

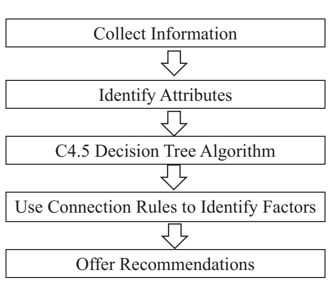
into valuable information is seldom performed. In the SMT process, 52 to 71 per cent of the solder defects originate in inappropriate solder paste deposition process (Wang, 2013; Zhou *et al.*, 2010). In recent studies, Chen *et al.* (2016) tried to minimize the maximum warpage of shielding frame by controlling reflow soldering control parameters. A surrogate-assisted optimization called adaptive Kriging high-dimensional representation model is used to solve the high-dimensional problem. The warpage of shield frame is significantly reduced. Moreover, Roman (2017) investigated the effect of solder alloying with a small amount of lanthanum and yttrium on bond formation with the Si and Cu substrates. Soldering was performed using a fluxless method in air and with ultrasonic activation. He found that in the process of ultrasonic soldering, the lanthanum and yttrium were distributed at the interface with Si and Cu substrates, which enhanced the bond formation. In previous studies, Qiao *et al.* (2012) observed that the miniaturized passive components provided in the tape, may become attached to the tape sealant due to electrostatic effects, resulting in placement error. Though some of these studies may identify factors that affect the quality of the production process, they generally explore only a single machine and do not analyze the entire production line to suggest comprehensive improvement measures. Cheng *et al.* (2008) also used a classification and regression tree (CART) to construct an SMT solder paste printing process quality control forecasting model. To determine the impact of solder paste printing volume, important key process attributes were defined as input variables, while the quality of the solder paste volume was used as the output variable to explore important attributes of the spacing, stencil thickness and squeegee material as key characteristics. However, scholars have rarely considered the attributes of the production environment. Further, when the input attribute is treated as a binary variable for analysis, the loss of some attribute information can occur.

This research examines the realistic production environment of the manufacturer, collects machine process data, inspection data, and big data from production reports, takes into account process environment-related attributes and analyzes this body of data to obtain valuable information. The purpose of this study is to construct a model of yield prediction, extract key attributes, and determine the correlation between the key attributes and causes of defects. Improvement measures are then suggested for use with the production parameters determined in this study, and debugging suggestions are offered.

2. Materials and methods

The process of this research is illustrated in Figure 1. First, we explore the manufacturing process and collect production information in the form of daily production reports, the first article confirmation report, raw material control report, fixture control report, and the production environment temperature and humidity records. We then obtain the production volume, the yield rate, the process parameter settings, material, and environmental information, and determine the factors that affect the yield rate. We first use a C4.5 decision tree algorithm to construct the prediction model for the yield prediction and identify key attributes which affect the process yield. The C4.5 algorithm is an algorithm used in data mining as a decision tree classifier which can be employed to generate a decision, based on a certain sample of data (Quinlan, 1993; Witten *et al.*, 2011). Then, we use the association rules to determine the correlation between the key attributes and product defects. Finally, we integrate the decision tree and the association rules to analyze the causes of problems in the production process, and construct intelligent decision-making rules and offer recommendations for production decision-making.

Figure 1 Research process



2.1 Information collection and production attributes

This research explores the domestic electronics assembly industry. It collects a wide range of production process attributes from quality control staff and AOI from the solder printing machine to the solder reflow oven. The reflow oven used is an infrared heater with ten heating zones. The assembled components include various sizes of capacitors and resistors, and 208 pin quad flat packages with a lead pitch of 0.5 mm.

