

# 應用約略集合理論建立電磁干擾診斷系統

## Application of Rough Set Theory to EMI diagnosis

黃承龍<sup>1</sup> Cheng-Lung Huang  
華梵大學資訊管理學系

李得盛<sup>2</sup> Te-Sheng Li  
明新科技大學工業工程與管理系

彭定國<sup>2</sup> Ting-Kuo Peng  
明新科技大學工業工程與管理系

施政男<sup>1</sup> Cheng-Nan Shih  
華梵大學資訊管理學系

<sup>1</sup>The Department of Information Management, Huafan University and <sup>2</sup>Department of Industrial Engineering and Management, Ming Hsin University of Science and Technology

(Revised July 18, 2003; Accepted September 22, 2003)

**摘要：**「電磁干擾」(EMI, Electromagnetic Interference) 可能造成工廠機器故障、錯誤動作而發生不可預料的情況。如果產品有EMI品質不良，必須針對不符合測試標準的產品進行修改，直到產品通過標準測試值之後，才能將產品出貨。EMI診斷由有經驗的工程師找到EMI問題的根源，此工作耗時且困難，建立一套EMI診斷決策系統有助於縮短診斷時間。本研究應用資料探勘領域中的知識發掘理論—約略集合理論 (RST, Rough Set Theory) 找到診斷知識。RST用來處理含糊與不精確資訊，此法不需要對資料集有先驗或是額外的資訊。以RST探勘輸入變數與不良位置之間存在的法則知識，其步驟如下：資料蒐集、資料前處理、屬性值離散化、屬性化簡、最小屬性集過濾、產生法則、過濾法則、做分類、計算分類正確率以及建立診斷系統等步驟。以國內知名主機板廠商的EMI歷史資料做資料探勘，透過上述步驟所建立的RST診斷法則平均有八成的正確率，已達到實務應用的要求。

**關鍵詞：**資料探勘、電磁干擾、約略集合理論、屬性化簡

**Abstract:** Electromagnetic emissions are radiated from every part of motherboards of personal computers, and thus electromagnetic interference (EMI) occurs. EMI has a bad effect on the

surrounding environment because EMI may cause malfunctions or fatal problems of other digital devices. EMI engineers diagnose EMI problems of motherboard from the electromagnetic noise data measured by the Spectrum Analyzer. It is time consuming to find out the sources (PS2, USB, VGA, etc.) of electromagnetic noise. Rough set theory (RST) is a new mathematical approach to data analysis. This paper constructs an EMI diagnostic system based on RST. There are the following steps: Data Collection, Data Preprocessing, Discretization, Attribute Reduction, Reduction Filtering, Rule Generation, Rule Filtering, Classification, and Accuracy Calculation. Historical EMI noise data, collected from a famous motherboard company in Taiwan, are used to generate diagnostic rules. The result of our research (average diagnostic accuracy of 80%) shows that RST model is a promising approach to EMI diagnostic support system.

**Keywords:** Data Mining, Rough Set Theory, Attributes Reduction, Electromagnetic Interference

## 1. 簡介

台灣的主機板產業大部分以代工為主，代工的客戶大多為世界知名的大廠，這些大廠對於產品的品質的要求特別重視，而在多家主機板廠商相互的競爭之下，所賺取的代工利潤有限，如能有效控制產品品質的不良率及迅速診斷品質不良的原因，必能為公司減少不必要的損失，增加公司的獲利空間。

EMI (Electromagnetic Interference) (Morgan, 1994; Mills, 1993)，中文譯為「電磁干擾」，電磁相互干擾可能造成工廠機器故障、錯誤動作而發生不可預料的情況，甚至對人體健康造成不可預期之影響。所以，在產品設計時就必須以各種方式以降低電磁干擾 (Archambeault, 2002)，而在正式量產前與量產後，如果產品有電磁干擾方面的品質不良，必須針對不符合測試標準的產品進行修改，直到產品通過標準測試值之後，才能將產品出貨。

近年來，歐、美、大陸、日本、韓國、新加坡以及台灣(經濟部標準檢驗局)等各個國家政府單位，對EMI的認證及測試極為重視。全球的各資訊大廠 (例如：IBM、NEC、HP、Dell、Fujitsu) 也對此項測試，列為產品測試規格的重點項目之一。所以，各家主機板公司對EMI品質的檢測會投入很多的成本。

解決EMI的困難程度相當高，非一般的EMI工程師在短時間內就能解決，EMI工程師通常在檢測發現產品EMI不良時，必須要再進一步花更多時間確認不良的來源 (例如，可能是網路卡、USB等)，進行反覆測試和修改，此工作非常耗時與花費高額成本。因此，品質部門迫切希望能找到診斷不良EMI根源零組件的診斷規則，以減少重複性的測試工作，藉此縮短研發測試階段的時間，及降低成本，提昇公司的競爭力。

依據現場工程師的經驗，組件搭配組合的關係及測試頻率等因素，與某特定位置組件的EMI不良有關連。對於這些複雜的關連知識，品質研發管理者無法用耗時的實驗設計方法得到快速有效之結果，一個好的方法便是：針對已開發完成的產品，蒐集相關的歷史資料進行資料探勘 (Data Mining)，以得到診斷規則，供產品設計以及EMI診斷工程師或品質決策者的參考。

結合資料探勘之EMI問題診斷決策流程如圖1所示，以EMI診斷的歷史資料進行資料探勘，找出診斷規則，爾後發生新的案例，便依據診斷決策支援系統的建議，優先進行該位置的診斷，一直到通過EMI測試為止，通過測試之後才可以申請電磁相容性 (EMC, Electromagnetic Compatibility) 認證。

本研究提出以資料探勘中知識發掘的方法—約略集合理論 (RST, Rough Set Theory) (Pawlak 1982, 1994, 1995; Kusiak, 2001) 來進行資料探勘，並配合EMI的專業知識來找出有關於造成EMI方面問題的知識，將所得到的解決方案提供給EMI工程師作為決策的參考。

