

模糊約略集合法於顧客與購買樣式之關聯分析

Fuzzy Rough Set Approach to Association Analysis of Customers and Purchasing Patterns

邱宏彬

Hung-Pin Chiu

張晉赫

Jin-He Chang

吳文盛

Wen-Sheng Wu

南華大學資訊管理學研究所

Department of Information Management, Nanhua University

摘 要

針對不同的顧客群給予不一樣的行銷策略，是提高企業競爭優勢的重要方法。在現今市場中，企業以商品組合方式銷售之行銷策略，已普遍運用。若成功將這些商品組合推薦給合適的目標顧客群，將可達到吸引顧客及提高銷售量之目的。關聯法則探勘以發現商品關聯性為主；少有文獻探討顧客屬性與購買樣式之間的關聯分析。在本研究中，我們先挖掘出顧客購買商品之間的關聯性，根據其購買樣式給予組合方式銷售。然後，透過改良後之模糊約略集合法找出顧客特徵與商品購買樣式之間的關聯性，作為目標行銷策略之依據。顧客數值屬性的模糊化，讓轉換過程保留較精確與完整的資訊，同時，本研究利用 α -cut 過濾雜訊，以便模糊約略集合法能挖掘出更有意義的法則。我們實作系統雛型，並且透過一些實驗來分析系統之可行性及有效性。

關鍵詞：顧客區隔、資料探勘、關聯分析、模糊約略集合理論。

Abstract

It is an important task that the enterprise must give different marketing strategies based on the different targeted customers to raise its competitive advantages. In the present market, the sale marketing strategy by merchandise combination is widely used by the enterprise. The enterprise can gain a lot of profit by recommending the relevant purchasing patterns to the targeted customers to promote their consumption desires. However, the previous researches on mining association rules only focus on discovering the relationships among items. Also, few researches have explored the association analysis between items and customer attribute values. In this paper, we first mine the frequent purchasing patterns from the transaction database. Then the association rules of customer characteristics and purchasing patterns are discovered through the Fuzzy Rough Set Approach. In addition, several modifications are proposed to mine the more useful rules. Finally, some experiments are conducted to evaluate the feasibility and effectiveness of the proposed system.

Keywords: Customer Segmentation、Data Mining、Association Analysis、Fuzzy Rough Set Theory

壹、緒論

在現今競爭激烈的市場，提供好產品或好服務只是顧客對企業的基本要求。由研究報告指出，若顧客保留率可提高5%，平均每位顧客的價值就能增加25%到100%；80%的利潤來自20%客戶(理查科克，1998)，由此可見保住既有客戶及提高其價值的重要性。傳統大量行銷之模式已無法滿足顧客之需求，而『個人化行銷』勢必成爲企業強化競爭優勢之主要行銷策略。近年來，商品以組合方式出售的行銷策略，已被普遍的運用於各種商品市場中(蔡玉娟等，2001)。企業若能根據不同的產品特色與消費者特性，擬定精確的

商品組合之行銷策略，而把這些商品組合成功推薦給適合的目標顧客群，不僅可達到吸引顧客及提高銷售量之目的，進而也可以提升企業之利潤與競爭力。

資料探勘技術可發掘隱藏於大量資料中有用的資訊，作爲決策之依據 (Han and Kamber, 2007)。約略集合法(Rough Set Approach, RSA)是從不完美的資料集中，發掘出條件屬性和決策屬性之關聯法則的有力方法；RSA在面對數值資料時，需要進行離散化(discretization)轉換，以利處理(Walczak and Massart, 1999)。在轉換過程中，通常運用區間切割的方式進行處理，然而以此方式，區間邊界的資料會被轉換到不同區間，此轉換誤差將會影響所

挖掘知識的準確性。模糊集合的平緩邊界觀念 (Zadeh, 1965), 是解決此種資料轉換缺失, 提高資料探勘法則精確性的自然方法。因此, 模糊約略集合法(Fuzzy Rough Set Approach, FRSA)因應而生, 不但保留了 RSA 可以有效地找出資料屬性關聯性的優點, 而且以模糊集合解決數值資料轉換的缺失(Hong *et al.*, 2000); 同時, 模糊語意的描述方式, 也提高了關聯法則的可讀性, 增加決策之方便性。

