Proceedings of the 8th European Conference on Industrial Engineering and Operations Management
Paris, France, July 2-4, 2025

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/EU08.20250171
Published: July 2, 2025

Association Rule Learning for Predictive Maintenance in Industry AI

Hsiao-Kang Lin and Tzu-Jou Liao
Department of Industrial Management
I-Shou University
Kaohsiung City 84001, Taiwan R.O.C
hklin@isu.edu.tw, hyliao@isu.edu.tw

Abstract

The evolution of industry Artificial intelligence (AI) technology in today's manufacturing system has the potential to be extremely helpful in applications such as quality assurance, process optimization, and predictive maintenance. Particularly, predictive maintenance in manufacturing uses data to forecast potential failures and optimize maintenance schedules, leading to increased efficiency, productivity, and sustainability. Industrial AI is an interdisciplinary area of research, encompassing fields such as Machine Learning (ML), natural language processing and robotics. Association rule learning, a rule-based ML method for discovering interesting relationships between variables in large databases, has been widely used across business intelligent applications for decision-making. However, there are few studies on predictive maintenance in manufacturing systems. This paper proposes Apriori-Based learning in R for providing the predictive analytics to improve fault detection and support predictive maintenance, thereby enhancing production stability and efficiency. The case study is conducted to illustrate the feasibility of the proposed method is demonstrated.

Keywords
Industry Artificial Intelligence (AI), Association Rule Learning, Apriori-Based Learning, arules Package in R, Predictive Maintenance.

1. Introduction
Industrial AI can be defined as a systematic discipline focusing on the development, validation, deployment and maintenance of AI solutions for industrial applications with sustainable performance. Hence, Industrial AI is an interdisciplinary area of research, encompassing fields such as Machine Learning (ML), natural language processing and robotics [Li et al. 2022] [Lee 2014]. Association rule learning is a rule-based ML technique used to discover interesting relationships among transactions encoded in a large database. It is one of the most important techniques of the knowledge discovery and is considered as significant subfield in knowledge and information management. It has been used to improve decision making in a wide variety of business intelligent applications, such as, market basket analysis [Omari et al. 2008], commercial sales analysis [Bing and Li 2009], customs risk management [Wang et al. 2010], and customer purchase intention [Kim et al. 2012] …etc. Association rule mining have also been applied in various smart manufacturing system, e.g. manufacturing resource planning system [Wang 2010] and manufacturing process control [Vazan et al. 2017] …etc. However, there are few studies on association rule mining for predictive maintenance in industry AI.

Association rule mining requires online interactive data mining tools for predictive analytics, analytical modelling and visualization [Peres et al. 2020]. R programming language has always been an outstanding environment for widely available data exploration, statistical computations, regression analysis, visualization and Big Data. Many useful R function come in packages, free libraries of code written by R's active user community. For examples, a series of R packages and an environment for statistical computing with Big Data. Moreover, the *arules* package for R provides the infrastructure for representing, manipulating and analyzing transaction data and patterns (e.g. frequent itemsets and association rules). The arules provides comprehensive functionality for Apriori algorithm. This paper proposes Apriori-Based rules learning in R for providing the predictive analytics to improve fault detection and support predictive maintenance, thereby enhancing production stability and efficiency. A case study is conducted. The main purpose of this study is to explore the association rules that will contain advanced analytics to conduct predictive modelling functionalities. Whether there are related rules for fault anomaly detections, inferring future fault events and diagnosing potential root causes of the problem.

2. Apriori-based learning towards intelligent association rules mining

The Apriori algorithm is the first association rule algorithm that proposed by [Agrawal et al. 1993] [Agrawal and Srikant 1994]. It is a technique used to discover relationships among very large set of variables in databases. Apriori is an algorithm for f*requent item set mining* and *association rule learning*.

2.1 Frequent item set mining

The concept of Apriori algorithm is used to extract Frequent Itemsets (FI) from large database, then to generate association rules from each FI. Let's take the following an example of transaction items as show in Table1 and Figure1.

Table 1. Example for 5 transactions in a database

Transaction ID	Items
1	ACD
2	CE
3	AB
4	ACF
5	ABC

Apriori algorithm is started with a single item set, repeat the steps of generating the candidate itemsets until all the high frequency itemsets are found. Figure 1 shows the sample usage of Apriori algorithm.

