

# 利用資料探勘技術制定封測廠的機台保養維修計畫

## A study for developing the machine maintenance plan of IC packaging and testing plant by used of data mining technology

林雅祺 Ya-Chi Lin<sup>1</sup>

王政鴻 Cheng-hung Wang<sup>2</sup>

蔡勝閔 Sheng-Hong Tsai<sup>3</sup>

蕭淑君 Shu-Jun Xiao<sup>4</sup>

彭定國 Ting-Kuo Peng<sup>5</sup>

### 摘要

在眾多的半導體產品測試作業下，除了產出客戶所需要的產品外，也產出了許多有關封裝測試的品質資料。由於半導體產業所使用的機台成本高，不可能輕易置換，使得廠商非常看重機台的保養與維修。傳統上，廠商要制定機台的保養與維修計畫時，受限於人員素質以及時間急迫等因素，大都只看機台的測試數據以及機台稼動率等數值，以量化的方式來找出有問題的機台。然而，半導體測試作業所牽涉到的機台多，機台之間也存在某些不易察覺的關聯性，如果只以測試數據來找出有問題的機台，只能改善到個別獨立的機台，而對彼此影響的機台卻成了漏網之魚，在提升測試效率上也會出現瓶頸。本研究以半導體封裝測試廠商的測試資料為主體，進行機台間潛在關聯法則的探勘。藉由資料探勘技術所得到的關聯法則，考慮到存在於機台間的相依性，彌補原先獨立有問題的機台保養維修的不足，使廠商能制定更完善的保養維修計畫，以提昇機台測試品質。

**關鍵字：**資料探勘、關聯法則、機台篩選

### Abstract

There are many testing information resulted from semiconductor product testing operations. Due to the high cost of the machine used in the semiconductor industry, many companies focus on developing a very high-value machine maintenance and repair plan instead of replacement. Traditionally, subject to the quality of personnel and time urgency factors, most decision makers

---

<sup>1</sup> 明新科技大學工業工程與管理系大學生。

<sup>2</sup> 明新科技大學工業工程與管理系大學生。

<sup>3</sup> 明新科技大學工業工程與管理系大學生。

<sup>4</sup> 明新科技大學工業工程與管理系大學生。

<sup>5</sup> 明新科技大學工業工程與管理系助理教授(聯絡地址：304 新竹縣新豐鄉新興路 1 號，聯絡電話：03-5593142 轉 3236，E-mail: pengtk@must.edu.tw)。

build the machine maintenance and repair plan based on the test data and the machine utilization rate. However, the semiconductor test job involves a lot of equipments. There exists some imperceptible relevance between the machines. Only the individual machine can be improved independently if only considering the test data of individual machine. The mutual influence of the machine has been ignored to lower the efficiency of the machine maintenance and repair plan. This paper proposes an association rule mining approach for semiconductor packaging and testing manufacturers to find the main potentially relevant between machines. By used of the association rule data mining techniques, dependencies existing between machines will be taken into account so that manufacturers can develop a more comprehensive maintenance plan to enhance the test quality.

**Keywords:** Data Mining, Association Rule, Machines Selection.

## 壹、緒論

在半導體封裝測試產業中，所面臨的挑戰便是測試品質必須滿足客戶的需求。而在眾多的產品測試條件與標準作業下，除了產出了客戶所需要的產品與結果外，也產出了許多的測試資料與客戶對測試結果的抱怨回饋。這些不斷產出的大量資料，在資訊經濟時代成為新商機的材料，也是新商業模式的基礎。在不同的思維下，原本看似平凡無奇的測試資料，透過不同的方式進行分析與探勘，便可帶來公司的創新，進而創造出新經濟價值，使公司保有市場的佔有率，展現競爭力。由於半導體封裝測試的作業速度快，機台的購置成本高而不可能輕易置換，再加以人員的高流動性，新進員工因不熟練其中環節而出錯，導致產品不良率高，造成半導體業者的無謂損失，由於機台的正確性會影響測試的結果，也使得封裝測試廠商非常看重機台的保養與維修。傳統上，封裝測試廠商要制定機台的保養與維修計劃時，多仰賴工程(品質)人員分析判讀測試資料，以找出造成品質低落的問題所在。然而受限於人員素質以及時間急迫等因素，工程人員大都只看機台的測試數據以及機台稼動率等數值，以量化的方式來找出有問題的機台，而後以此判定改善的措施。然而，半導體測試作業所牽涉到的機台多，機台之間也存在某些不易察覺的關聯性，如果只以測試數據來找出有問題的機台，只能改善到個別獨立的機台，而對彼此影響的機台卻成了漏網之魚，在提升測試效率上也會出現瓶頸。

