

利用正規化概念分析探勘分類關聯法則之研究

黃錦法 彭浩政 徐博恩

雲林科技大學資訊管理所

{ Huangcf, M10123058, M10023055 }@yuntech.edu.tw

摘要

資料探勘可以從大量資料中找出一些有意義的資訊。資料探勘的模型主要有資料分類、資料關聯、資料分群及循序樣式探勘等。其中，資料關聯(關聯法則)是探討資料項目間的關係，找出在某一事件或是資料中會同時出現的項目。資料分類是依據資料的屬性來將資料分門別類，有結合資料關聯的關聯法則分類法。正規化概念分析是一種從資料集中發現概念結構的資料分析理論，它可以將表格格式的正規化本文轉換成圖形化的概念點陣。

本研究提出以正規化概念分析為基礎探勘分類關聯法則的方法，對資料做分類。本研究利用正規化概念分析建立概念點陣，並由概念點陣推導分類關聯法則，且對分類關聯法則做修剪及排序，以優先順序越高的法則越先使用。為了評估資料分類的成效，本研究以 UCI 網站上所提供的資料集進行實驗，評估指標為分類正確率及所需時間。實驗結果顯示，本研究的方法資料分類的速度相當的快，經由最小支持度和最小信賴度的調整，也能提高資料集的正确率。另外，本研究也在網際網路上架設分類預測系統，使用者透過瀏覽器即可使用。

關鍵詞： 正規化概念分析、概念點陣、資料探勘、資料分類、關聯法則。

Abstract

Data mining can find some interest information from large amounts of data. The models of data mining are data classification, data association, data clustering, sequential pattern mining, and so on. Data association (association rules) can find associations among data items. Data classification distinguishes every data from a data set or group, and it also can combine data association. Formal concept analysis is a data analyzing theory that discovers concept structure in data sets. It can transform formal context into concept lattice.

This study applies association rules for classification based on formal concept analysis to classify the data. The proposed method creates concept lattice by using formal concept analysis, and generates association rules for classification from concept lattice. The rules will be pruned and sorted, and it will be used by following the priority order. In order to estimate the performance of data

classification, experiments are done through a data set from UCI website. The evaluation indicators are correct rate and execute time. The result of experiments shows that the execute time of data classification is short; and the correct rate can increase after adjusting minimum support and minimum confidence. Furthermore, this study also develops a classification forecast system on the Internet, and the user can use this system by a web browser.

Keywords: Formal Concept Analysis, Concept Lattice, Data Mining, Data Classification, Association Rules.

1. 前言

隨著資訊科技及網際網路的進步，人們可獲得的資料量與日俱增。如何從大量的資料中，萃取潛藏的資訊，幫助人們做決策是一個很重要的議題。資料探勘(Data Mining)可以從資料庫中發現知識，將一些隱藏的、事先未知的、潛在的資訊，從資料庫中擷取出正確且有意義的資訊[1]。

資料探勘的模型主要有以下 4 種:資料分類(Data Classification)、資料關聯(Data Association)、資料分群(Data Clustering)以及循序樣式探勘(Sequential Pattern Mining)[2]。其中，資料關聯是探討資料項目間的關係，找出在某一事件或是資料中會同時出現的項目，也叫做關聯法則探勘。資料分類是指根據已知的資料及其類別屬性來建立資料的分類模型。分類模型的建立可以讓我們了解屬於各種類別屬性的資料具備哪些特徵，同時也可以用來預測新進資料的類別屬性。資料分類方法主要有決策樹、貝氏分類法、記憶基礎推論法及類神經網路等[2]，也有結合資料關聯的關聯法則分類法[3]~[5]。

正規化概念分析(Formal Concept Analysis, FCA)是一種從資料集中發現概念結構的資料分析理論，它可以將表格格式的正規化本文(Formal Context)轉換成圖形化的概念點陣(Concept Lattice)[6]。