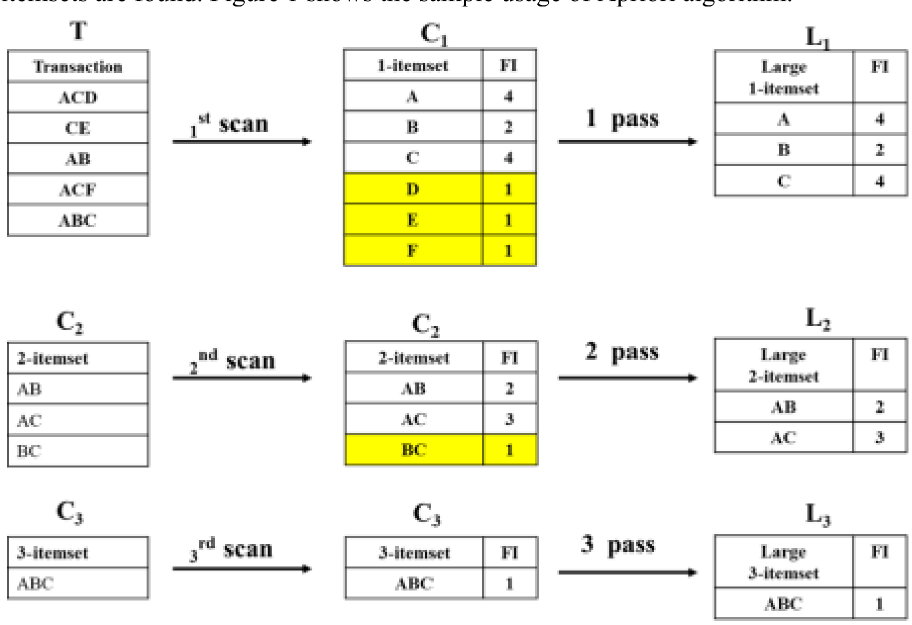

Figure 1. The sample usage of Apriori algorithm

2.2 Association rule learning

Sometimes, the association of the data in the database is not known. Association Rule Learning (also called as Association Rule Mining) is a common technique used to pick out the unknown inter-dependence of the data and finding out the rules between those items. In other words, involve finding the rules that may govern associations and causal objects (antecedent and consequent) between sets of items. In simple words, the rules for association rule analysis refer to: X⇒Y, X is called as antecedent and Y is called as consequent. That is, the rule of if X then Y.

[Agrawal et al. 1993] mentioned that association rules mining is usually required to have *support* and *confidence* greater than or equal to the user-specified minimum support and a minimum confidence respectively. *Support* can be item frequency percentage; *Confidence* is used to judge the possibility of B occurring under the condition of item A.

For rule A⇒B (antecedent ⟹consequent)

✓ Support (A) : the percentage of all transactions in the database containing A.

$$\text{Support (A)} = \frac{\text{A}}{\text{All transaction}} = \frac{4}{5} = 0.8$$

✓ Support (B) : the percentage of all transactions in the database containing B.

$$\text{Support (B)} = \frac{\text{B}}{\text{All transaction}} = \frac{2}{5} = 0.4$$

✓ Support (A⋂B): the percentage of all transactions in the database containing A and B

$$\text{Support (A}\cap\text{B)} = \frac{\text{A} \cap \text{B}}{\text{All transaction}} = \frac{2}{5} = 0.4$$

✓ Confidence (A⇒B): the percentage of those transactions (antecedent support) containing A that also contain B.

$$\text{Confidence (A}\Rightarrow\text{B)} = \frac{\text{Support (A}\cap\text{B)}}{\text{Support (A)}} = \frac{0.4}{0.8} = 0.5$$

However, the *support* and *confidence* are high, does not necessarily mean that the events referred to by this association rule have high correlation with each other. According to **[**Liu et al. 2000**]** [Peng and Zhou 2012]**,** association-rule-mining algorithms tend to produce huge numbers of rules, most of which are of no interest. Users have considerable difficulty analyzing so many rules to identify the truly interesting ones. Lift is a measure of the strength of the association between two items, taking into account the frequency of both items in the dataset. It is calculated as the confidence of the association divided by the support of the second item. Lift is used to compare the strength of the association between two items to the expected strength of the association if the items were independent.