近年來，許多廠商逐漸懂得資料的重要性，紛紛引進資料分析技術或是委託資料分析專家，把工廠生產的巨量資料轉換成有用的資訊，以便能作為改善作業程序的參考依據。然而，資料分析專家往往缺乏對半導體特定領域的了解，以致於分析的結果無法反映出廠商的真正需求；而廠商本身藉由訓練內部工程人員，結合內部資料處理人員，建立工程資料分析體系試圖解決工廠所遇到的問題，企圖不假外人之手，以確保工廠內的機密(know-how)沒有外流之虞，卻不知如何進行資料探勘，以致於成效亦有限。本研究便以半

導體封裝測試廠商的測試資料為主體，進行機台間潛在關聯法則的探勘。希望藉由資料探勘技術所得到的關聯法則，考慮到存在於機台間的相依性，彌補原先獨立有問題的機台保養維修的不足，將所篩選的機台適當地納入到保養維修計畫中，以期能更完整的改善機台測試品質。

## 貳、文獻探討

### 一、封測產業的關鍵成功因素

由於封測產業已走向微利時代，在單價不高的情形下，要保持其原本的競爭優勢，必須將公司稀有資源投入於關鍵成功因素上，每家公司會因不同的定位，核心能力，策略目標，其關鍵成功因素的取捨也有所不同。陳梧桐(民 89)在其研究中指出，台灣封測廠的競爭力來自於以下幾個因素：產品型態的組合、客戶關係管理、一元化服務、技術上的研發、人力資源管理、規模經濟以及範疇經濟等。其中，由於封測產業是無自有產品的製造服務業，再加上設備投資昂貴，若能達到規模經濟，便能降低原料成本，學習曲線提高生產力等優勢，使產品價格有競爭力，成為公司的主要獲利來源。林顯堂(民 90)在其研究中指出封測產業之核心能力包括：技術提昇與改善、技術累積能力、跨部門整合能力、成本掌控能力、問題解決能力、知識管理能力，並指出公司可以透過全面品質管理(TQM)的系統化方法，產生公司有價值的能力，並藉由顧客價值、競爭差異化與延展性加以篩選，提昇公司的核心能力。許日明(民 94)針對封測產業的移動障礙與經營績效差異進行探討，發現到積極領先的企業遇到的主要移動障礙為：研發與製程能力、投資金額、成本優勢、垂直整合能力、對關鍵材料的掌控能力等。吳彩如(民 97)則在其論文中提到，封測產業競爭力的重要因素在封測的產出能力，若想要提昇其產出能力，可以從現行的品質問題著手，藉由詢問員工意見，集合工程師、管理團隊的共同經驗，運用群體合作的學習方法，找出真正的解決方法。

從上面的文獻可以歸納出，封測產業的關鍵成功因素可以分成幾個方面：

- ◎ 品質(Quality)掌控—這包括了客戶關係的建立、產品良率的掌控。
- ◎ 技術(Technique)掌控—關鍵技術掌握能力、製程技術累積能力。
- ◎ 成本(Cost)掌控—成本掌控能力、規模經濟下的價格競爭力。
- ◎ 人力(Human)素質—問題解決能力、知識管理能力。