Each case of PCB testing after the AOI and reflow soldering includes insufficient solder, solder gaps, bridges, floating, component offset, and other solder defects. The information collected on each firm's production environment includes each machine's work orders and the reports filed by each worker, as follows:

- *Daily production reports* – These reports currently use International Organization for Standardization (ISO) auditing. Among these reports, factors affecting production may include Line, Customer, Number of Components, Yield, Production Time, and Production Period. Three production lines (L1, L2, and L3) are included. L1 is for the pure solder paste process, while L2 and L3 are for the solder paste process and the dispensing process. Materials for the components and PCB are specific to the customer; thus, different work orders for different customers may specify materials of differing quality. Number of components is the number of electronic parts contained in the circuit board. Yield is the number of products processed in SMT process. Production time is the time required to complete the work order, defined as end time minus start time in minutes. Production period refers to the SMT process shift, including the following shifts: 00 a.m. – 8 a.m., midnight shift (MS); 8 a.m. – 4 p.m., day shift (DS); 4 p.m. – 00 a.m., night shift (NS); and all day. Because each shift has different individuals responsible, their differing physical and mental attributes may affect production quality.
- *SMT First Article Assurance Form* – The testing of a PCB board must first be conducted to confirm the process parameters and quality stability, then mass production can begin. Related factors include: Stencil Printing Speed (mm/s) – The speed of the squeegee on the stencil

pushing the solder paste into the stencil aperture, the impact of the squeegee, and stencil aperture contact time may affect the amount of solder; Wafer Wiping Frequency (Piece/Time) – After each printing, the residual solder may accumulate in the wall of the stencil, the corner, or the back, and be transferred to the next printed circuit board surface, or result in solder deficiency; Paste Type – Sn63/Pb37 solder paste and lead-free solder paste with an alloy composition of SAC305 (Sn96.5%/Ag3.0%/Cu0.5%) were considered; Reflow Oven Conveyor Speed (cm/min) – Too fast may lead to uneven heating of solder paste, resulting in cold soldering, non-melting of solder, tombstones, and other phenomena. If the speed is too slow, components may suffer damage due to long heating time; and High/Low Temperatures in Differing Areas of the Reflow Oven (°C) – Too low a temperature may cause solder skips and poor soldering, too high a temperature may damage components.

- *Solder paste report* – Solder paste not opened can be placed in the refrigerator before placement in cold storage. The paste control report records the time from removal from the refrigerator to the opening of the can and use. The relevant properties are as follows: Solder Paste Temperature – Fridge out to the time before the cans are opened and used, usually more than 4 hours; Solder Paste Warm-Up Before Printing – Reduce temperature differences, and prevent moisture condensation caused by solder voids; and Solder Paste Re-Use Decision – Recover extra solder paste resulting from production line stoppages on holidays and return it to the refrigerator. However, solvent evaporation within the solder paste may result in solder skips, tombstoning, and short circuits.
- *Ambient Temperature Report* – This report consisted of daily morning and afternoon records of the temperature of the solder paste refrigerator, and monitoring of the ambient temperature and humidity of the production site. Relevant properties include Refrigerator Temperature (°C) and Ambient Temperature (°C), Humidity (% RH). Refrigerator Temperature (°C) – ISO specifies a temperature range of 5°C±3°C. Too high a temperature may cause the alloy powers in the solder paste to undergo chemical reactions, increasing the viscosity and affecting printing. Too low a temperature prevents the rosin ingredients from crystallizing, resulting in solder paste deterioration. Ambient Temperature (°C), Humidity (% RH) – ISO specifies a temperature range of 24°C±5°C and a humidity of <70% RH. When the ambient temperature is too high, solder paste located on the stencil may severely oxidize, resulting in poor soldering. If the solder paste absorbs moisture, it may exhibit solder voids.

2.2 Data exploration

This study integrates the decision tree and the association rules to perform the data exploration. Since the former has the ability to clearly indicate the best attributes, this research uses feature extraction to identify key attributes. However, when the decision tree is too large, the error rate increases, resulting in an inability to display the degree of influence each attribute has on undesirable phenomena. Therefore, the association rules are applied in the study to show the expected yield rate under the combination of specific attribute conditions. The weakness of the association rules is that they have too many rule categories. The decision tree helps narrow the scope of decision-making.