✓ Lift (A⇒B): the percentage of those transactions (consequent support) containing A that must contain B.

$$\text{Lift (A}\Rightarrow\text{B)} = \frac{\text{Confidence (A}\Rightarrow\text{B)}}{\text{Support (B)}} = \frac{0.5}{0.4} = 1.25$$

- A lift value of 1 indicates that the two items are independent,
- A Lift value greater than 1 indicates a positive association, While
- A lift value less than 1 indicates a negative association

3. R package arules for production performance management - case study

This paper proposes Apriori algorithm with R arules for providing the predictive analytics to improve fault detection and product quality with production efficiency. A case study is conducted to illustrate the feasibility of the proposed method is demonstrated. The semiconductor packaging, assembling and test manufacturing company that provides fully integrated semiconductor backend service is explored. Over the last few years, the market for substrates is growing and the cost of substrates as a percentage of the total packaging process is getting more significant, especially for advanced packages such as Ball Grid Array (BGA) packages. Moreover, BGA rework is one of the most challenging parts of Printed Circuit Board *(PCB)* assembly and repair. There are variable BGA defects to be detected and classified including stain, scratch, solder-mask, dark ball, no ball and pinhole…etc, as shows in Figure 2.

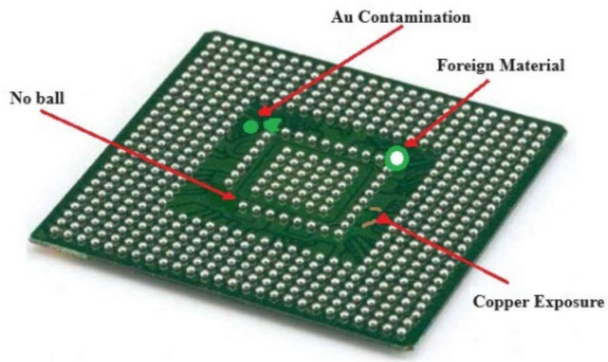

Figure 2. the variable BGA defects to be detected

Due to the BGA's complexity, many potential defects can occur during the soldering process for a BGA and the substrate gold-plated area. Therefore, this paper explores BGA mainly for solder mask and surface finishes (gold plating) are inspected for defects.

- **Solder Mask (S/M)**: The main purpose of making the "solder proof layer" is to distinguish between the welded assembly area and the non-welded area. In addition, it is also possible to prevent the copper surface from being oxidized and to meet the aesthetic requirements. Many potential defects can occur during the soldering process for a BGA.
- **Surface Finishes (Gold Plating)** - Electrical gold Ni/Au/Cu plating: It consists of an electroless nickel plating covered with a thin layer of immersion gold, which protects the nickel from oxidation.

The abnormal conditions of the BGA product are:
- **Solder Mask (S/M)**:
  - ✓ Lack of S/M (Insufficient solder)
  - ✓ S/M *uneven* surface
  - ✓ S/M peeling (*Peel*-off from the surfaces),
  - ✓ S/M scratches,
  - ✓ S/M Foreign Material (S/M FM);

- **Gold plating [upward (U:)* [* the front surface of gold plating]**
  - ✓ U:Au Contamination,
  - ✓ U:Au Dent,
  - ✓ U:Au Bump,
  - ✓ U:Via not filled,
  - ✓ Gold Foreign Material (U:GFM),
  - ✓ Copper Exposure(U: CU), ;

- **Gold plating [downward (D:)) [# the back surface of gold plating]**
  - ✓ Gold Foreign Material (D:GFM),
  - ✓ Copper Exposure (D: CE);

- **Others**: Substrate Broken (SB) and OSP Contamination.