其中，封測產業的工業化基礎已從自然資源轉移到知識資產，而品質的控管妥當以及技術的掌控，皆有賴於人力素質的提昇以及善用組織內的知識，以維持本身的競爭力。由於資訊科技的發達，使得知識的編碼、儲存與分享，均較以往更為容易且所需成本亦為較低。因此若能有效地進行資料分析，發掘組織中最具價值的資產—知識(Knowledge)，便能有效提昇廠商本身的技術能力。

## 二、資料探勘

近年來，資訊技術的提昇以及儲存技術的進步，資料倉儲內存放了各式各樣的大量資料，資料分析的工作已非人力所能完成，為了能快速且有效地因應不同的分析需求，資料探勘的技術乃因應而生。資料探勘可視為知識發掘(Knowledge Discovery in Databases, KDD)其中一個重要的分析步驟。知識發掘是從資料中辨別潛在有用的法則或關係，並形成最終可理解的模式的過程；而資料探勘則是知識發掘在可接受的計算效率限制內透過特定的演算法形成特定模式的一個步驟。資料探勘(Data mining)有多種不同的翻譯，諸如數據挖掘、資料挖掘、資料採礦。儘管通常資料探勘應用於資料分析，但是像人工智慧一樣，資料探勘可用於不同的領域，因此此詞彙具有豐富含義，學者紛紛提出不同的資料探勘定義：Frawley(1992)依據過去文獻，將資料探勘定義為「從資料庫中提取出隱含的、前所未知以及潛在有價值資訊的過程」。Grupe 與 Owrang (1995)則認為資料探勘是指「從已經存在的資料庫當中挖掘出專家仍未知的新事實」。而 Berry 與 Linoff (1997, 2000) 認為「資料採礦是為要發現出有意義的樣型或規則，而必須從大量資料之中以自動或是半自動的方式來探索和分析資料」。Han 與 Kamber (2012)將資料探勘定義為「大型資料庫中萃取出決策人員感興趣資訊或型態(interesting information or patterns)的過程」。Hand 等人(2001)則將資料探勘定義為「一門從大量資料或者資料庫中提取有用資訊的科學」。

一般而言，資料探勘過程可分為四個階段：資料彙總整理、資料轉換篩選、資料型態擷取、以及資料驗證呈現。首先，在彙總整理階段刪除了錯誤、重複以及不完整的資料。而後在轉換篩選階段選取探勘所需要的欄位或資料，並對資料進行格式的轉換或衍生新的變數。在型態擷取階段選擇適當的探勘方法進行型態擷取，倘若得到的型態不理想，則再次進行資料的彙整與節選；倘若已產生理想的型態，則進入資料驗證呈現階段，以領域知識為基礎驗證探勘出來的型態，並將探勘出來的型態以容易理解的方式(如以視覺化工具轉換成扼要易懂的圖表)呈現出來，以利提供決策支授之用。由於資料存在多種特性，資料探勘的方法也相當多樣化。主要有關聯規則分析 (association rules analysis)、分類與預測(predictive and classification)、分群分析(cluster analysis)與趨勢分析(trend and evolution analysis) 等模式(Chen et. al, 1996; Han 與 Kamber, 2012)。

關聯式規則在不同領域有相當多的應用，較常見的例子是在顧客的購物分析上。一個比較常被提出的例子為「**尿布與啤酒**」，就是消費記錄中發現許多顧客買了尿布就會買啤酒。經過研究分析後發現：「在美國，許多年輕母親都留在家中照顧嬰兒，而父親匯到超市買尿布，但往往就會順便買啤酒」。從「尿布與啤酒」的例子可以發現，在實際的生活上，存在著許多表面上看似毫無關連的事物，但其實有著密不可分的關係，而 Apriori 便是在資料探勘的領域中，用來找尋關聯式規則的經典演算法之一。