先前對於關聯法則探勘的研究, 以發現商品關聯性為主; 少有文獻探討顧客特性與商品購買樣式之間的關聯分析(Tsai and Chen, 2004)。我們可以先將顧客資料模糊化, 當作關聯分析之條件屬性。然後, 從交易資料庫中挖掘出購買商品的關聯性, 作為組合銷售的依據, 此即關聯分析之決策屬性。很自然地, 我們可以利用 FRSA 找出某種特質的顧客, 將會購買某種商品組合的模糊規則, 並以這些關聯法則, 進行商品組合行銷策略的擬定。

綜合上述, 為了有效的區隔目標顧客群, 以及針對目標顧客群之特性, 找出最合適的商品組合行銷策略, 本研究提出一個利用資料探勘技術進行顧客與購買樣式之關聯分析的系統模式, 方法如下:

(1) 購買樣式的挖掘: 利用 Apriori 演算法 (Agrawal and Srikant, 1994), 從交易資料庫中挖掘出購買商品的關聯性, 然後根據此購買樣式給予組合方式銷售。例如: 我們得到一法則為:

牛奶→麵包, 因此我們可根據該法則給予組合方式銷售: 買某種牛奶一瓶, 贈送某種麵包或者加幾元即可購得某種麵包。

- (2) 顧客資料的模糊化: 模糊邏輯的特點就是允許以接近自然語言的詞語(語意變數)來表達資料特性 (Yen and Reza, 1999)。將顧客資料的模糊化, 可以讓擷取出來的知識更易於了解, 並且讓資料在轉換過程中, 保持較完整的資訊。
- (3) 顧客與購買樣式之關聯分析: 本研究利用模糊約略集合法來找出顧客特徵與商品組合樣式之間的關聯性, 讓挖掘出的法則具有高可讀性, 並且解決資料轉換時所產生的問題。同時, 本研究利用 α -cut 克服資料模糊化所造成的資訊雜訊, 以便挖掘出的關聯法則更具有意義。另外, 我們透過系統實作與一些實驗來分析系統之可行性及有效性。

貳、文獻探討

一、資料探勘與資料庫行銷

資料庫記錄了大量的消費者購買行為和基本資料, 也因此引起一陣資料庫行銷 (Database marketing) 之風潮 (Ha and Park, 1998)。資料庫行銷是資料探勘(Han and Kamber, 2007)最常見的商業用途, 企業在現有的資料中, 找尋客戶喜好及習

性，依據發現的知識選擇目標客戶，並且設計不同產品及行銷組合策略，以滿足不同需求的消費者。

關聯法則主要是從大量的資料中，找出某些項目會引發其他項目出現的規則，也就是找出資料項目的相關性。其最典型的例子為購物籃分析，主要藉由找出顧客放置於購物籃內不同物品間之關聯性來分析顧客之購買行為，以作為市場行銷策略擬定之依據 (Changchien and Lu, 2001)。

Z. Pawlak 在 1982 年提出約略集合理論(Rough Set Theory)，用來從資料庫中發現分類規則。其主要概念為，將資料庫中的屬性分為條件屬性和決策屬性；資料庫中的所有物件，根據各個屬性不同之屬性值分成相應的子集，然後對條件屬性劃分的子集與決策屬性劃分的子集之間之上下近似關係生成判定規則。約略集合進行資料分析時具有許多之優點 (Pawlak, 2002；Pawlak, 1982)，例如，能夠找出最小資料集(data reduction)、產生決策規則、提供獲得結果最直接的解釋等。

二、模糊集合

Zadeh (1965) 提出模糊集合的數學概念，它是一種定量表達工具，用來表現某些無法明確定義的模糊性概念。透過歸屬函數我們可以對模糊集合進行量化來描述模糊集合的性質。若 X 為宇集合(universal set)，而 A 為一個模糊集合，則