後續章節將在第2節介紹資料探勘與RST理論；在第3節以主機板工廠實務資料進行診斷知識之資料探勘；最後在第4節做結論。

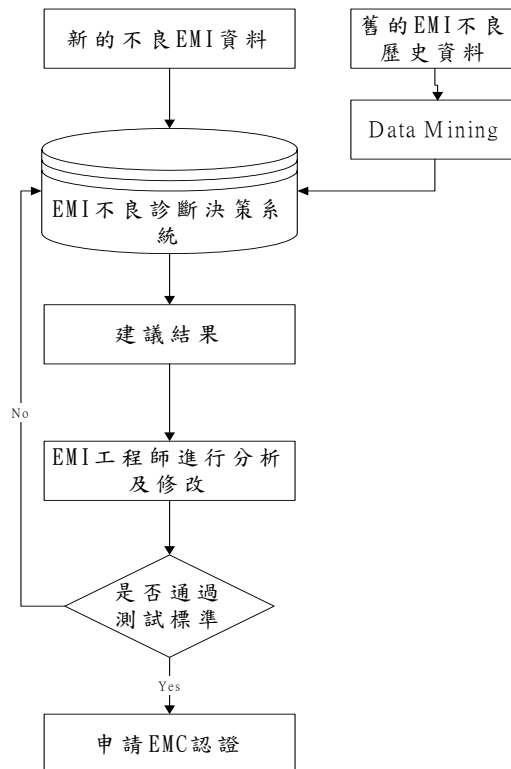


圖1 結合資料探勘之EMI問題診斷決策流程

## 2. 資料探勘與約略集合理論

### 2.1 資料探勘

資料探勘 (Data Mining) 是由大量資料中探勘出不明確的、未知的以及潛在的有用資訊之過程 (Han, 2000)。資料探勘有不同的功能，如做分類 (Classification)、預測 (Prediction)、關聯法則 (Association)、群集化 (Clustering) 等。

許多研究者將資料探勘應用在很多不同的領域，解決不同的問題 (Sushmita, 2002; Kusiak, 2000; Kusiak, 2001; Bayardo, 1999; Hui, 2000; McKee, 2002; Ben-Arieh, 1998; Yoshida, 1999)，也有許多研究者運用資料探勘解決品質方面的問題，例如，Gardner (2000) 使用SOM (Self-Organizing Map Neural Networks) 及法則歸納 (Rule Induction) 兩種演算法對資料作探勘，研究結果指出某個製程步驟和所使用的晶圓材質是造成良率下降的主要因素，而明顯改善電晶體中的集極與射極的漏電流過高的品質不良。簡禎富 (民90) 利用Kruskal-Wallis檢定以及決策樹方法，針對半導體晶圓允收測試及相關製程資料，發展製程事故診斷資料探勘。也還有許多研究針對半導體廠晶圓圖 (Wafer Map) 進行資料探勘 (Chen 2000; Lee, 2001; Skinner, 2002)。

還有很多的資料探勘應用研究，在此無法一一羅列。截至目前為止，應用資料探勘在EMI診斷系統方面，目前尚未有相關的研究。本研究嘗試使用資料探勘中知識發掘的方法—約略集合理論來進行資料探勘，相關理論說明在次小節。

### 2.2 約略集合理論

約略集合或粗糙集合理論 (Rough Set Theory, RST) 是由波蘭籍Z. Pawlak在1982年提出 (Pawlak, 1982)，用來處理含糊與不精確 (Vagueness and imprecise) 資訊的問題，此方法不需要對資料集有先驗或是額外的資訊，例如：機率分佈或是模糊理論的歸屬函數 (Pawlak, 1991)。RST可以在數據集合當中做知識探勘，進而產生決策法則知識。亦即RST可以用來 (1)做屬性化簡、(2)找到資料的隱藏樣式、以及 (3)產生決策法則 (Pawlak, 1982; Kusiak, 2001)。RST在近十多年來已經被使用在許多決策領域的知識探勘 (Dimitras, 1999; Kusiak, 2001; Ahna, 2000; McKee, 2002)。本研究的資料型態適合RST來處理，並且因為RST可以找出決策法則，因此透過RST來找出EMI不良的診斷規則。

#### (1) 資訊系統 (Information System)

資訊系統也稱為決策表 (Decision Table) 或是資訊表 (Information Table)，定義為 $S = \langle U, A, V, f \rangle$  (Pawlak, 1991)， $U$ 是有限物件集合 (Universe)，例如， $U = \{x_1, x_2, \dots, x_{10}\}$ 。 $A$ 是屬性 (Attributes) 的集合，例如： $A = \{a_1, a_2, a_3\} = \{\text{體重, 性別, 血型}\}$ 。 $V$ 是各屬性的值域，例如： $V_{a_1}$ 是屬性 $a_1$ 的值域， $V_{a_1} = \{\text{重, 中, 輕}\}$ 。定義資訊函數 (Information Function)  $f : U \times A \rightarrow V$ ，表示各物件與各個屬性值

之關係，亦即  $f(x, a) \in V_a, \forall a \in A, \forall x \in U$ ，可以用一個表格加以表示。在所有屬性中，有一個是決策屬性 (Decision Attribute)，其餘為條件屬性 (Condition Attributes)。

### (2) 「不可區別關連」(Indiscernibility Relation)

有一個資訊系統  $S = \langle U, A, V, f \rangle$ ，如果  $P$  為一個屬性子集合， $P \subseteq A$ ，而且  $x, y \in U$ ， $\text{Ind}(P)$  稱為「不可區別關係」，定義如下： $x$  與  $y$  是「不可區別」，如果滿足以下條件： $f(x, a) = f(y, a), \forall a \in P$ ，亦即  $\text{Ind}(P)$  是由各屬性值相同的許多群組所構成。依據屬性子集合  $P$ ，這些屬性值相同的群組稱為基本集合 (Elementary Sets)，群組內的物件表達為  $[x_i]_{\text{Ind}(P)}$ 。

### (3) 集合的界線(Approximation of Sets)

「下界」(Lower Approximation)與「上界」(Upper Approximation)是RST的兩個重要觀念。如果  $X$  是  $U$  的某個物件子集合， $X \subset U$ ， $P \subseteq A$ ，則  $X$  在  $P$  中的下界定義為：

$$\underline{PX} = \{x_i \in U \mid [x_i]_{\text{Ind}(P)} \subset X\}$$

亦即下界是由  $x_i$  物件所組成，而  $x_i$  物件是由「不可區別關係」群組中，包含於  $X$  者所組成的， $\underline{PX}$  稱之為  $X$  在  $P$  的下界 (P-Lower Approximation)。