## 參、研究方法

本研究主要目的是藉由現場機台的資料來找出檢測機台之間的潛在關聯性。一般而

言，現場都會將機台測試資料保存下來，但是面對一大堆的測試結果數據，讓人眼花瞭亂，不知從何處開始分析。在整個資料探勘過程中，本研究採取了兩個階段進行：資料的預處理與關聯法則探勘。在資料的預處理階段，將每一批量產品對不同檢測機台的錯判情形，如果此一批量在機台被檢測結果誤判，便標示為「V」，如表 1 所示。

表 1. 產品檢測資料

|              |   |   |   |       |   |
|--------------|---|---|---|-------|---|
| 機台編號<br>批量編號 | 1 | 2 | 3 | ..... | m |
| 1            | V | V |   | ...   | V |
| 2            |   |   | V | ...   |   |
| 3            |   | V |   | ...   | V |
| :            | : | : | : |       | : |
| :            | : | : | : | ...   | : |
| :            | : | : | : |       | : |
| N            |   |   | V | ...   |   |

將表 1 的資料表示轉換成全部產品檢測資料集  $D$ ，令全部的機台所成的集合為  $M$ ， $M = \{S_i, i = 1, 2, \dots, m\}$ ，如表 2 所示。以  $T_i$  來標示每批產品的檢測記錄識別標籤， $D = \{T_i, i = 1, 2, \dots, n\}$ 。每筆記錄中都有每批產品檢測誤判機台所成的集合  $F_i \subseteq M, i = 1, 2, \dots, n$ 。

表 2. 全部產品檢測資料集

| 產品批量  | 誤判機台集合 |
|-------|--------|
| $T_1$ | $F_1$  |
| $T_2$ | $F_2$  |
| :     | :      |
| :     | :      |
| :     | :      |
| $T_n$ | $F_n$  |

在關聯法則探勘階段，本研究採用 Apriori 演算法來進行誤判機台篩選，流程如圖 1 所示。

步驟 1. 設定  $k=1$  來標示疊代的階段。並給定誤判可容忍的次數，作為演算所需的最小支持度  $Min\_sup$ 。

- 步驟 2. 從全部產品檢測資料集  $D$  中找出含有  $k$  個誤判機台的組合  $C(k) \subseteq M$ 。
- 步驟 3. 掃描資料集  $D$ ，分別計算各個誤判機台組合  $C(k)$  的支持度(support)，即在全部的資料集中，包含誤判機台組合的檢測記錄數量： $Support(c(k)) = O(c(k))$ 。
- 步驟 4. 剔除  $C(k)$  中未達到  $Min\_sup$  的機台組合，便可得到含有  $k$  個機台組合之高風險集合  $L(k) \subseteq S$ 。
- 步驟 5. 合併  $L(k)$  來產生  $C(k+1)$ ，倘若  $L(k)=C(k+1)$ ，即沒有辦法產生其他組合，則誤判機台篩選演算結束。
- 步驟 6.  $k=k+1$ ，跳到步驟 2。