在正規化概念分析前，需將資料集合轉換成正規化本文，正規化本文是由物件、屬性及物件與屬性間二元關係所組成的。正規化本文經正規化概念分析後會產生正規化概念(Formal Concept)，正規化概念是由涵義(Intent)及範圍(Extent)所組成的。涵義是由屬性所組成的，且涵義中的所有屬性必須歸屬於範圍中的物件；範圍是由物件所組成的，且範圍

中的所有物件必須擁有涵義中的屬性。概念點陣是將正規化概念間最小屬性集合排序而產生的，是一種圖形結構。

本研究首先利用原始資料集產生訓練資料集及測試資料集，接著將訓練資料集轉換成正規化本文，然後對訓練資料集的正規化本文做正規化概念分析產生正規化概念及概念點陣。並且，由概念點陣推導出分類關聯法則，也對分類關聯法則做修剪及排序。本研究的分類模型是以分類關聯法則為主，並利用分類關聯法則的優先順序，判斷測試資料集的資料分類。

以下各章節的內容如下：第 2 章是文獻探討，第 3 章是研究方法，第 4 章是系統實驗與試用，第 5 章是結論。

2. 文獻探討

文獻探討主要分成資料分類與正規化概念分析等 2 部分。

2.1 資料分類

資料分類是依據資料的屬性來將資料分門別類。資料分類方法主要有決策樹、貝氏分類法、記憶基礎推論法及類神經網路等，也有結合資料關聯的關聯法則分類法。

(1) 決策樹

決策樹是一種樹狀結構，樹的非葉(Non-Leaf)節點代表屬性的測試條件，樹的分支(Branch)代表屬性的測試結果，樹的葉節(Leaf)點代表分類屬性的分類結果。基本的決策樹分類法有 ID3、C4.5 及 CART 等。

(2) 貝氏分類法

貝氏分類法是一種以機率、統計學為基礎的分類法，它利用事件發生機率來推測未知資料類別，適合用在預測未知樣本的類別，而不適合用來找出資料分類的原因。

貝氏分類法的理論基礎是貝氏定理。

$$P(C | X) = \frac{P(X | C)P(C)}{P(X)}$$

X 代表某個未知案例，C 代表某一類別，X 案例屬於 C 類別的機率=(C 類別中出現 X 案例的機率)×(C 類別出現的機率)/(X 案例出現的機率)。

(3) 記憶基礎推論法

記憶基礎推論法是根據 Roger Schank 的動態記憶法所提出的一種分類法，它從過去的經驗知識中擷取相似案例來解決問題。

記憶基礎推論法的成功關鍵:1)選取合適的訓練資料集;2)正確的資料精簡處理;3)決定適當的距離函數、組合函數以及鄰近樣本個數。

(4) 類神經網路

類神經網路是一種人工智慧，它模擬人類大腦神經細胞的運作方式。類神經網路經過簡化成為一

個單向多連結的架構，通常分為輸入層、隱藏層與輸出層三部分(圖 1)。

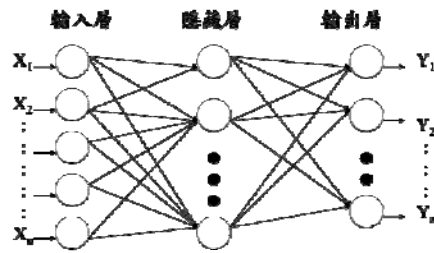


圖 1 類神經網路範例

類神經網路在資料分類上的應用包含訓練階段及測試階段等 2 個階段。訓練階段主要調整網路內部各節點連結的權重值，使得輸入資料屬性值經過網路計算之後能得到目標的輸出分類屬性值。測試階段則為驗證網路的分類準確度或是利用訓練完成的網路進行分類預測。