2.2.1 Constructing the decision tree model

In the study, a total of 984 pieces of production data were collected over three months of the SMT process, as shown in Table I. Considering the attributes of the parameters collected in the daily on-site reports, and using a negative number to represent the number of quality characteristics such as solder skips and short circuits, the C4.5 algorithm is used to construct the prediction model, and the key factors in poor processing are identified.

The production data of Table I were randomly divided into 10 equal parts, 9 of which were used as the training materials. The remaining part was used as the test data for analysis by the C4.5 algorithm of the SAS Enterprise Miner software decision tree. First, we used Formula (1) to calculate the total amount of information Info (D), in which the probability of occurrence of each category is $P_j = X_j/N$, where X_j is the number of occurrences without adverse phenomena, and N is the data number. Next, information bifurcation of attribute A is performed. Formula (2) is used to calculate the information amount $Info(A_i)$ of attribute A_i, where X_{i1} is the total amount of information divided by the attribute values A_i, and X_{i1}, and X_{i2} are the attribute values of A_i for the occurrence and nonoccurrence of defect numbers, respectively. We then used Formula (3) to calculate the amount of information for attribute A, $Info_A(D)$. Finally, the sum of the amount of information $Info_A(D)$ minus the amount of information after the branch $Info_A(D)$ becomes the degree of contribution of attribute A (Formula 4). The above steps are repeated for each attribute. We select the attribute with the largest information gain as the attribute branch (Yeh *et al.*, 2011; Huang and Lin, 2013):

$$\mathrm{Info(D)} = -\sum_{j=1}^{k} P_j \times \log_2(P_j) \tag{1}$$

$$\mathrm{Info}(A_i) = -\frac{X_{i1}}{X_{i\cdot}}\log_2\left(\frac{X_{i1}}{X_{i\cdot}}\right) - \frac{X_{i2}}{X_{i\cdot}}\log_2\left(\frac{X_{i2}}{X_{i\cdot}}\right) \tag{2}$$

$$\mathrm{Info_A}(D) = \sum_{i=1}^{l} \frac{X_{i\cdot}}{N} \times \mathrm{Info}(A_i) \tag{3}$$