而  $X$  在  $P$  中的上界定義為：

$$\overline{PX} = \{x_i \in U \mid [x_i]_{\text{Ind}(P)} \cap X \neq \emptyset\}$$

亦即上界是由那些  $x_i$  物件所組成，而  $x_i$  物件是由「不可區別關係」群組與  $X$  之交集不為空集合者所組成， $\overline{PX}$  稱為  $X$  在  $P$  的上界 (P-Upper Approximation)。而  $X$  在  $U$  的邊界 (Boundary) 定義為： $\text{PNX} = \overline{PX} - \underline{PX}$ 。

如果  $\text{BX} = \underline{\text{BX}}$ ，則集合  $X$  是可被定義的，亦即  $X$  可以從  $U$  中被完全區別出來，否則有4種型態「無法被定義集合」(Undefinable Sets):

- 1) 如果  $\underline{PX} \neq \emptyset$ ，而且  $\overline{PX} \neq U$ ，則稱  $X$  在  $U$  中可以被粗略定義(Roughly definable in  $U$ )。
- 2) 如果  $\underline{PX} \neq \emptyset$ ，而且  $\overline{PX} = U$ ，則稱  $X$  在  $U$  中在外部無法被定義(Externally undefinable in  $U$ )。
- 3) 如果  $\underline{PX} = \emptyset$ ，而且  $\overline{PX} \neq U$ ，則稱  $X$  在  $U$  中在內部無法被定義(Internally undefinable in  $U$ )。
- 4) 如果  $\underline{PX} = \emptyset$ ，而且  $\overline{PX} = U$ ，則稱  $X$  在  $U$  中完全無法被定義(Totally undefinable in  $U$ )。

接著要計算錯誤分類的比例有多高？作法是以下界的物件數除上界的物件個數：

$$\mu_p(X) = \text{card}(\underline{PX}) / \text{card}(\overline{PX}), 0 \leq \mu_p(X) \leq 1$$

### (4) 屬性化簡 (Reduction of Attributes)

如果  $\text{Ind}(A) = \text{Ind}(A - a_i)$ ，則屬性  $a_i$  就是多餘的，否則  $a_i$  就是必要的 (Dispensable in  $A$ )。也就是說如果抽掉屬性  $a_i$ ，不可區別群組的個數仍舊一樣，表示  $a_i$  屬性是多餘的可以拿掉。

如果  $S$  表示一個資訊系統(或資訊表)， $P$  是  $A$  的屬性子集合，令  $\Psi = \{X_1, X_2, \dots, X_n\}$ ,  $X_i \subset U$ ,  $X_i \cap X_j = \phi$ ,  $\cup X_i = U$ ， $\Psi$  表示  $U$  的分類 (Classification)， $X_i$  稱為類別 (Classes)，這個類別以另外一個決策屬性 (目標屬性)  $d$  代表之。 $\Psi$  的上界與下界定義為： $\underline{P\Psi} = \{\underline{PX_1}, \underline{PX_2}, \dots, \underline{PX_n}\}$ ， $P\Psi = \{PX_1, PX_2, \dots, PX_n\}$ 。 $\Psi$  分類的品質 (Quality of Classification) 定義為：

$$\beta_p(\Psi) = \sum_{i=1}^n \text{card}(\underline{PX_i}) / \text{card}(U)$$

若有一個屬性集合  $P$ ， $P \subseteq A$ ，而屬性集合  $P$  與  $A$  兩者具有相同的  $\beta_p(\Psi)$ ，則屬性集合  $P$  即為縮減的屬性。何謂最小屬性集合？如果無法再找到屬性集合  $R$ ， $R \subseteq P \subseteq A$ ，使得  $\beta_r(\Psi) = \beta_p(\Psi)$ ，此屬性集合  $P$  就是最小屬性集合，以  $RED_\Psi(P)$  表示之。

資訊系統的最小屬性集合 (Minimal Subsets of Attributes) 有多少個？一個資訊系統可能有多個最小屬性集合。多個最小屬性集合的交集所得到的屬性集合，稱為核 (Core)，亦即  $CORE_\Psi(P) = \cap RED_\Psi(P)$ ，核是資訊系統最顯著的屬性集合。

而求算最小屬性集合的過程與方法也是 RST 的重要理論之一。常見的方法是採用「可區別矩陣」(Discernibility Matrix) (Walczak, 1999)，並進行布林運算得到屬性化簡的結果，在 2.3 節中有實例說明。

### (5) 決策法則 (Decision Rules)

當屬性化簡完成之後，即可得到一個決策表，分別由條件屬性與決策屬性所構成，該決策表可能產生很多條規則，這個決策表亦即代表著決策分類法則，說明於 2.3 節中。

## 2.3 RST 實例說明

以下舉一個簡單實例說明 RST 做屬性化簡，以及產生決策法則的過程。有 8 位應徵者的基本背景與錄取與否之資料(表 1)，具有三個條件屬性， $A = \{\text{Diploma, Experience, French, Reference}\} = \{a_1, a_2, a_3, a_4\}$ ，以及一個決策屬性  $D = \{\text{Decision}\} = \{d\}$ ，各屬性之屬性值以及其對應的編碼如下，整理得到表 2：

- Diploma : {MBA, MCE, MSc} = {1, 2, 3}
- Experience : {High, Medium, Low} = {1, 2, 3}
- French : {Yes, No} = {1, 2}
- Reference : {Excellent, Good, Neutral} = {1, 2, 3}
- Decision : {Accept, Reject} = {1, 2}

表1 一個決策表 (或資訊系統)

$U$	Diploma	Experience	French	Reference	Decision
$x_1$	MBA	Medium	Yes	Excellent	Accept
$x_2$	MBA	Low	Yes	Neutral	Reject
$x_3$	MCE	Low	Yes	Good	Reject
$x_4$	MSc	High	Yes	Neutral	Accept
$x_5$	MSc	Medium	Yes	Neutral	Reject
$x_6$	MSc	High	Yes	Excellent	Accept
$x_7$	MBA	High	No	Good	Accept
$x_8$	MCE	Low	No	Excellent	Reject

表2 轉換後的一個決策表

$U$	a1	a2	a3	a4	d
$x_1$	1	2	1	1	1
$x_2$	1	3	1	3	2
$x_3$	2	3	1	2	2
$x_4$	3	1	1	3	1
$x_5$	3	2	1	3	2
$x_6$	3	1	1	1	1
$x_7$	1	1	2	2	1
$x_8$	2	3	2	1	2