(5) 關聯法則分類法

關聯法則分類法為資料探勘模型中的資料關聯(關聯法則)與資料分類的結合。關聯法則分類主要利用分類關聯法則來判斷或預測資料的分類。目前有以高頻項目集為基礎探勘分類關聯法則的方法 [3],[4]，也有以概念點陣為基礎探勘分類關聯法則的方法 [5]。

2.2 正規化概念分析

正規化概念分析是由 Rudolf Wille 於 1982 年提出來的，它是一種從資料集中發現概念結構的資料分析理論。

正規化概念分析，可以由物件及物件所擁有的屬性所組成的正規化本文，導出正規化概念，並且可以將正規化概念轉換成概念點陣。以下將針對正規化本文、正規化概念、概念點陣及正規化概念分析工具等做個別的說明。

(1) 正規化本文

正規化本文是由(G, M, I)三個部分所組成，表示成 Context: C=(G, M, I)。G 為物件的集合，M 為屬性的集合，而 I 則為 G 和 M 所構成之二元關係的組合，定義為(g, m)I 或是 gIm，意思為「物件 g 的屬性為 m」。

一個正規化本文可以用一個二元的表格來呈現，稱為交叉表 [7]。表格中的列代表物件，欄則代表屬性，而列與欄的相交處即代表該物件具有該屬性，正規化本文的表示如表 1 所示。這是一個以人類為物件，其喜好食物為屬性的範例，表中的 X 代表某人喜好某食物。

(2) 正規化概念

當正規化本文建立之後，就可以導出正規化概念，一個正規化概念可以表示成一組(A, B)，其中 A 為正規化本文中物件的子集合，而 B 則是正規化本文中屬性的子集合，一個以物件為觀點，另一個以

屬性為觀點。

表1 正規化本文範例

屬性 物件	Fish	Beef	Pork	Chicken
Fred	X			X
Jess	X	X	X	
Bob	X		X	
Mel	X		X	X

為了要使(A, B)成為正規化概念, B 所有屬性必須歸屬於物件 A, 稱為涵義, 而 A 所有物件必須擁有 B 中的屬性, 稱為範圍。兩種關係的定義如下:

A' 為所有屬性必須歸屬於物件 A, 稱為涵義, 定義為:

$$A' := \{m \in M \mid (g, m) \in I \text{ for all } g \in A\}$$

B' 為所有物件必須擁有 B 中的屬性, 稱為範圍, 定義為:

$$B' := \{g \in G \mid (g, m) \in I \text{ for all } m \in B\}$$

任何一組正規化概念(A, B)必須滿足下列條件:

$$A = B', B = A'$$

以表 1 為測試的範例, 若從物件觀點來看, 將 Jess、Bob 與 Mel 視為物件子集, 在試圖尋找可能的正規化概念時, 透過定義可找出三物件所共同屬性群有 Fish 及 Pork, 其餘屬性皆無法同時吻合; 另一方面, 從屬性觀點來看, 將 Fish 與 Pork 視為屬性子集, 透過定義即可找出兩屬性所共同具有的物件群為 Jess、Bob 與 Mel。因此可以得出一正規化概念 $\{\{Jess, Bob, Mel\}, \{Fish, Pork\}\}$, 此一概念的範圍為 Jess、Bob 與 Mel, 概念的涵義為 Fish 與 Pork。

(3) 概念點陣

概念點陣是透過概念之間最小屬性集合排序而產生。為了產生概念點陣, 首先須找出正規化概念之間子概念-父概念的關係。其定義為:

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 (\Leftrightarrow B_2 \subseteq B_1)$$

其中, (A_1, B_1) 稱為 (A_2, B_2) 的子概念, 而 (A_2, B_2) 則稱為 (A_1, B_1) 的父概念, 關聯符號 \leq 稱為概念的階層排序(Hierarchical Order)[8], 代表這兩者之間存在繼承關係。