對於每個 X 中的元素 x ，其屬於 A 的程度為 $\mu_A(x)$ ，此 $\mu_A(x)$ 稱之為歸屬函數(membership function)，其值介於 0 和 1 之間，其數學表示式如下：

$$\mu_A(x): X \rightarrow [0, 1] \quad (1)$$

而模糊集合 A 可表示為：

$$A = \{(x, \mu_A(x)) \mid x \in X\} \quad (2)$$

三角形與梯形歸屬函數是描述模糊集合最常用的函數 (Yen and Reza, 1999)。

三、模糊約略集合理論(Fuzzy Rough Sets Theory)

由於約略集合僅對類別型資料有處理的能力，因此，數值型資料通常透過區間切割的方式來轉換，換言之，資料數值落於某一區間，即以該區間之語意來表示 (Walczak and Massart, 1999)。此種轉換方式會產生資訊誤差的問題，所以模糊約略集合因應而生(Hong *et al.*, 2000)。以下為模糊約略集合理論之定義，表 1 為其符號解釋表 (Hong *et al.*, 2000)：

表 1 模糊約略集合法符號解釋表

符號解釋表	
U	所有物件的範圍
N	U 訓練樣本之總數
$Obj^{(i)}$	第 i 個訓練樣本， $1 \leq i \leq n$
A	U 之所有屬性集合
M	A 之所有屬性個數
B	包含於 A 之特定屬性子集
A_j	第 j 個屬性
$ A_j $	第 j 個屬性之語意範圍個數
R_{jk}	第 j 個屬性之第 k 個語意值
$v_j^{(i)}$	$Obj^{(i)}$ 之 A_j 的數值
$f_j^{(i)}$	$v_j^{(i)}$ 模糊化後之模糊集合
$f_{jk}^{(i)}$	$v_j^{(i)}$ 之歸屬度
c	等價類別之數目
X_l	第 l 個類別
$B_k(x)$	第 k 個模糊條件等價類別，其中 $1 \leq k \leq B(x) $
$ B(x) $	所有模糊條件等價類別之個數

(一) 模糊不可辨識關係 (Fuzzy indiscernibility relation)

當有兩個模糊物件，分別為 $Obj^{(i)}$ 和 $Obj^{(r)}$ 分別屬於 U ，其屬性為 $A_j \subseteq A$ ，並且有相同之語意值 R_{jk} ，其歸屬度分別為 $f_{jk}^{(i)}$ 和 $f_{jk}^{(r)}$ 並且大於 0，因此我們可定義其在屬性 A_j 中具有模糊不可辨識關係或模糊等價關係 (Fuzzy equivalence relation)，而在此關係中之歸屬度為 $\min(f_{jk}^{(i)} \cap f_{jk}^{(r)})$ 。

相同的，當 $B \subseteq A$ 時，在 B 屬性集中， $Obj^{(i)}$ 和 $Obj^{(r)}$ 兩個模糊物件分別屬於 U ，其屬性為 $B_j \subseteq B$ ，並且有相同之語意值 R_{jk} ，其歸屬度分別為 $f_{jk}^{(i)}$ 和 $f_{jk}^{(r)}$ 並且大於 0，因此我們可定義其在 B 屬性集中具有模糊不可辨識關係或模糊等價關係，在此關係中其歸屬度為 $\min(f_{jk}^{(i)} \cap f_{jk}^{(r)})$ 。

透過模糊等價關係我們可將 U 分割成多個模糊子集，其結果表示為 U/B 。

基於 B 且包含 $Obj^{(i)}$ 之分割集可表示為 $B(Obj^{(i)})$ 。

其 $B(Obj^{(i)}) = \{(B_1(Obj^{(i)}), \mu_{B_1}(Obj^{(i)})), \dots, (B_r(Obj^{(i)}), \mu_{B_r}(Obj^{(i)}))\}$ ，其中 r 為 $B(Obj^{(i)})$ 中分割之個數； $B_j(Obj^{(i)})$ 為在 $B(Obj^{(i)})$ 中之第 j 個分割集；並且 $\mu_{B_j}(Obj^{(i)})$ 為第 j 個分割集之歸屬度。

(二) 模糊下近似空間 (Fuzzy Lower Approximation)