$$Gain(A) = \mathrm{Info(D)} - \mathrm{Info_A}(D) \tag{4}$$

3. Results and discussion

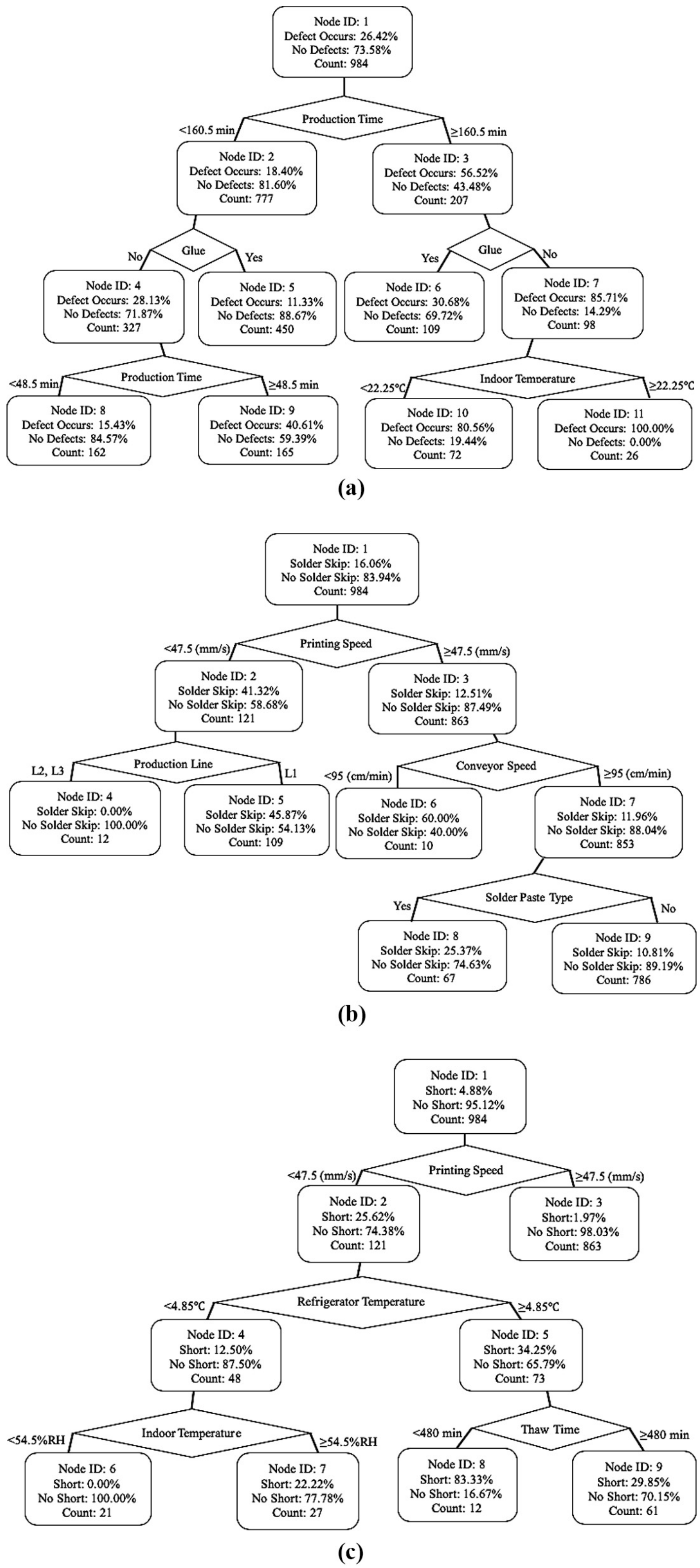
Figure 2 (a) shows that the key attributes of the defect phenomena include production time, solder dispensing, and ambient temperature. The negative number in the yield prediction model is the sum of the number of defect occurrences, but the reasons for the defects are different. Therefore, the prediction model is established for solder skips

Table I Mass production information table

Basic information				Production environment			Attributes Solder paste operations				Printing parameters				Reflow oven		Quality characteristics Defect count		
Customer	Production volume	No. of components	Processing time	Shift	Ambient temperature	Ambient humidity	Refrigerator temperature	Solder paste	Warm-up time	Reuse	Production line	Printing speed	Cleaning frequency (Wipes)	Conveyor speed	Assembly process highest temperature	Assembly process lowest temperature	Solder skip	Short circuit	PPM
A	240	14	230	DS	20.4	68	5.7	Lead-Free	140	N	L3	100	10	100	260	110			0
B	500	34	110	DS	29.3	69	5.5	Lead	120	N	L3	100	10	100	260	110		1	57
A	200	118	90	NS	21.8	73	5	Lead-Free	155	N	L2	90	10	100	275	140	3	1	163
A	180	74	120	NS	24.4	73	4.6	Lead-Free	180	N	L1	100	10	180	285	110			0
B	400	156	630	MS	28.6	71	5.1	Lead-Free	175	N	L1	110	5	100	260	110			0
C	600	7	85	NS	29.1	72	4.9	Lead	195	Y	L2	100	5	100	290	140			0
C	3000	53	15	NS	25.4	70	5.1	Lead-Free	155	Y	L2	100	10	180	285	110	3		1000
D	2000	8	80	DS	21.9	68	4.8	Lead-Free	160	Y	L3	100	5	100	275	160			0
A	900	18	20	MS	30.0	67	5.4	Lead-Free	125	Y	L1	90	10	100	290	110			0
C	300	3	50	MS	20.7	66	5.6	Leaded	195	N	L3	100	10	100	260	110	2		467
D	400	26	75	MS	21.5	71	5.1	Lead	200	N	L2	100	10	180	275	140	4		165
B	1000	66	28	MS	23.4	72	4.6	Lead-Free	130	N	L1	100	5	180	290	110	1	1	205
A	500	23	30	MS	25.6	71	5	Lead	180	N	L3	100	10	100	260	140			0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Soldering & Surface Mount Technology
Volume 31 · Number 4 · 2019 · 271–278