當要展現圖形化的概念點陣時, 必須決定每個概念的前任者和後繼者, 前任者可以由每個涵義的最大子概念得到, 而後繼者則可以由每個涵義的最小父概念得到[9]。當概念點陣形成之後, 最大的子概念會放在概念點陣的最上方, 稱為上確界(Supremum), 而最小的子概念會放在最下方, 稱為下確界(Infimum)。

圖 2 為表 1 正規化本文範例轉換而成的概念點陣。每個節點在概念點陣上都代表一個概念, 節點上方表示的為此概念的涵義, 下方則是範圍。由圖 2 可以得知 $\{Fish\}, \{Fred, Jess, Bob, Mel\}$ 為上確界, 而 $\{Fish, Beef, Pork, Chicken\}, \{\}$ 為下確界。

(4) 正規化概念分析工具

目前正規化概念分析領域中, 已有相當多的開

發工具, 而且大部分的工具都是使用 Java 語言來實作的, 具有跨平台的特性。例如, Concept Explorer、ToscanaJ 及 Galicia 等都是屬於此類的正規化概念分析工具[10], 這三種工具的功能比較表可以參考文獻[11]。

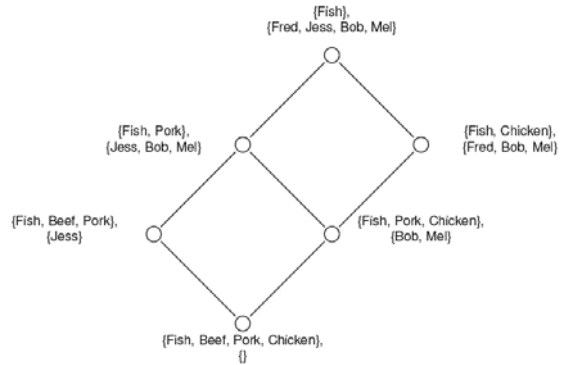


圖 2 概念點陣範例

3. 研究方法

3.1 系統架構

本研究的系統架構由資料準備、分類訓練、分類測試及分類預測等 4 個階段所組成(圖 3)。

- (1) 資料準備階段是由原始資料集來產生訓練資料集及測試資料集, 以供分類訓練階段及分類測試階段使用。
- (2) 分類訓練階段是由訓練資料集建立正規化本文, 接著使用正規化概念分析產生概念點陣檔, 然後轉換概念點陣檔建立概念點陣資料庫, 最後利用概念點陣資料庫推導分類關聯法則, 以供分類測試階段使用。
- (3) 分類測試階段是利用分類關聯法則及其優先順序來測試分類測試資料集, 並將結果放在分類結果資料庫; 結果分析是統計分析分類結果, 並將結果顯示出來。
- (4) 分類預測階段是使用者可以輸入單筆資料, 將利用分類關聯法則及其優先順序來做分類預測, 並即時顯示分類預測結果給使用者參考。此階段可以透過瀏覽器來線上執行。

3.2 資料準備階段

資料準備階段分為產生訓練資料集及產生測試資料集等 2 個部分。產生訓練資料集是由原始資料集中取一定比例的資料來當訓練集; 而產生測試資料集是以除了訓練集以外的原始資料集的資料來當測試集。當原始資料集有提供訓練集及測試集時則直接使用。