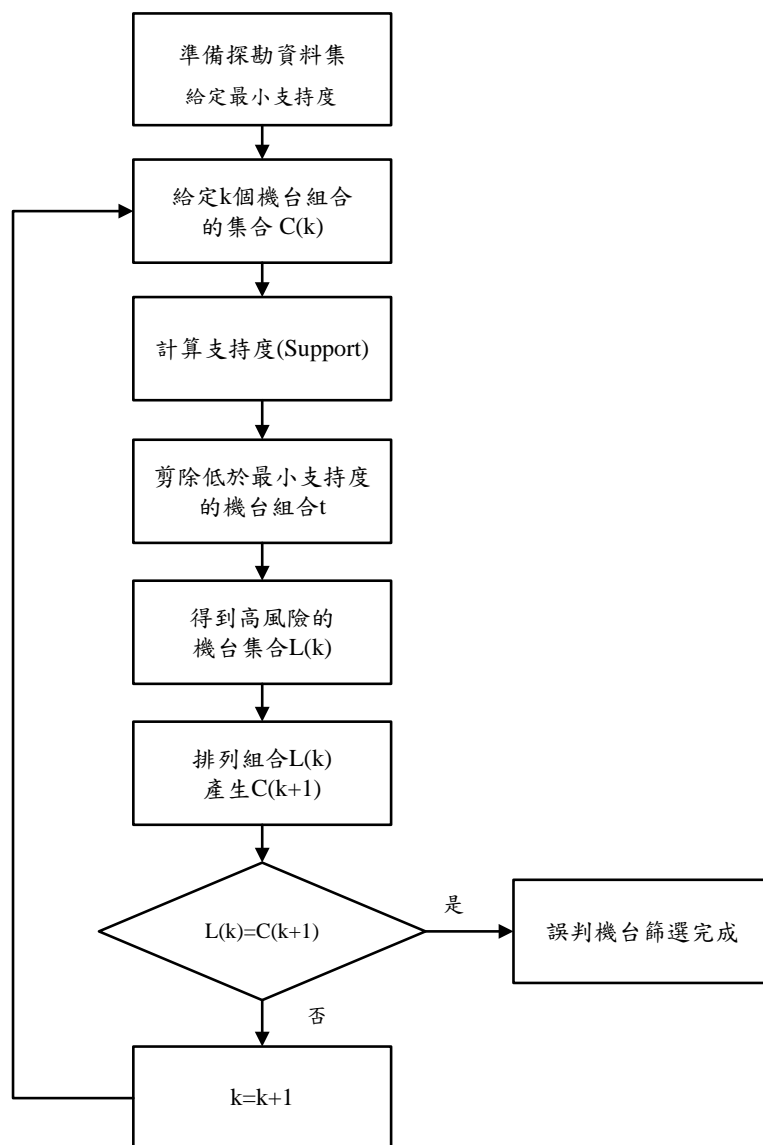


圖 1. 誤判機台篩選流程

## 肆、範例演練

本研究以某半導體封裝測試廠商的 20 筆檢測資料(如表三所示)來做為範例。第 1 欄是標示檢測記錄編號，第 2 欄到第 9 欄分別代表檢測所需使用到的 8 部機台，打“V”者表示機台在該檢測記錄中是檢測誤判。

表 3. 產品檢測資料

| 機台識別<br>檢測記錄 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------|---|---|---|---|---|---|---|---|
| 1            | V | V |   | V | V |   | V | V |
| 2            |   |   |   |   |   | V |   |   |
| 3            |   | V |   | V | V | V |   | V |
| 4            | V | V | V | V |   |   |   | V |
| 5            |   |   |   | V |   |   |   |   |
| 6            | V | V | V | V | V |   |   | V |
| 7            | V | V | V | V |   | V |   | V |
| 8            | V |   |   |   |   | V | V |   |
| 9            |   | V |   |   |   |   |   |   |
| 10           | V |   | V |   |   |   |   |   |
| 11           | V |   |   | V |   | V | V |   |
| 12           |   | V |   |   |   |   |   |   |
| 13           | V |   |   |   |   | V |   |   |
| 14           | V |   | V |   | V |   |   |   |
| 15           | V | V |   | V |   | V | V | V |
| 16           |   |   |   | V |   | V | V |   |
| 17           |   | V |   | V |   | V |   | V |
| 18           | V |   | V |   |   |   |   |   |
| 19           | V |   | V | V |   |   |   |   |
| 20           |   | V | V | V |   | V | V | V |

表 3 的產品檢測資料若以傳統的誤判數(如表 4 所示)多寡來決定改善維修的重點，礙於人力與時間的限制，工廠管理者會選擇誤判率前四名的機台(1, 2, 4, 6)做為改善的重點。然而，從誤判數的資料來看，機台 3 與機台 8 的誤判率似乎也有一定的程度，但卻因而排到較後的改善維修。