Figure 2 Defect frequency models



Notes: (a) Overall defect frequency model; (b) solder skip frequency model; (c) short circuit frequency model and short circuit. Figure 2 (b) shows the solder skip prediction model. The results show that the key attributes of poor soldering include printing speed, production line, reflow conveyor speed and solder paste type. Figure 2 (c) shows the short circuit quality prediction model. The key attributes of short circuits include printing speed, refrigerator temperature, ambient humidity and solder paste thaw time.

3.1 Constructing the associated rules model

The K-means method was used to group the key attribute values. The original continuous variables were changed into discrete variables. The grouped key attribute values for solder skips and short circuits were then divided as shown in Table II(a) and (b).

This study uses the *a priori* algorithm of SAP Predictive Analytics software to construct association rules, which are algorithms for mining association rules for high frequency item sets (Chen and Weng, 2008; Hu *et al.*, 2003; Han, 2001). The implementation procedure is as follows: First, we define the minimum support (Support) and the minimum confidence (Confidence). We then find the first-order project set and calculate the support of the project and the minimum support, and then perform a comparison, assuming that the highest value for support is greater than the minimum support threshold. Such items are removed and treated as the first order high frequency item set. The first-order high-frequency project set is then combined to obtain the second-order project set. We calculate its support and compare it with the smallest support, and then produce the second-order high-frequency project set. Finally, to determine whether to search all the item sets, we calculate the collection of the project set of the reliability and gain value (Gain). Based on the gain value ranked from largest to smallest, we find the significant association rules.

3.1.1 Application results for solder skip association rules

There are a total of 15 rules of association for solder skips, based on the ranking from greatest to smallest gain. Table III shows the results for these rules, which are summarized as follows:

- If the speed of the reflow conveyor is greater than 106 cm/min, solder skips are likely to occur. When the reflow conveyor is too fast, the solder paste does not receive enough heat, causing uneven heating. Improvement measures include reducing the speed of the reflow conveyor so that the solder paste has enough time to heat.
- When the stencil printing speed is greater than 46 mm/s, solder skips are prone to occur. When the printing speed is too fast, less paste is deposited. Improvement measures include reducing the printing speed of the stencil to increase the amount of solder.
- Lead-free solder paste is also prone to produce solder skips. SAC305 lead-free solder paste has a higher melting point, and solder skips occur when the reflow temperature is not high enough. Improvement measurers include evaluating the use of low temperature melting lead-free solder paste, or moderately adjusting the reflow temperature curve to ensure that the solder is fully melted during the reflow process.

Table II Key attribute groups

Attribute name	Attribute value
(a) Solder skip	
1 Printing Speed (mm/s)	<25, 26-45, 46-60, >61
2 Production Line	L1, L2, L3
3 Conveyor Belt Speed (cm/min)	<90, 91-105, 106-120, >121
4 Solder Paste Type	Leaded, Lead-free
(b) Short circuits	
1 Printing Speed (mm/s)	<20, 21-35, 36-50, 51-65, >66
2 Refrigerator Temperature (°C)	<4.2, 4.3-5.25, 5.26-6.2, >6.3
3 Ambient Humidity (% RH)	<45, 46-51.5, 51.6-56.5, 56.6-60.5, >60.6
4 Thaw time (min)	<450, 451-650, 651-980, 981-1280, >1281