其中 $X_{d=1} = \{x_1, x_4, x_6, x_7\}$ ,  $X_{d=2} = \{x_2, x_3, x_5, x_8\}$ 。比較兩兩物件之屬性值，如果屬性不相同表示該屬性可以將兩物件區別分開，於是將這個屬性寫到表3之可區別矩陣內，例如，物件 $x_1$ 與 $x_2$ 之屬性值不相同為a2與a4，於是將a2與a4寫到可區別矩陣內，其餘依此類推。但是如果屬於同一類別的物件，例如， $\{x_1, x_4, x_6, x_7\}$ ，因為屬於同一類別，因此，彼此之間不必加以區別，所以在可區別矩陣中不必加以處理，而以“-”符號加以忽略。

接著將可區別矩陣內的所有屬性值做布林函數 (Boolean function) 計算： $a_1 a_2$ 表示  $a_1 \wedge a_2$ ，亦即a1 and a2。 $(a_1, a_2)$ 表示為  $(a_1 + a_2)$ ，也可以表達為  $(a_1 \text{ or } a_2)$ 。運用布林運算吸收律的觀念，例如： $(a_1 + a_2 + a_3)a_2 = a_2$ ,  $a_2(a_1 + a_2 + a_3)(a_1 + a_3)(a_1 + a_3) = a_2(a_1 + a_3) = a_2 a_1 + a_2 a_3$ 。因此得到化簡如下：

表3 屬性集合A之可區別矩陣

	<i>x1</i>	<i>x2</i>	<i>x3</i>	<i>x4</i>	<i>x5</i>	<i>x6</i>	<i>x7</i>	<i>x8</i>
<i>x1</i>	—							
<i>x2</i>	a2, a4	—						
<i>x3</i>	a1,a2,a4	-	—					
<i>x4</i>	-	a1,a2	a1,a2,a4	—				
<i>x5</i>	a1, a4	-	-	a2	—			
<i>x6</i>	-	a1,a2,a4	a1,a2,a4	-	a2,a4	—		
<i>x7</i>	-	a2,a3,a4	a1,a2,a3	-	a1,a2,a3,a4	-	—	
<i>x8</i>	a1,a2,a3	-	-	a1,a2,a3,a4	-	a1,a2,a3	a1,a2,a4	—

$$\begin{aligned}
 f_A(D) &= (a2, a4) (a1,a2,a4) (a1, a4) (a1,a2,a3) (a1,a2) (a1,a2,a4) (a2,a3,a4) (a1,a2,a4) \times \\
 &\quad (a1,a2,a4) (a1,a2,a3) (a2) (a1,a2,a3,a4) (a2,a4) (a1,a2,a3,a4) (a1,a2,a3) (a1,a2,a4) \\
 &= (a1, a4) a2 = a1a2 + a2a4
 \end{aligned}$$

化簡的結果得到屬性集合為 $a1a2+a2a4$ ，表示以屬性集合 $a1a2$ 或是 $a2a4$ 就可以代表整個資訊表的資訊，亦即利用 $a1a2$ 或是 $a2a4$ 屬性就可以將決策變數予以區別出來。以「 $a1a2$ 」為例，去除屬性 $a3$ 與 $a4$ 後，可得到表4。

表4 化簡後的決策表 (以 $a1a2$ 為例)

<i>U</i>	Diploma (a1)	Experience (a2)	Decision(d)
<i>x1</i>	1	2	1
<i>x2</i>	1	3	2
<i>x3</i>	2	3	2
<i>x4</i>	3	1	1
<i>x5</i>	3	2	2
<i>x6</i>	3	1	1
<i>x7</i>	1	1	1
<i>x8</i>	2	3	2

為了得到決策規則，必須將表4中不重要的屬性值忽略，所謂不重要的屬性是指該屬性值放到決策法則中是多餘的。我們必須將屬性集合{ $a1,a2$ }之可區別矩陣寫在表5，並針對每一行 (或列) 進行以下的布林運算：



表5 屬性集a1與a2之可區別矩陣

	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$	$x7$	$x8$
$x1$	—	a2	a1,a2	-	a1	-	-	a1,a2
$x2$	a2	—	-	a1,a2	-	a1,a2	a2	-
$x3$	a1,a2	-	—	a1,a2	-	a1,a2	a1,a2	-
$x4$	-	a1,a2	a1,a2	—	a2	-	-	a1,a2
$x5$	a1	-	-	a2	—	a2	a1,a2	-
$x6$	-	a1,a2	a1,a2	-	a2	—	-	a1,a2
$x7$	-	a2	a1,a2	-	a1,a2	-	—	a1,a2
$x8$	a1,a2	-	-	a1,a2	-	a1,a2	a1,a2	—

$$f_1(D) = a2(a1,a2)a1(a1,a2) = a1a2$$

$$f_2(D) = a2(a1,a2)(a1,a2)a2 = a2$$

$$f_3(D) = (a1,a2)(a1,a2)(a1,a2)(a1,a2) = a1+a2$$

$$f_4(D) = (a1,a2)(a1,a2)a2(a1,a2) = a2$$

$$f_5(D) = a1a2a2(a1,a2) = a1a2$$

$$f_6(D) = (a1,a2)(a1,a2)a2(a1,a2) = a2$$

$$f_7(D) = a2(a1,a2)(a1,a2)(a1,a2) = a2$$

$$f_8(D) = (a1,a2)(a1,a2)(a1,a2)(a1,a2) = a1+a2$$

彙整布林運算的結果得到表6，在表6中有些屬性值依據布林運算結果予以忽略的有 $x2$ ,  $x4$ ,  $x6$ ,  $x7$ ；必須保留兩個屬性值的有： $x1$ ,  $x5$ ；其中 $x3$ 與 $x8$ 所代表的法則之前提為 $a1=2$ 或 $a2=3$ 兩個屬性值，但是因為同屬 $d=2$ 類別中，已經有 $x2$ 所代表的法則為：「 $a2=3 \rightarrow d=2$ 」，因此， $x3$ 與 $x8$ 所代表法則的屬性 $a1$ 之值可以予以省略(省略的屬性值以“-”表之)，而僅以屬性 $a1$ 之值代表即可。