A set of R-arules analysis the abnormal factors and association rules that may cause product's failure during the processes for applying solder mask and gold plating. As shown in Figure 3, the abnormal conditions of BGA metadata, **“1” indicates abnormality**; **“0” indicates normality**. Before importing data into R, all abnormal metadata data types field (col-type) must first change from double to text type (Character),

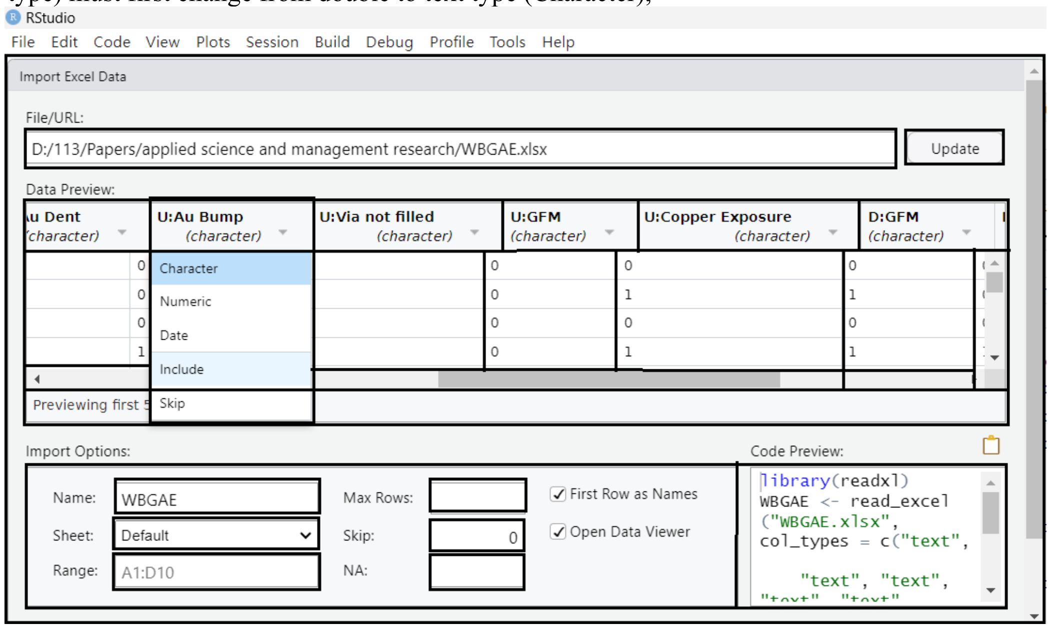

Figure 3. The abnormal conditions of BGA metadata

In this study, the association rules that satisfy minimum support (s ~≧0.5), minimum confidence (c ~≧0.5) and "lift" value is ≧1. As shows in Figure 4, we can find that the **single item set rules** for abnormal condition of BGA product are:

[1] { }⇒{**U: Copper Exposure =1**}:

U:Copper exposure **on the front surface** of gold plating**= 1** with 0. 5078864 support and 0.5078864 confidence. That is, the percentage of the "**abnormality**" for copper exposure on the front surface of gold plating is 51%. On the contrary, the normality is 49%.

[3] { }⇒{**D:GFM =1**}:

D:Gold foreign material **on the back surface** of gold plating=1 with 0.6056782 support and 0.6056782 confidence. That is, the percentage of the "**abnormality**" for foreign material on the back surface of gold plating is 61%. On the contrary, the normality is 39%.

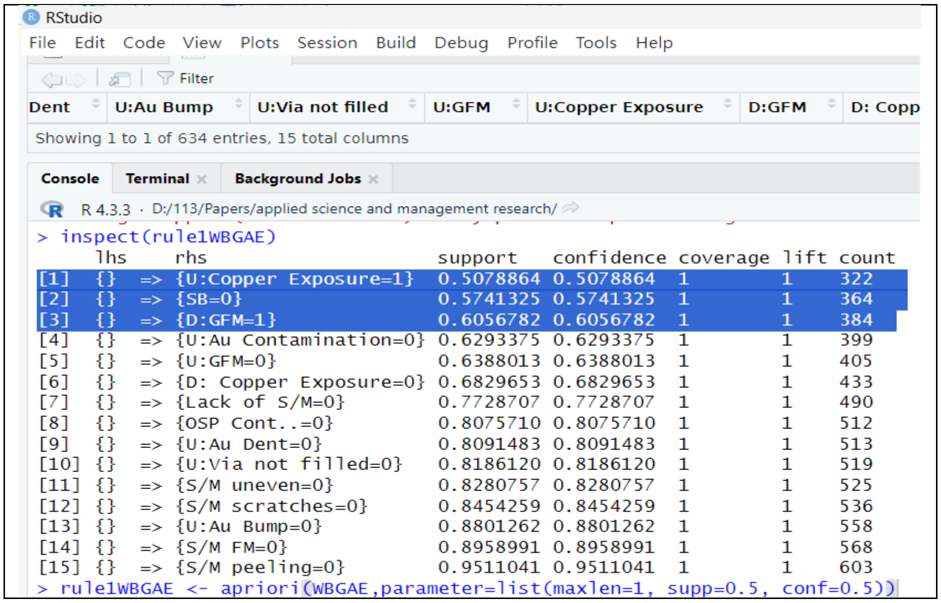

Figure 4. The single item set rules for normal and abnormal condition of BGA product