一般常用的比例是 2/3 當訓練集, 1/3 當測試集; 但也可以選取 10%~90% 當訓練集, 90%~10% 當測試集。

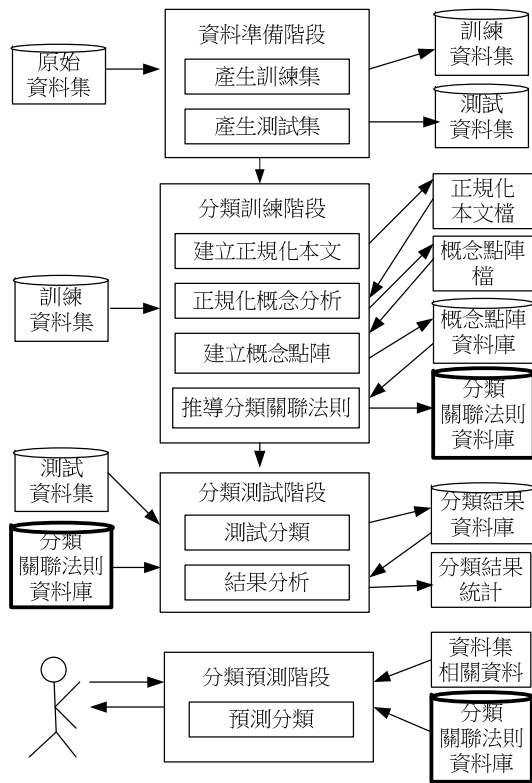


圖 3 系統架構圖

3.3 分類訓練階段

分類訓練階段有建立正規化本文、正規化概念分析、建立概念點陣及推導分類關聯法則等 4 個步驟。

(1) 建立正規化本文

將訓練資料集的資料轉換成適合正規化概念分析的正規化本文，正規化本文是一種 XML 檔案，它是根據 Galicia 的檔案格式。

(2) 正規化概念分析

將訓練資料集經由正規化概念分析，產生概念點陣。本研究使用正規化概念分析工具 Galicia 來完成此步驟。Galicia 中有許多正規化概念分析的演算法可以使用，本研究使用 Godin(1995) 演算法。概念點陣也是一種 XML 檔案。

Galicia 所產生的概念點陣檔包含有物件、屬性及概念節點等資訊。其中，概念節點包含範圍(Extent)、內涵(Intent)及父節點(Super_Node)等資訊。

(3) 建立概念點陣

將概念點陣檔轉換成資料庫方式儲存在概念點陣資料庫中。此步驟是為了後續方便處理概念點陣，因此將檔案格式轉換成資料庫格式，並且將產生子節點資訊，以利概念點陣的搜尋。

Galicia 所產生的概念點陣只有父節點資訊，並沒有子節點資訊，因此在檔案轉換成資料庫儲存的同時，產生各節點的子節點資訊。

(4) 推導分類關聯法則

利用概念點陣資料庫推導分類關聯法則，主要分成法則產生及分類器建構等 2 個步驟。

1) 法則產生

由上而下將概念點陣中每一個節點搜尋一次。節點內的範圍(Extent)就是物件、內涵(Intent)就是屬性。利用節點的物件計算支持度，判斷每一個節點有沒有大於或等於最小支持度，把符合條件(大於或等於)的節點留下。從留下來的節點中，利用節點的物件和類別屬性值計算信賴度，判斷有沒有大於或等於最小信賴度，符合條件(大於或等於)的節點將推導出分類關聯法則。

假設某一個節點，它的物件為{何偉業,是、張明雄,是、林葳婷,否、田信程,是}有 4 筆，屬性為{年齡大於 30}有 1 個。訓練資料集有 6 筆資料，最小支持度設為 0.2、最小信賴度設為 0.6。則此節點的支持度為 4/6 (0.67)，大於最小支持度。其信賴度有 2 個，類別屬性值為{是}的有 3 筆，信賴度為 3/4 (0.75)、類別屬性值為{否}的有 1 筆，信賴度為 1/4 (0.25)。類別屬性值為{是}的信賴度大於最小信賴度，因此推導出分類關聯法則{年齡大於 30→是}。

支持度及信賴度的定義如下：

$$\text{支持度(Support)} = \frac{|A|}{|D|}$$

$$\text{信賴度(Confidence)} = \frac{|A \cap B|}{|A|}$$

其中，|A| 是指節點的物件(資料)筆數；|D| 表示在訓練資料集中所有資料(物件)筆數。而 |A ∩ B| 是指節點的物件(資料)中類別屬性值 B 筆數。

分類關聯法則的定義如下：

$$\text{分類關聯法則} = \{\text{Cond}_i \rightarrow B_i, i \geq 0\}$$

其中，分類關聯法則可以由 0 條或多條所組成。Cond_i 是指第 i 條法則的條件部分，它是由節點的屬性所構成；B_i 則是第 i 條法則的結論部分，它是節點的物件中類別屬性值 B_i。