表6 產生決策法則 (“-”表示忽略該屬性值) (以a1a2為例)

$U$	Diploma (a1)	Experience (a2)	Decision
$x1$	1	2	1
$x2$	-	3	2
$x3$	-	3	2
$x4$	-	1	1
$x5$	3	2	2
$x6$	-	1	1
$x7$	-	1	1
$x8$	-	3	2

因此，由表6中整理得到4個決策(分類)規則如下：

- 1) Experience = 1(High) → Decision = 1 (Accept)
- 2) Experience = 3(Low) → Decision = 2 (Reject)
- 3) Experience = 2(Medium) and Diploma = 1(MBA) → Decision = 1 (Accept)
- 4) Experience = 2(Medium) and Diploma = 3(MSc) → Decision = 2 (Reject)

以同樣的方式處理屬性「a2a4」，而得到表7與表8，由表8中整理得到4個決策(分類)規則，規則如下：

- 1) Experience = 1(High) → Decision = 1 (Accept)
- 2) Experience = 3(Low) → Decision = 2 (Reject)
- 3) Experience = 2(Medium) and Reference = 1(High) → Decision = 1 (Accept)
- 4) Experience = 2(Medium) and Reference = 3(Low) → Decision = 2 (Reject)

表7 化簡後的決策表 (屬性a2a4)

<i>U</i>	Experience (a2)	Reference (a4)	Decision (d)
<i>X1</i>	2	1	1
<i>X2</i>	3	3	2
<i>X3</i>	3	2	2
<i>X4</i>	1	3	1
<i>X5</i>	2	3	2
<i>X6</i>	1	1	1
<i>X7</i>	1	2	1
<i>X8</i>	3	1	2

表8 產生決策法則 (“-”表示忽略該屬性值) (屬性a2a4)

<i>U</i>	Experience (a2)	Reference (a4)	Decision(d)
<i>X1</i>	2	1	1
<i>X2</i>	3	-	2
<i>X3</i>	3	-	2
<i>X4</i>	1	-	1
<i>X5</i>	2	3	2
<i>X6</i>	1	-	1
<i>X7</i>	1	-	1
<i>X8</i>	3	-	2

### 3. 以RST建立EMI診斷系統

#### 3.1 電磁干擾量測資料

EMI標準實驗室使用頻譜分析儀 (Spectrum Analyzer)，測試產品所得到的EMI資料及圖形如圖2所示，在圖中的輸出頻率 (Frequency) 數據資料是由30MHz到1GHz的範圍，每1MHz取得一數值 (即垂直高度)，此數值的單位稱之為dB，經一段測試時間後此dB數值會累增到最高，將其所測得每一點MHz的dB單位之最高數據值取出，並結合每一點連成一個曲線圖。曲線圖的每一個點都不得超過規定的標準高度，即30MHz~230MHz是30dB以下，231MHz~1GHz是37dB以下。如果有任何一點的最高值超出了此標準高度，則被判定為不符合EMI法規標準值。EMI工程師必須對此頻率點，找出造成不良的真正原因並進行修正，直到每點的頻率數值皆通過法規所訂定的標準值以下。

在圖2的頻率波形圖中所標示數字，其詳細數據資料顯示在下半部的表中，表中第一欄位為頻率點數字編號，第二欄位的「Freq.(MHz)」為不良EMI頻率點，最後一欄「Warning Mark」的標示中，符號“X”代表測試不合格，符號“!”代表在危險邊緣，至於沒有標記的頻率點，則是列為觀察對象。

本研究即是蒐集這些現場的EMI測試歷史資料，做為資料探勘之依據。

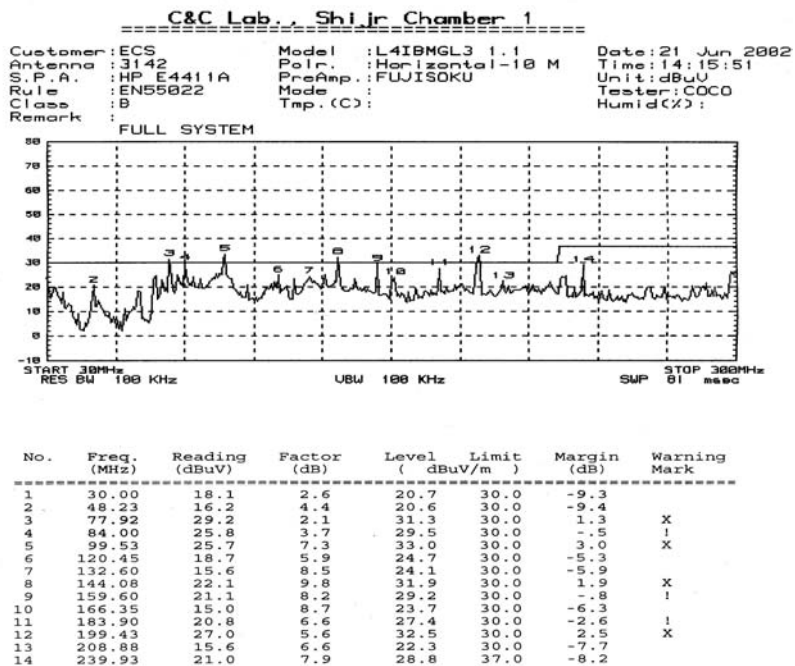


圖2 EMI 頻率波形圖與數據資料

### 3.2 EMI診斷規則探勘步驟

欲以RST建立EMI診斷系統，其步驟如圖3所示，分別經過資料蒐集 (Data Collection) → 資料前處理 (Data Preprocessing) → 屬性值離散化 (Discretization) → 屬性化簡 (Attribute Reduction) → 最小屬性集過濾 (Reduction Filtering) → 產生法則 (Rule Generation) → 過濾法則 (Rule Filtering) → 做分類 (Classification) → 計算分類正確率 (Accuracy Calculation) → 建立診斷系統 (Diagnostic System)，分別說明如下：