3.1.2 Application results for short circuit association rules

There are a total of 15 association rules for short circuits, ranked from largest gain to smallest. These rules are shown in Table IV and summarized as follows:

- When the stencil printing speed is less than 35 mm/s, short circuits are prone to occur. If the printing speed is too slow, although there is plenty of time for the solder paste to fill, the contact time between the steel sheet and the PCB contact time is too long, resulting in the solder paste moving to the opposite side of the stencil, leading to the separation of the stencil and the solder paste, and in turn to solder balling and short circuits. Improvement measures include enhancing the steel printing speed and shortening the stencil and PCB contact time to reduce the occurrence of solder balls.
- Indoor humidity greater than 56.5 (% RH) can cause short circuits. When the solder paste absorbs moisture from the environment, reflow leads to solder paste spattering. Improvement measures include setting up a hygrometer to control the ambient humidity in the production environment.
- When the solder paste warm up is less than 450 (min) short circuits are prone to occur. When the temperature of the solder paste from the refrigerator is lower than the room temperature, if it is not properly warmed, it will easily absorb moisture from the environment, leading to spattering in the reflow process and solder beads. Improvement measures include placing the solder paste at room temperature for at least 4 hours to fully warm up, after removing it from the refrigerator. The thawing time should be indicated on the solder paste tank and recorded in the solder paste control report.

4. Conclusions

This study collects a huge amount of data in the manufacturing process environment from the inspection data and production reports. It discusses the process-related attributes of the process environment, identifies data on defects obtained by the AOI machine and the quality control personnel, and then integrates the decision tree and the association rule-based method for data exploration. First, the decision tree C4.5 algorithm is used to extract the features to determine the key attributes. The association rules for the solder skip prediction model include

Soldering & Surface Mount Technology
Volume 31 · Number 4 · 2019 · 271–278

Table III Solder skip association rules

Input variable	Target variable	Support level	Confidence	Gain
1 {Conveyor Speed => 121}		0.00	1.00	6.23
2 {Solder Paste = Lead-Free}and{Conveyor Speed = 106-120}		0.00	1.00	6.23
3 {Printing Speed => 61}		0.00	1.00	6.23
4 {Printing Speed => 61}and{Conveyor Speed => 121}		0.00	1.00	6.23
5 {Printing Speed = 46-60}and{Conveyor Speed = 106-120}		0.04	0.51	3.20
6 {Solder Paste = Lead-Free}and{Conveyor Speed => 121}		0.04	0.47	2.95
7 {Printing Speed = 46-60}		0.04	0.42	2.62
8 {Solder Paste = Lead-Free}	Solder Skip Occurrence	0.04	0.40	2.46
9 {Printing Speed = 46-60}and{Conveyor Speed => 121}		0.04	0.39	2.43
10 {Conveyor Speed = 106-120}		0.04	0.39	2.42
11 {Printing Speed = 46-60}and{Solder Paste = Lead-Free}		0.04	0.37	2.30
12 {Conveyor Speed = 91-105}		0.02	0.34	2.12
13 {Production Line=L1}and{Printing Speed = 46-60}		0.02	0.30	1.87
14 {Production Line=L1}		0.02	0.28	1.75
15 {Production Line=L1}and{Solder Paste = Lead-Free}		0.06	0.17	1.08