2) 分類器建構

當分類關聯法則推導完成後，對多餘及衝突的法則做修剪(Pruning)，並將法則做排序，主要目的是在減少不必要的法則及增加分類器的效率。分類器建構主要有以下 6 個步驟：

1. 刪除多餘(Redundancy)法則

對推導出的分類關聯法則做修剪，以避免發生法則多餘的情況。多餘法則的刪除方法如下：

檢查每一條法則 XY → C_i，假如已經存在 X → C_i，就把多餘的 XY → C_i 刪掉，保留 X → C_i。

2. 對分類關聯法則做排序(Sort)

為了解決法則衝突時之優先順序，先將分類關聯法則做排序。假設有 2 條法則 R_i 和 R_j，若符合下列條件時，R_i 的優先順序大於 R_j 的優先順序(R_i > R_j)。

- R_i 的信賴度大於 R_j 的信賴度。
- 如果 R_i 的信賴度等於 R_j 的信賴度，但 R_i 的支持度大於 R_j 的支持度。
- 如果 R_i 和 R_j 的信賴度和支持度都相同，但 R_i 左邊條件比 R_j 左邊條件少。

- d. 比較 R_i 和 R_j 的類別，在訓練資料集中 R_i 的類別比較多。
- e. 比較 R_i 及 R_j 的使用，在分類訓練集資料時 R_i 會較先使用。
- f. R_i 或 R_j 隨機選， R_i 被選出。

3. 刪除衝突(conflicts)法則

對排序完成的分類關聯法則做檢查，以避免發生法則衝突的情況。衝突法則的刪除方法如下：

檢查每一條法則 $X \rightarrow C_i$ ，假如已經存在 $X \rightarrow C_j$ ，且 $X \rightarrow C_j$ 的優先順序比 $X \rightarrow C_i$ 來的小，則刪掉衝突的 $X \rightarrow C_j$ 。

4. 刪除多餘衝突的法則

對排序完成的分類關聯法則做檢查，以避免發生法則多餘且衝突的情況。多餘衝突法則的刪除方法如下：

檢查每一條法則 $XY \rightarrow C_j$ ，假如已經存在 $X \rightarrow C_i$ ，且 $X \rightarrow C_i$ 的優先順序比 $XY \rightarrow C_j$ 來的大，則刪掉多餘衝突的 $XY \rightarrow C_j$ 。

5. 檢查分類關聯法則的適用

利用修剪排序完成的分類關聯法則來分類訓練集，在分類的過程中有使用的法則將被註記，完全沒有使用的法則將被刪除，以減少分類關聯法則的數量。

6. 選定預設類別(Default Class)

為了防止無法分類的情况，設置預設類別。利用修剪排序完成的分類關聯法則來分類訓練集後，從無法被分類的訓練資料中選出類別數較多的當作預設類別。如果沒有無法被分類的訓練資料時，則從所有的訓練資料中選出資料筆數最多的類別當作預設類別，若資料筆數最多的類別有多個時則從中隨機選出。

3.4 分類測試階段

分類測試階段有測試分類及結果分析等2個步驟。

(1) 測試分類

利用分類關聯法則及其優先順序來測試分類測試資料集中的每一筆資料，並將結果儲存在分類結果資料庫中。分類結果主要包含每一筆資料分類的類別及平均每一筆資料執行的時間。

測試分類方法主要是根據分類關聯法則的優先順序來對測試資料做分類，優先順序越高的法則，會越先被使用。當測試資料的屬性滿足法則的所有條件時，則可以利用此法則將資料分類，若有任一個屬性不滿足時，則此法則無法適用，將以優先順序次高的法則來分類，當所有法則都無法適用時，則分類為預設類別