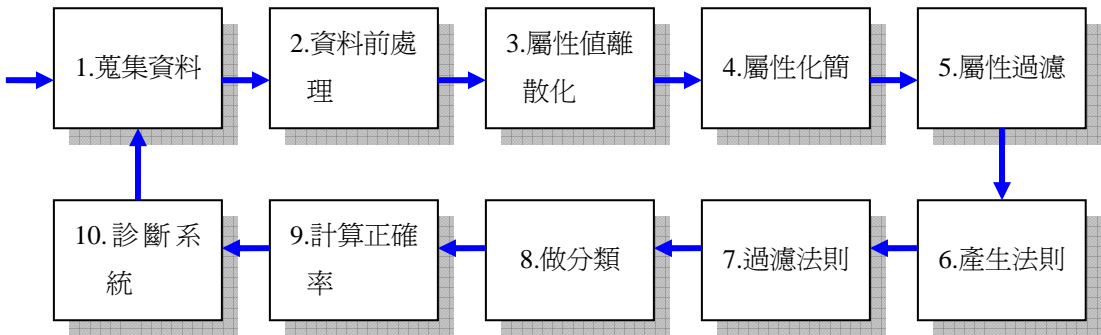


圖3 以RST建立EMI診斷系統的步驟

#### (1) 資料蒐集

從測試資料庫中蒐集EMI不良問題之相關資料，並進行整理轉換建立資料倉儲。蒐集測試不良的主機板型式等11個輸入屬性及1個目標屬性 (如表9所示)，包括產品資料的詳細規格、使用的週邊設備、以及頻譜分析儀之各頻率測試結果的數據。共蒐集國內某知名主機板製造公司的EMI測試不良資料共415筆資料，其中高頻有134筆，低頻有281筆。

#### (2) 進行資料前處理

對蒐集的資料進行整理，將來自於不同檔案的資料合併為一個資料表，將有遺漏值的資料去除，決定各屬性值的編碼。並且將可能產生問題的Port，依據位置來區分為三區 (如圖4)，分別為：A區 (Printer, Ps2, Com)、B區 (VGA)、C區 (LAN, USB, Audio, Game)，以此三區為目標屬性的值。

#### (3) 屬性值離散化

所有屬性只有「測試頻率點」這個變數為連續數值變數，其餘變數皆為名目尺度或是順序尺度。因此將「測試頻率點」離散化為10個間斷的區間，分區的原則是依據領域工程師對於頻率點可能的分佈情況而決定。

表9 輸入屬性與目標屬性

屬性名稱	說明	屬性值個數			性質	代號
		低 頻	高 頻	全 頻		
Chipset	晶片組廠牌	3	3	3	名目尺度	輸入屬性a1
IO Chipset	I/O晶片組(Printer,Com,PS2)	7	3	7	名目尺度	輸入屬性a2
Audio	音效晶片廠牌	3	2	3	名目尺度	輸入屬性a3
Memory	記憶體規格	2	2	2	名目尺度	輸入屬性a4
PCB-size	主機板尺寸規格	2	2	2	名目尺度	輸入屬性a5
CPU-slot	CPU插槽規格	5	3	5	名目尺度	輸入屬性a6
USB	USB規格	2	2	2	名目尺度	輸入屬性a7
LAN	LAN晶片廠牌	4	4	4	名目尺度	輸入屬性a8
VGA	VGA位置	4	3	4	名目尺度	輸入屬性a9
Failure-Level	EMI Noise等級	3	3	3	順序尺度	輸入屬性a10
Frequency-Level	測試頻率點	8	5	3	連續數值轉成順序尺度	輸入屬性a11
Failure-Area	EMI不良位置	3	3	3	名目尺度	目標屬性d

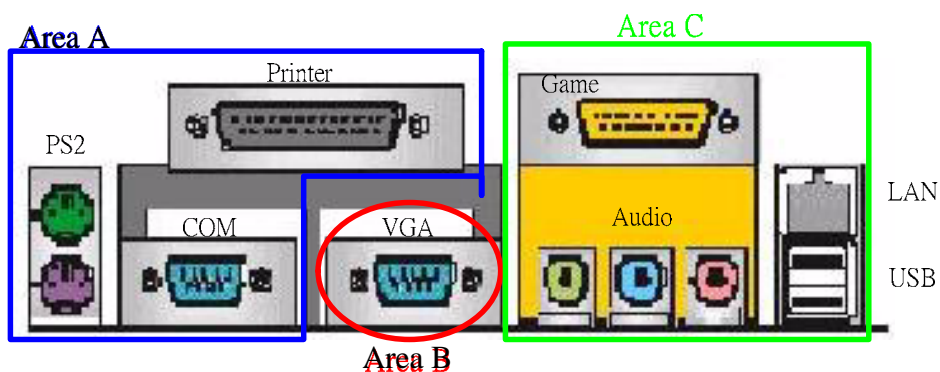


圖4 將目標屬性的值區分為A、B、C三個區域。

#### (4) 屬性化簡

國際規定位於30-1G MHz之間必須測量EMI之dB值，區分為低頻 (30-230 MHz)，高頻 (230-1G)，兩種頻率有不同的標準，因此依據領域工程師的意見，分別針對低頻、高頻做屬性化簡。

屬性化簡的方法有多種 (Johnson, 1974; Vinterbo, 2000)，本研究採用後者的基因演算法，其演算法的配適函數 (Fitness function) 設計的原則為屬性集合可以包含到所有案例的比率。Rosetta (2001) 是一套以約略集合理論為基礎的軟體，此軟體將這些方法實作在內，經由此軟體的協助計算，得到屬性化簡的結果為：

低頻得到2個Set：

{ Chipset, Audio, Memory, PCB-size, CPU-slot, LAN, Failure-Level, Freq-Level }

{ Chipset, Audio, PCB-size, CPU-slot, LAN, VGA, Failure-Level, Freq-Level }

高頻得到3個Set：

{ Chipset, VGA, Failure-Level, Freq-Level }

{ Chipset, Memory, LAN, Failure-Level, Freq-Level }

{ Chipset, PCB-size, LAN, Failure-Level, Freq-Level }

#### (5) 最小屬性集合過濾

屬性化簡得到的答案不唯一，可能得到多個最小屬性集合，到底要採用哪一個屬性集合，除了要比較分類結果的精確度之外，也要參考領域專家的意見。

本研究經由與具備現場經驗的管理者的討論，低頻選擇 {Chipset, Audio, Memory, PCB-size, CPU-slot, LAN, Failure-Level, Freq-Level} 最小屬性集合當作輸入屬性。高頻選擇 {Chipset, VGA, Failure-Level, Freq-Level} 最小屬性集合當作輸入屬性。