Table IV Short circuit association rules

Input variable	Target variable	Support level	Confidence	Gain
1 {Printing Speed =< 20}		0.00	0.27	5.59
2 {Printing Speed = 21-35}		0.00	0.24	5.01
3 {Ambient Humidity => 60.5}		0.00	0.22	4.59
4 {Printing Speed =< 20}and{Ambient Humidity => 60.5}		0.00	0.20	4.04
5 {Printing Speed = 21-35}and{Ambient Humidity = 56.5-60.5}		0.00	0.18	3.66
6 {Warm Up Time =< 450}		0.02	0.09	1.76
7 {Refrigerator Temperature = 5.26-6.2}	Short Circuit Occurrence	0.01	0.07	1.51
8 {Ambient Humidity = 56.5-60.5}		0.02	0.07	1.40
9 {Warm Up Time =< 450}and{Ambient Humidity => 60.5}		0.02	0.07	1.40
10 {Printing Speed = 21-35}and{Warm Up Time =< 450}		0.01	0.07	1.36
11{Printing Speed = 21-35}and{Ambient Humidity = 46-51.5}		0.01	0.06	1.20
12 {Warm Up Time =< 450}and{Ambient Humidity = 46-51.5}		0.01	0.06	1.17
13 {Ambient Humidity = 46-51.5}		0.04	0.05	1.12
14 {Refrigerator Temperature = 4.3-5.25}		0.01	0.05	1.05
15 {Printing Speed = 21-35}and{Refrigerator Temperature = 4.3-5.25}		0.01	0.05	1.02

stencil printing speed, production line, reflow conveyor speed and solder paste type. The association rules for the short circuit prediction model include stencil printing speed, refrigerator temperature, ambient humidity, and warm up time. Then, the association rules with the *a priori* algorithm are used to construct the model. The successful application of the methodology is assured by reconfirming the rules for solder skip and short circuit occurrence and their causes. Finally, improvement measures like production parameters and debugging advice were offered for incorporation into the manufacturing process to raise yield rates.

References

Bath, J., Jay, R., Bennett, L., Tan, S.K., Chih, P.W. and Lucuta, A. (2007), "Repeatability analysis of EDXRF equipment for RoHS compliance screening for soldering materials used in PCBA manufacturing", *Electronic Manufacturing Technology Symposium*, San Jose, CA, pp. 339-145.

Borecki, J. and Serzysko, T. (2016), "Mechanical reliability of solder joints in PCBs assembled in surface mount technology", *Soldering & Surface Mount Technology*, Vol. 28 No. 1, pp. 18-26.

Chen, L., Li, E. and Wang, H. (2016), "Time-based reflow soldering optimization by using adaptive Kriging-HDMR method", *Soldering & Surface Mount Technology*, Vol. 18 No. 2, pp. 101-113.

Chen, Y.L. and Weng, C.H. (2008), "Mining association rules from imprecise ordinal data", *Fuzzy Sets and Systems*, Vol. 159 No. 4, pp. 460-474.

Cheng, W., Zhang, X., Wang, K. and Dai, X. (2008), "Integrating classification and regression tree (CART) with GIS for assessment of heavy metals pollution", *Environmental Monitoring and Assessment*, Vol. 158 Nos 1/4, pp. 419-431.

Han, J. (2001), *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Francisco.

He, X., Liu, Z., Cai, J., Chen, Y., Tan, L. and Wang, Q. (2014), "Characterization of stencil printing parameters for fine pitch wafer bumping", *Electronic Packaging Technology (ICEPT), Chengdu*, pp. 80-85.

Hu, Y.C., Chen, R.S. and Tzeng, G.H. (2003), "Discovering fuzzy association rules using fuzzy partition methods", *Knowledge-Based Systems*, Vol. 16 No. 3, pp. 137-147.

Huang, C.Y. and Huang, H.H. (2014), "Process optimization of SnCuNi soldering material using artificial parametric design", *Journal of Intelligent Manufacturing*, Vol. 25 No. 4, p. 813.

Huang, C.Y. and Lin, Y.H. (2013), "Applying CHAID algorithm to investigate critical attributes of void formation in QFN assembly", *Soldering & Surface Mount Technology*, Vol. 25 No. 2, pp. 117-127.

Huang, C.Y. (2015), "Innovative parametric design for environmentally conscious adhesive dispensing process", *Journal of Intelligent Manufacturing*, Vol. 25 No. 1, pp. 1.