(2) 結果分析

統計分析測試分類結果，並顯示分類正確率及每筆資料平均所需時間。當測試分類完成後分類結果資料庫中已有相關資料，此步驟可以不用重複執行測試分類，即可觀看測試分類的結果。

3.5 分類預測階段

分類預測階段主要是利用前3階段訓練測試完成的資料分類模型(分類關聯法則及其優先順序)來預測欲分類的資料。欲分類的資料可以由使用者任意輸入，預測分類系統將利用本研究的資料分類模型來預測輸入資料所屬的類別，並即時顯示分類結果以供使用者參考。

4. 系統實驗與試用

4.1 系統實驗

本研究的實驗資料集是 UCI 機器學習庫網站 [12] 所提供 MONK's Problems 資料集中的 monks-1。系統實驗共有 2 項，分別是(1) monks-1 資料集分類測試及(2)不同參數設定對分類結果的影響。實驗的評估指標為資料分類的正確率和所需時間，正確率是正確分類資料筆數除以資料總筆數，所需時間是分類一筆資料所需的時間。

(1) 實驗 1：monks-1 資料集分類測試

[實驗目的]

評估利用正規化概念分析探勘的分類關聯法則，是否能有效分類 monks-1 資料集。

[實驗方法]

資料取樣 - 此資料集有提供訓練集及測試集。

最小支持度及最小信賴度設定 - 最小支持度設為 0.1，最小信賴度設為 0.5。

分類方法 - 分類關聯法則及其優先順序。

評估指標 - 分類正確率及分類所需時間。

[實驗結果]

實驗結果如表 2 所示，此資料集有 432 筆訓練資料、屬性有 6 個及類別有 2 個，測試資料有 124 筆。本研究探勘出的分類關聯法則有 5 條，分類正確率是 0.73、平均每筆所需時間為 0.04 秒。

(2) 實驗 2：不同參數設定對分類結果的影響

[實驗目的]

在不同參數設定下，評估利用正規化概念分析探勘的分類關聯法則，是否能更有效分類 monks-1 資料集。

[實驗方法]

資料取樣 - 此資料集有提供訓練集及測試集。

最小支持度及最小信賴度設定 - 先調整最小支持度到達最佳正確率後，再調整最小信賴度。

分類方法 - 分類關聯法則及其優先順序。

評估指標 - 分類正確率及分類所需時間。

[實驗結果]

圖 4-1 為調整最小支持度之實驗結果，最小支持度由 0.05 到 0.5 調整時，正確率由 0.73 下降到 0.5。圖 4-2 為調整最小信賴度之實驗結果，最小支持度 0.1、最小信賴度由 0.4 到 0.8 調整時，正確率由 0.73 上升到 1.0，可以達到相關研究[13]相同的正確率。

表 2 實驗 1：monks-1 資料集分類測試結果

欄位 資料集	Instances	Attributes	Class	Test cases	Rules	Correct	time	Nodes	Correct	time
monks-1	432	6	2	124	5	0.73	0.04	2881	1.0	0.94