表 4. 機台誤判數

| 機台編號 | 1  | 2  | 3 | 4  | 5 | 6  | 7 | 8 |
|------|----|----|---|----|---|----|---|---|
| 誤判數  | 12 | 10 | 8 | 12 | 4 | 10 | 6 | 8 |

本研究採用資料探勘的方式找出機台之間潛在的關聯法則，以便找出潛在應改善維修但被忽略的機台。首先，將表 3 的資料轉換成旅遊篩選資料集  $D$ ，如表 5 所示。

表 5. 全部產品檢測記錄

| 產品批量 | 誤判機台集合             |
|------|--------------------|
| T1   | {1, 2, 4, 5, 7, 8} |
| T2   | {6}                |
| T3   | {2, 4, 5, 6, 8}    |
| T4   | {1, 2, 3, 4, 8}    |
| T5   | {4}                |
| T6   | {1, 2, 3, 4, 5, 8} |
| T7   | {1, 2, 3, 4, 6, 8} |
| T8   | {1, 6, 7}          |
| T9   | {2}                |
| T10  | {1, 3}             |
| T11  | {1, 4, 6, 7}       |
| T12  | {2}                |
| T13  | {1, 6}             |
| T14  | {1, 3, 5}          |
| T15  | {1, 2, 4, 6, 7, 8} |
| T16  | {4, 6, 7}          |
| T17  | {2, 4, 6, 8}       |
| T18  | {1, 3}             |
| T19  | {1, 3, 4}          |
| T20  | {2, 3, 4, 6, 7, 8} |

接下來給定最小支持度  $Min\_sup=8$ ，而後依據先前所述的演算法流程來進行誤判機台潛在關聯法則探勘。

在第一次疊代( $k=1$ )中，由單一機台組成項目子集合  $C(1)$ 。並掃描資料集  $D$ ，找出包含  $C(1)$  的檢測記錄，以得到個別機台的誤判數，並將將  $C(I)$  中誤判數未達到最小支持度 ( $Min\_sup=8$ ) 的機台 {5}, {7} 剔除，得到含有單一機台之高風險集合  $L(I)$ 。如圖 2 所示。



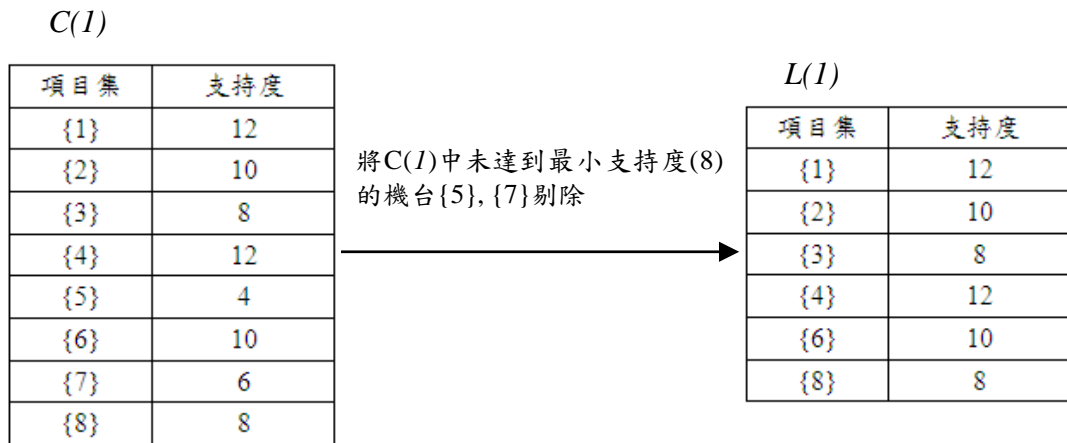


圖 2. 單一機台的篩選結果

從圖 2 的結果可以發現， $L(1)$ 的數量不只一個，可以合併  $L(1)$ 項目集來產生不同的  $C(2)$ 項目集。而後進行第二次疊代( $k=2$ )。