As shows in Figure 5, we can find that the **two item set rules** for abnormal condition of BGA product are:

[3]{D:GFM=**1**}⇒{U: Copper Exposure =**1**}

if {Gold foreign material on the back surface of gold plating=1}, then {Copper exposure on the front surface of gold plating=1} with 0.40 support and 0.66 confidence. That is, the 40% of foreign material on the back surface is "abnormality", which contain also copper exposure on the front surface of gold plating is likely "abnormality".

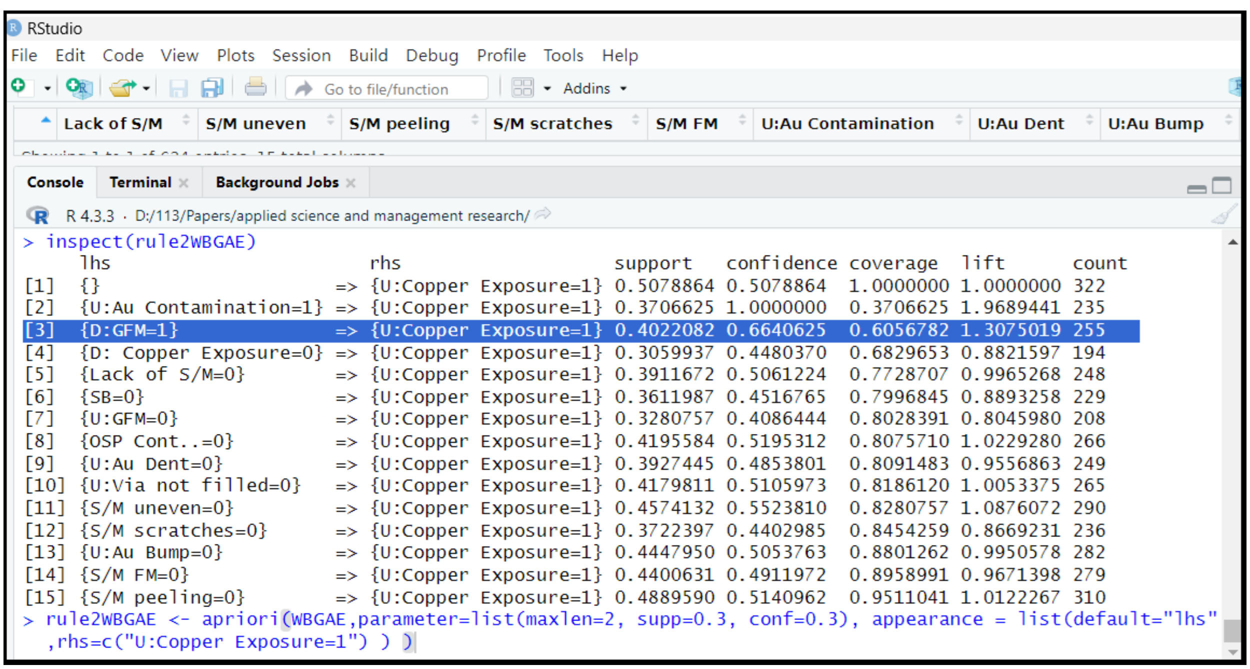

Figure 5. The two item set rules for normal and abnormal condition of BGA product

There are no abnormality of three and four item set. Therefore, when there is a problem with foreign material on the *back* surface, we can give priority to whether there is a problem of copper exposure on the *front* surface of gold plating. In addition, during BGA production, we can strengthen the inspection in the gold surface on the back process, and explore whether there are problems in personnel, methods, materials, and machinery to cause abnormal conditions in this process.