#### (6) 產生法則

前述低頻最小屬性集合產生166條法則，高頻產生22條法則。其中一條法則範例如下：

Chipset (Via) AND Audio (CMI9738) AND Memory (SDRAM) AND PCB-size (Micro-ATX) AND CPU-slot (Socket478) AND LAN (None) AND Fail-Level (C) AND Freq-Level(2) => Fail-area (A)

這個法則的解釋是：如果主機板的Chipset是「Via」，而且Audio是「CMI9738」，而且Memory是「SDRAM」，而且PCB-size是「Micro-ATX」，而且CPU-slot是「Socket478」，而且LAN是「None」，而且Fail-Level為「C」，而且Freq-Level為「2」，則不良的區域在「A區」。其意義是主機板所用的組件搭配 (Chipset, Audio,...等)、產生EMI的頻率點 (Freq-Level)、以及其不良的嚴重性 (Fail-Level)可以判斷EMI不良發生的位置在A區 (可能是Printer, Ps2或Com)。

#### (7) 過濾法則

必要時可以對法則做過濾，將預測能力較差的法則過濾掉，保留預測能力較佳的法則。通常此步驟要結合領域專家的知識，以判斷這些法則的合理性，而決定法則是否刪除或保留。

#### (8) 做分類

當有新的物件資料要分類時，依據的原則有以下幾種狀況 (Slowinski, 1994)：

- 1) 新的物件剛好符合一條確定性規則 (Deterministic Logical Rule)。

- 2) 新的物件符合多條規則，但是其決策屬性值都相同，此時不會產生混淆問題（多對一）。但是如果其決策屬性值不相同（多對多），此時必須計算哪一個決策屬性值的可能性比較高，以較多樣本數支持的法則為準。
- 3) 新的物件剛好符合一條未確定規則（Non-deterministic Logical Rule），亦即，符合一條相同的前提屬性值，但是其決策屬性值卻有多個（一對多），這樣的法則此時必須計算哪一個決策屬性值的可能性比較高，以較多樣本數支持的法則為準。
- 4) 新的物件不符合任何一條規則，此時必須找「鄰近」較為類似的法則加以運用。

(9) 計算正確率

本研究以前一個步驟所得到的分類法則對這些資料做分類，得到分類正確率為低頻90.39%，高頻為71.64%（詳如表10及表11）。

表10 低頻分類準確度

目標屬性值	A	C	B	準確度
A	72	3	0	96%
C	23	179	0	88.61%
B	0	1	3	75%
準確度	75.79%	97.81%	100%	90.39%

表11 高頻分類準確度

目標屬性值	A	C	B	準確度
A	46	12	0	79.31%
C	18	42	0	70%
B	4	4	8	50%
準確度	67.65%	72.41%	100%	71.64%

(10) 建立診斷與預防系統

將EMI診斷法則建立一個診斷知識庫系統，當EMI工程師測量得到EMI品質不良主機板的數據以及主機板搭配的零組件資訊後，將輸入屬性資料輸入到EMI診斷決策系統，根據知識庫內的診斷法則，系統會推論出不良位置的建議，工程師優先診斷這個位置，平均而言有81%的正確率（低頻90%，高頻72%）。

當新的EMI不良資料累積之後，必須將新的資料納入訓練集合，並排除過時的資料，重新回到步驟(1)，這樣可以讓診斷法則保有最新的診斷知識。

### 3.3 結果分析與比較

決策樹 (Decision Tree) 是最常用的一種資料探勘方法，有許多文獻採用決策樹獲致頗佳的分類準確率。一方面因為其實用性，另一方面因為決策樹適用在本研究的名目尺度資料型態的探勘，因此，本研究也採用決策樹建立EMI的診斷規則，並與EMI診斷規則的準確性做比較。

決策樹的方法有多種，ID3 (Quinlan, 1986) 與C4.5 (Quinlan, 1993) 是常被使用的兩種決策樹方法，這兩種方法的最新後繼產品版本是“C5.0” (<http://www.rulequest.com/>)，本研究採用C5.0建立EMI的診斷法則。

本研究以資料集的三分之一做為測試資料 (Testing Set)，另外三分之二做為訓練資料 (Training Set)，得到如表12的平均測試正確率結果。以C5.0決策樹得到的法則做為診斷法則，平均有77%測試準確度，使用11個輸入屬性資訊 (表12)。雖然RST的準確度只有高出一些，但是RST使用較少的屬性個數 (低頻8個，高頻4個)，而得到差不多的分類準確度，因此，對本研究個案公司之研究資料而言，RST探勘方法之結果頗令人滿意。

表12 與決策樹得到之結果比較

	使用之屬性個數	低頻	高頻	平均測試正確率
RST	低頻：8 高頻：4	88.4%	70.6%	79.5%
決策樹	低頻：11 高頻：11	85.4%	68.7%	77%

## 4. 結論

本研究針對EMI診斷問題，蒐集EMI不良歷史資料，應用資料探勘中常用的知識發掘方法—RST找到診斷規則知識，並建立診斷決策系統。EMI工程師依據此診斷規則找出EMI不良之可能根源，可以迅速解決在研發與製造階段產生的EMI不良問題，對縮短研發時間與降低品質成本有重大的助益。

以國內知名主機板廠商的實務資料為對象，採用RST理論，依循本研究所提出的九個步驟進行，所建立的診斷決策系統，其平均準確度約為八成，達到實務應用上可接受的要求程度。

在實際運用時，因為生產現場隨時動態變化，若有新的案例產生，必須將最新的案例加入訓練案例中，重新進行前述探勘的步驟(1)~(9)，以更新診斷規則，得到最符合現況的診斷系統。

後續研究著重在改良方法，或是結合不同的探勘方法，得到整合系統 (Hybrid System)，以增加準確率。例如，當新的案例輸入，找不到對應的診斷法則，或是找到的診斷法則有多種可能