在第二次疊代( $k=2$ )中，由第一次疊代所得高支持度集合  $L(1)$ 產生  $C(2)$ 。而後掃描資料集  $D$ 以得到  $C(2)$ 項目的支持度，並將將  $C(2)$ 中未達到最小支持度( $Min\_sup=8$ )的機台組合剔除，得到兩部機台組合之高支持度集合  $L(2)$ 。如圖 3 所示。

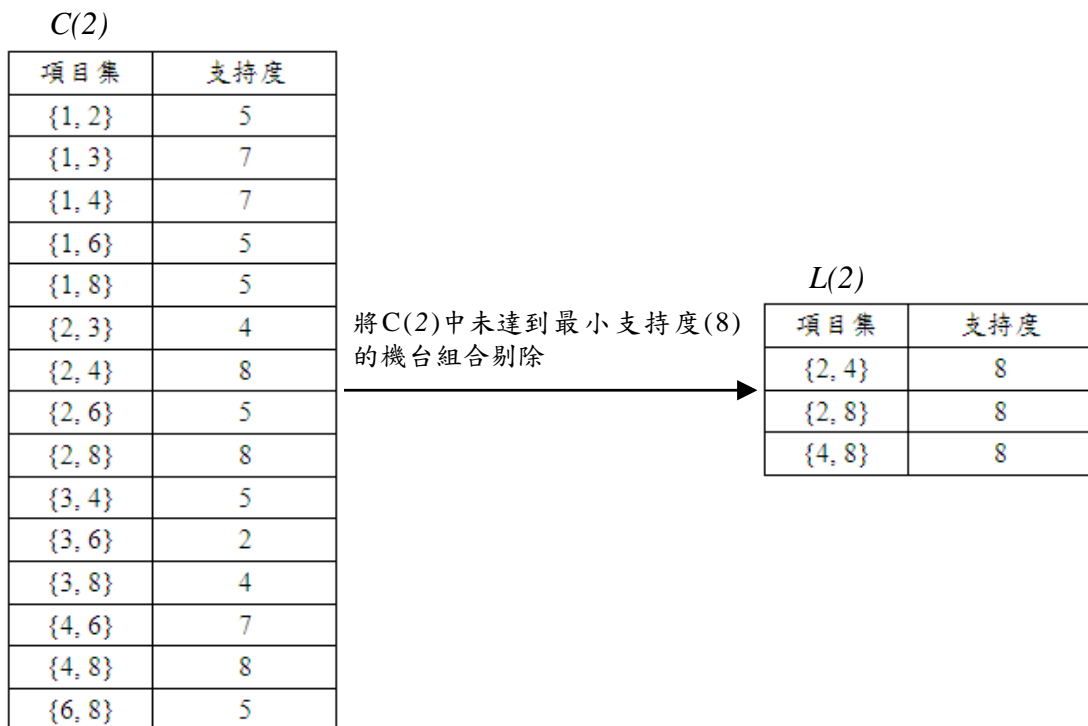


圖 3. 兩部機台的篩選結果

從圖 3 的結果可以發現， $L(2)$ 的數量不只一個，可以合併  $L(2)$ 項目集來產生不同的  $C(3)$ 項目集。而後進行第三次疊代( $k=3$ )。

在第三次疊代( $k=3$ )中，由第二次疊代所得高支持度集合  $L(2)$  產生  $C(3)$ 。而後掃描資料集  $D$  以得到  $C(3)$  項目的支持度，並將將  $C(3)$  中未達到最小支持度( $Min\_sup=8$ )的機台組合剔除，得到含有三部機台之高支持度集合  $L(3)$ 。如圖 4 所示。由於  $L(3)$  只餘一個組合，無法再組合出不同的  $C(4)$  組合，所以沒有再繼續向下探勘的必要，因此可以得到高支持度的機台組合為  $\{2, 4, 8\}$ 。即  $\{2, 4, 8\}$  三部機台中存在著潛在的關聯性。