Huang, C.Y., Huang, H.H. and Ying, K.C. (2012), "Sn-Cu-Ni soldering process optimization using multivariate analysis", *IEEE Transactions on Components, Packaging and Manufacturing Technology*, Vol. 2 No. 3, pp. 527-535.

Huang, C.Y., Lin, Y.H., Ying, K.C. and Ku, C.L. (2011), "The solder paste printing process: critical parameters, defect scenarios, specifications, and cost reduction", *Soldering & Surface Mount Technology*, Vol. 23 No. 4, pp. 211-223.

Kune, R., Konugurthi, P.K., Agarwal, A., Chillarige, R.R. and Buyya, R. (2016), "The anatomy of big data computing", *Software: Practice and Experience*, Vol. 46 No. 1, pp. 79-105.

Lau, C.S., Khor, C.Y., Soares, D. and Teixeira, J.C. (2016), "Thermo-mechanical challenges of reflowed lead-free solder joints in surface mount components: a review", *Soldering & Surface Mount Technology*, Vol. 28 No. 2, pp. 41-62.

Lin, J.W., Selvi, A. and Arul, J. (2016), "A performance-satisfied and affection-aware MapReduce allocation scheme for intelligent information applications", in *Knowledge Engineering and Applications (ICKEA), IEEE International Conference*, pp. 46-51.

Merkle, T. and Götzen, R. (2015), "Millimeter-Wave surface mount technology for 3-D printed polymer multichip modules", *IEEE Transactions on Components, Packaging and Manufacturing Technology*, Vol. 5 No. 2, pp. 201-206.

Qiao, C., Shi, Y., Vicera, N.G., Poon, M., Li, W., Chen, H. and Wu, J. (2012), "Improvement of pick & place yield in carrier tape packaging system through materials selection and cavity structure optimization", *Electronic Materials and Packaging (EMAP), Lantau Island*, pp. 1-4.

Quinlan, J.R. (1993), *C4.5: Programs for Machine Learning*, San Francisco: Morgan Kaufmann.

Roman, K. (2017), "Study of direct soldering of silicon and cooper substrate by use of Bi-based solders with lanthanides addition", *Soldering & Surface Mount Technology*, Vol. 29 No. 3.

Soto, H.P. (1998), "Study of the thermal behavior of a printed circuit board during the reflow process inside a surface mount technology oven", Unpublished master dissertation, American University of PuertoRico.

Srimani, P.K. and Prathiba, V. (2016), "Adaptive data mining approach for pcb defect detection and classification", *Indian Journal of Science and Technology*, Vol. 9 No. 44.

Tsai, T.N. (2011), "Improving the fine-pitch stencil printing capability using the Taguchi method and Taguchi fuzzy-based model", *Robotics and Computer-Integrated Manufacturing*, Vol. 27 No. 4, pp. 808-817.

Tsai, T.N. and Tsai, C.W. (2014), "Development of a closed-loop diagnosis system for reflow soldering using neural networks and support vector regression", *International Journal of Industrial Engineering*, Vol. 21 No. 1.

Wang, S.T. (2013), "Integrating the Taguchi method and the multiattribute decision-making method to optimize the surface mount technology solder paste printing thickness process", *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, Vol. 227 No. 8, pp. 1209-1217.

Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J. (2011), *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. San Francisco: Morgan Kaufmann, San Francisco, p. 191.

Yeh, D.Y., Cheng, C.H. and Hsiao, S.C. (2011), "Classification knowledge discovery in mold tooling test using decision tree algorithm", *Journal of Intelligent Manufacturing*, Vol. 22 No. 4, pp. 585-595.

Zhou, Y., Al-Bassyiouni, M. and Dasgupta, A. (2010), "Harmonic and random vibration durability of SAC305 and Sn37Pb solder alloys", *IEEE Transactions on Components and Packaging Technologies*, Vol. 33 No. 2, pp. 319-328.

Corresponding author

Chien-Yi Huang can be contacted at: jayhuang@ntut.edu.tw