結果時，此時可以使用其它的替代法則（由不同的探勘方法得到，例如，神經網路、基因演算法等），以補足RST的診斷法則所無法含括的部分。

## 5. 致謝

本研究為國科會計畫編號NSC-92-2213-E-211-012之部分成果。

## 參考文獻

- [1] 簡禎富、林鼎浩、彭誠湧、徐紹鐘，「建構半導體晶圓允收測試資料挖礦架構及其實證研究」，工業工程學刊，第 18 卷，第 4 期，民國 90 年，第 37-48 頁。
- [2] Archambeault, B. and Connor, S., "Measurements and Simulations for Ground-to-Ground Plane Noise for SDRAM and DDR RAM Daughter Cards and Motherboards for EMI Emissions," *IEEE International Symposium on EMC*, 2002, pp.105-108.
- [3] Ahna, B. S., Cho, S. S. and Kim, C. Y., "The Integrated Methodology of Rough Set Theory and Artificial Neural Network for Business Failure Prediction," *Expert Systems with Applications*, Vol. 18, 2000, pp.65-74.
- [4] Bayardo, R. J., Rakesh, A. and Dimitrios G., "Constraint-Based Rule Mining in Large, Dense Databases," *Proceeding of the 15th International Conference on Data Engineering*, 1999, pp.188-197.
- [5] Ben-Arieh, D., Chopra, M. and Bleyberg, M. Z., "Data Mining Application for Real-time Distributed Shop Floor Control," *IEEE International Conference on Systems, Man, and Cybernetics*, vol. 3, 1998, pp.2738-2743.
- [6] Chen, F.-L. and Liu, S.-F., "A Neural-Network Approach To Recognize Defect Spatial Pattern In Semiconductor Fabrication," *IEEE transactions on semiconductor manufacturing*, 2000, Vol. 13, No. 3, pp.366-373.
- [7] Dimitras, A. I. Slowinski, R., Susmaga, R. and Zopounidis, C., "Business Failure Prediction Using Rough Sets," *European Journal of Operational Research*, Vol. 114, 1999, pp.263-280.
- [8] Gardner, R. and Bieker, J., "Solving Tough Semiconductor Manufacturing Problems Using Data Mining," *IEEE/SEMI Advanced Semiconductor Manufacturing Conference*, 2000, pp. 46-55.
- [9] Han, J. and Kamber, M., *Data Mining: Concepts and Techniques*, San Francisco, CA: Morgan Kaufmann Publishers, 2001.
- [10] Hui, S. C. and Jha, G., "Data Mining for Customer Service Support," *Information &*

- Management*, Vol. 38, 2000, pp.1-13.
- [11] Johnson, D. S., "Approximation Algorithms for Combinatorial Problems." *Journal of Computer and System Sciences*, Vol. 9, 1974, pp.256–278.
- [12] Kusiak, A., "Decomposition in Data Mining: An Industrial Case Study," *IEEE Transactions on Electronics Packaging Manufacturing*, Vol. 23, No. 4, 2000, pp.345-353.
- [13] Kusiak, A., "Rough Set Theory: A Data Mining Tool for Semiconductor Manufacturing," *IEEE Transactions on Electronics Packaging Manufacturing*, Vol. 24, No. 1, 2001, pp.44-50.
- [14] Lee, J.-H., Yu, S.-J. and Park S.-C., "Design of Intelligent Data Sampling Methodology Based on Data Mining," *IEEE transactions on semiconductor manufacturing*, 2001, Vol. 17, No. 5, pp.637-649.
- [15] McKee, T. E. and Lensberg, T., "Genetic Programming and Rough Sets: A Hybrid Approach to Bankruptcy Classification." *European Journal of Operational Research*, Vol. 138, pp.436 – 451, 2002.
- [16] Mills, J. P., *Electromagnetic Interference*. New Jersey: Prentice Hall, 1993.
- [17] Sushmita, M., Sankar, K. P. and Pabitra, M., "Data Mining in Soft Computing Framework: A Survey," *IEEE Transactions on Neural Networks*, Vol. 13, No. 1, 2002, pp.3-14.
- [18] Morgan, D., *A Handbook for EMC Testing and Measurement*, IEEE Electrical Measurement Series 8, London: Peter Peregrinus, 1994.
- [19] Pawlak, Z., "Rough Sets." *International Journal of Information and Computer Sciences*, 11, 1982, pp.341-356.
- [20] Pawlak, Z., *Rough Sets: Theoretical Aspects of Reasoning About Data*, Boston: Kluwer Academic Publishers, 1991.
- [21] Pawlak, Z., and Slowinski, R., "Rough Set Approach to Multiattribute Decision Analysis." *European Journal of Operational Research*, Vol. 72, 1994, pp.443-459.
- [22] Pawlak, Z., Grzymala-Busse, J. W., Slowinski, R. and Ziarko, W., "Rough Sets." *Communications of the ACM*, Vol. 38, No. 11, 1995, pp.89-95.
- [23] Rosetta, version 1.4.41 (2001. 5. 25), <http://www.idi.ntnu.no/~aleks/rosetta/>
- [24] Quinlan, J. R., "Induction of Decision Trees," *Machine Learning*, Vol. 1, No. 1, 1986, pp.81-106.
- [25] Quinlan, J. R., *C4.5: Programs for Machine Learning*, San Francisco, CA: Morgan Kaufmann Publishers, 1993.
- [26] Skinner, K. R., Montgomery, D. C., Runger, G. C., Fowler, J. W., McCarville, D. R. Rhoads, T. R. and Stanley, J. D., "Multivariate Statistical Methods for Modeling and Analysis of Wafer

- Probe Test Data,” *IEEE transactions on semiconductor manufacturing*, 2002, Vol. 15, No. 4, pp.523-530.
- [27] Slowinski, R. and Stefanowski, J., “Rough Classification with Valued Closeness Relation,” in: E. Diday, Y. Lechevallier, M. Schader, P. Bertrand, B. Burtschy Eds., *New Approaches in Classification and Data Analysis*, Berlin: Springer, 1994.
- [28] Vinterbo, S. and Øhrn, A., “Minimal Approximate Hitting Sets and Rule Templates.” *International Journal of Approximate Reasoning*, Vol. 25, No.2, 2000, pp.123–143.
- [29] Walczak, B. and Massart, D. L., “Rough Set Theory”, *Chemometrics and Intelligent Laboratory Systems*, Vol. 47, 1999, pp.1–16.
- [30] Yoshida, T. and Touzaki, H., “A Study on Association Among Dispatching Rules in Manufacturing Scheduling Problems,” *IEEE International Conference on Emerging Technologies and Factory Automation*, vol. 2, 1999, pp.1355–1360.