4. Conclusion

In this paper, we illustrated the association rules find out the main conditions of abnormalities in various production lines, and whether there are related rules for product failures and abnormal conditions. The abnormal association rules of the products and provide the company with a systematic, scientific and quantitative reference information through the analysis of the company's accumulated data. The company manager can guess which part of the problem is wrong according to the specific products, abnormal conditions and the relationship. At the same time, it proposes improvement measures to prevent the failure from happening again, thereby improving liability for the products by offering customers some form of guarantee in form of either guaranteed function, product availability or results.

The experimental results show that the proposed algorithm is successful in detecting and classifying the defects on gold-plating regions. The recognition speed becomes faster and the system becomes more flexible in comparison to the previous system. The proposed method, using unsophisticated and economical equipment, is also verified in providing highly accurate results with a low error rate.

References

Agrawal, R., Imieliński, T. and Swami A., Mining Association Rules Between Sets of Items in Large Databases, in *International Conference on Management of Data, SIGMOD.*ACM Press, 1993.

Agrawal, R. and Srikant R., Fast Algorithms for Mining Association Rules in Large Databases, in *Proceedings of the 20th International Conference on Very Large Data Bases*, Morgan Kaufmann Publishers Inc. pp.487-499, 1994,.

Bing, H. and. Li, Y.B.,Research and Application of Association Rules Methods in Data Mining for Commercial Sales Analysis. in *International Conference on Networking and Digital Society*, 2009.

Kim, J. W., Han, S.Y. and Kim, D.S., Association rules application to identify customer purchase intention in a real-time marketing communication tool. in *Fourth International Conference on Ubiquitous and Future Networks (ICUFN)*. 2012.

Lee, J. *Industrial AI*, Springer, Singapore, ISBN-10: 9811521433, ISBN-13: 9789811521430, 10.1109/ CCIS.2012.6664374, 2020.

Li, C., Chen, Y., and Shang, Y., A review of industrial big data for decision making in intelligent manufacturing*, an International Journal of Engineering Science and Technology* (JESTECH), Volume 29, 2022,

Liu,B., Hsu,W., Chen, S. and Ma, Y., Analyzing the subjective interestingness of association rules, in *IEEE Intelligent Systems and their Applications*, vol. 15, no. 5, pp. 47-55, doi: 10.1109/5254.889106, Sept.-Oct. 2000.

Omari, A., Langer, R. and Conrad, S., TARtool: A Temporal Dataset Generator for Market Basket Analysis. in *4th International Conference on Advanced Data Mining and Applications (ADMA)*, Chengdu, China, 2008..

Peres, R.S, Jia X., Lee J., Sun K., Colombo A.W. and Barata, J., "Industrial Artificial Intelligence in Industry 4.0 Systematic Review, Challenges and Outlook," in *IEEE Access*, vol. 8, pp. 220121-220139, 2020

Peng, Y. and Zhou, T., Research on the Apriori algorithm in extracting the key factor, *2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems*, Hangzhou, China, pp. 90-93, 2012

Wang, Y. and Song Y., Classification Model Based on Association Rules in Customs Risk Management Application. in *International Conference on Intelligent System Design and Engineering Application*, 2010.

Wang, F., Application of multidimensional association rule techniques in manufacturing resource planning system, *on Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, Yantai, China, pp. 1433-1437, *2010*

Vazan, P., Janikova, D., Tanuska, P., Kebisek, M. and Cervenanska, Z. Using data mining methods for manufacturing process control in *20th IFAC World Congress*, pp.6178-6183, 2017.

Biographies

Hsiao-Kang Lin is an Assistant Professor in Department of Industrial Management at I-Shou University, Kaohsiung, Taiwan R.O.C. She obtained her Ph.D. degrees from the Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, UK. Her expertise includes tools to support Knowledge Sharing within collaborative teams, semantic web application for global enterprise integration, Knowledge Discovery, Big Data and Artificial Intelligence applications, …etc. She has a wide range of international academic publications, such as Procedia CIRP, Scientific Research and Essays, International Journal of Production Research, Computers in Industry, Proceedings of the Institution of Mechanical Engineers Part B: Journal of Engineering Manufacture, etc.

Tzu-Jou Liao is currently an Associate Professor at I-Shou University in Kaohsiung, Taiwan. She holds a Ph.D. in Management from Rutgers, The State University of New Jersey. Her expertise mainly lies in organizational behavior and organizational theory. Her research has been published in both Taiwanese and international journals.