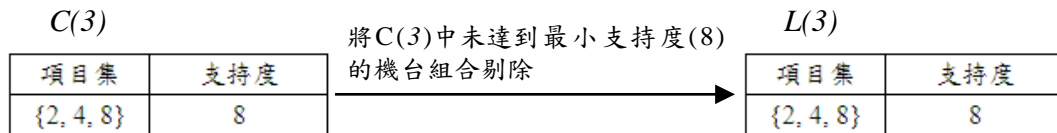


圖 4. 三部機台的篩選結果

因此，根據現前的測試資料，可以確定這一階段需要納入保養維護計畫的機台為 1, 2, 4, 6, 8。如果依據傳統的篩選準則，只進行機台 1, 2, 4, 6 的改善維修，卻忽略了機台 8，然而，機台 2 與 4 的測試結果仍會受到機台 8 的影響，機台 8 的檢測品質與效率仍未提昇，反而降低了改善維修措施的效益。

## 伍、結論

資料探勘是從大量資料中找出潛在有用的知識與關聯法則，本研究透過資訊技術的支援以及現場實地產出的資料，透過適當的資料探勘技術可以找出機台之間的潛在關聯法則，並做出適當的機台維修改善計畫，使半導體封裝測試效率得以提昇，使產品能夠更符合客戶的要求。傳統的維修改善或許只重視個別機台的測試狀況，卻忽略了機台之間的相依性，而使得某些機台與人員逃過了維修與保養的責任。這些機台雖然沒有出現高誤判率，但是卻與其他高誤判率的機台存在著相依關係，換言之，其他機台的高誤判率很可能是由這些潛伏機台所引發。期望藉由本研究的資料探勘規劃，讓工程單位在未來規劃保養維修計畫時，不要只考慮到單一機台的狀況，能夠更進一步考量多個機台之間的關聯，以便能將有問題的機台一網打盡。在此研究中所舉的範例，測試資料尚未顯現巨量資料的魅力，後續研究可以納入更多的測試記錄，讓整個研究更臻完善，即使在大量產生的測試資料下，機台之間的關聯法則也能更有效地擷取出來，以擴大本研究的實用範圍。

## 參考文獻

- 陳梧桐(民 89)，「封裝產業之關鍵成功因素」，*工研院電子所半導體產業情報*，第 7 期，頁 1-5。
- 林顯堂(民 90)，*IC 封裝業核心能力建構模式—以日月光公司為例*，中山大學高階經營碩士學程專班論文，高雄市。
- 許日明(民 94)，*台灣地區封裝產業策略群組與經營績效之關連研究*，東海大學管理碩士在職專班論文，台中市。
- 吳彩如(民 97)，*晶圓凸塊產品組合最優化之遺傳基因演算法研究*，成功大學碩士在職專班

論文，台南市。

- Berry, M., and Linoff, G. (1997), *Data Mining Techniques for Marketing, Sales and Customer Support*, John Wiley & Sons, New York.
- Berry, M., and Linoff, G. (2000), *Mastering Data Mining: The Art & Science of Customer Relationship Management*, John Wiley & Sons, New York.
- Chen, M. S., Han, J. and Yu, P. S. (1996), "Data Mining: An Overview from a Database Perspective, " *IEEE Trans. on Knowledge and Data Engineering*, Vol. 8, No. 6, pp.866-883.
- Frawley, W. J., Shapiro, G. P., and Mathews, C. J. (1992), "Knowledge discovery in databases: An overview," *AI Magazine*, Vol. 13, No. 3.
- Grupe G. H., and Owrannng, M. M. (1995), "Database Mining discovering new knowledge and cooperative advantage," *Information Systems Management*, Vol. 12, No. 4, pp.26-31.
- Han, J., and Kamber, M. (2012), *Data Mining Concepts and Techniques (3<sup>rd</sup> ed.)*, Morgan Kaufmann, San Francisco, CA.
- Hand, D., Mannila, H., Smyth, P. (2001), *Principles of Data Mining*, MIT Press, Cambridge, MA.