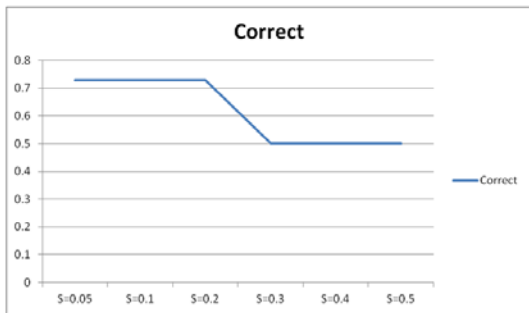


圖 4-1 調整最小支持度之實驗結果

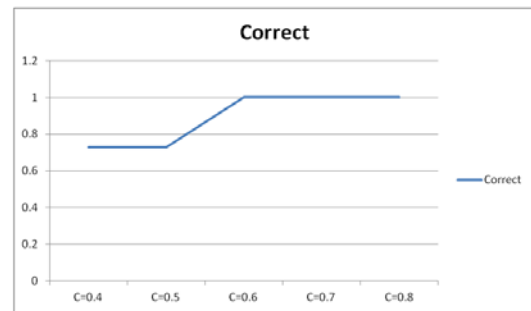


圖 4-2 調整最小信賴度之實驗結果

4.2 分類預測系統試用

本研究分類預測階段的支援系統是架設在網際網路的網頁伺服器 IIS 上，一般使用者只需利用瀏覽器即可以使用分類預測系統。

圖 5 是 monks-1 資料集分類預測畫面。在圖中，使用者首先選擇各屬性值(a1~a6)，然後按下分類按鈕。系統將顯示預測分類的結果(Class、適用規則與所需時間)。

分類預測系統

資料集: monks-1

a1: 1

a2: 1

a3: 2

a4: 2

a5: 3

a6: 2

分類 重選 回資料集選擇

Class: 0

適用規則: 3

所需時間: 0.01

訊息欄: 1.1.2.2.3.2 分類完成

圖 5 分類預測系統實例

5. 結論

本研究利用訓練資料集的正規化概念分析所產生的概念點陣，推導出分類關聯法則並修剪及排序。本研究的資料分類是以分類關聯法則為主，並利用其優先順序，判斷測試資料集及預測新進資料的類別屬性。

本研究以 monks-1 資料集為實驗對象，此資料集有提供訓練集及測試集。實驗結果顯示，最小支持度 0.1 及最小信賴度 0.5 的分類正確率為 0.73，

經由最小支持度及最小信賴度的調整正確率可以達到 1.0。另外，本研究也提供網頁版的分類預測系統，使用者可以很容易來使用分類預測系統。

初步的實驗結果顯示，利用正規化概念分析探勘的分類關聯法則確實可以應用於資料分類上。未來將測試其他的資料集找出應用上的一些問題，加以改善。

參考文獻

- [1] 魏巧宜、馬麗菁、李中芬，2012，運用資料探勘方法於憂鬱症傾向之研究，第二十三屆國際資訊管理學術研討會。
- [2] 曾憲雄等著，2006，資料探勘，旗標。
- [3] Liu, B., Hsu, W., and Ma, Y., 1998, "Integrating classification and association rule mining", In KDD'98, New York.
- [4] Li, W., Han, J., and Pei, J., 2001, "CMAR: accurate and efficient classification based on multiple class-association rules", ICDM 2001, IEEE Computer Society, San Jose, California.
- [5] Nguyen, L.T.T., Vo, B., Hong, T.P., and Thanh, H.C., 2012, "Classification based on association rules: A lattice-based approach", Expert Systems with Applications, Vol. 39, No. 13, p. 11357-11366.
- [6] Ganter, B. and Wille, R., 1998, "General Lattice Theory (2nd edition)", Birkhauser Verlag.
- [7] Wille, R., and Ganter, B., 1999, "Formal Concept Analysis: Mathematical Foundations", Springer.
- [8] Kim, M., 2003, "Document Management and Retrieval for Specialised Domains: An Evolutionary User-Based Approach", University of New South Wales, Ph.D. thesis.
- [9] Richards, D. C., 1998, "The Reuse in Ripple Down Rule Knowledge Based Systems", University of New South Wales, Doctorate thesis.
- [10] Tilley, T., 2004, "Tool Support for FCA", Springer Berlin/Heidelberg.
- [11] 吳慶陽，2010，基於概念點陣瀏覽搜尋中文資料之研究—以計算機概論試題為例，國立雲林科技大學，碩士論文。
- [12] UCI Machine Learning Repository 網站，<http://archive.ics.uci.edu/ml/>.
- [13] 陳博承，2012，正規化概念分析應用於資料分類之研究，國立雲林科技大學，碩士論文。