令 X 為 U 的子集，即 $X \subset U$ ，且 $B \subseteq A$ ，則 X 在 B 中的模糊下近似空間定

義為：

$$B_k(X) = \{B_k(x), \mu_{B_k(x)} \mid x \in U, B_k(x) \subseteq X, 1 \leq k \leq B(x)\} \quad (3)$$

由於在模糊下近似空間中， $B_k(x) \subseteq X$ ，也就是說模糊等價類別 $B_k(x)$ 完全包含於 X 裡，所以其信賴度為 1.0。此外，在模糊約略集合中，其模糊等價類別之歸屬度，可作為模糊下近似空間所產生法則的效度。

(三) 模糊上近似空間 (Fuzzy Upper Approximation)

令 X 為 U 的子集，即 $X \subset U$ ，且 $B \subseteq A$ ，則 X 在 B 中的模糊上近似空間定義為：

$$B^*(X) = \{B_k(x), \mu_{B_k(x)} \mid x \in U, B_k(x) \cap X \neq \emptyset, 1 \leq k \leq B(x)\} \quad (4)$$

在模糊上近似空間中， $B_k(x) \cap X \neq \emptyset$ ，也就是說模糊等價類別 $B_k(x)$ 部分包含於 X 裡，其信賴度主要是運用元件之歸屬度加以計算，其公式如下：

$$p(B_k(x)) = \frac{\sum_{x \in (B_k(x) \cap X)} \mu_{B_k(x)}}{\sum_{x \in B_k(x)} \mu_{B_k(x)}} \quad (5)$$

參、系統架構與功能

為了有效協助企業進行市場的區隔化來辨認目標顧客群，並針對目標顧客群之特性，以選取最好的行銷模式，我們提出一套完整資料探勘系統，來因應企業之需求。本研究首先運用 Apriori 演算法挖掘出商品間的關聯性，並搭配模糊約略集合法來分析商品購買樣式與顧客特性的關聯法則，以提高資料探勘之精確度，進而探勘出商品與消費者的有用關聯知識。

一、系統架構及流程

本研究中之系統架構如圖 1 所示。

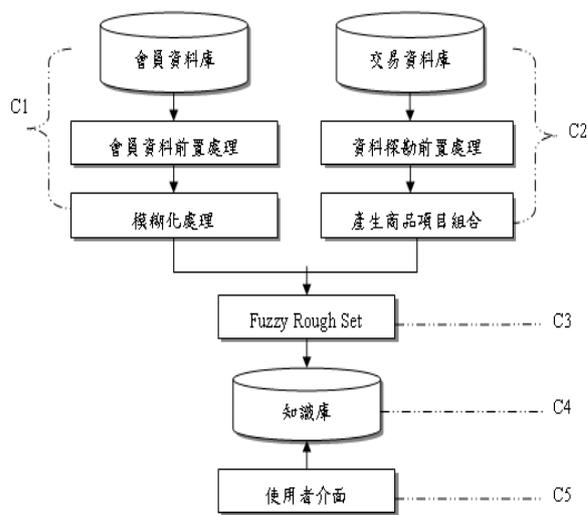


圖 1 資料探勘之系統架構圖

C1：會員資料模糊化流程

先將會員資料庫進行前置處理，篩選出想要挖掘之會員屬性，例如：年齡與收入。接下來，透過管理者或是專家訂定歸

屬函數，利用歸屬函數進行資料模糊化，例如：收入金額為 55000 元，轉換成收入為高，且歸屬度=0.9。

C2：產生商品組合之流程

將原本以交易時間編號(Time_id)為鍵值之交易資料庫，轉為以交易編號(TID)為鍵值之資料庫。接下來，利用 Apriori 演算法進行資料探勘，找出高頻項目組，以產生商品購買樣式。

C3：產生模糊約略法則

當會員資料處理完畢及商品項目組合產生完成後，必須找出其彼此間的關聯性，在本研究中運用模糊約略集合法進行資料的分類，以衍生出模糊約略法則。

C4：知識庫

透過模糊約略集合法所產生之法則，將所訓練出之法則儲存於知識庫中，以提供管理者決策查詢。

C5：使用者介面

在使用者介面中，除了可提供管理者查詢一般會員資料及商品資料外，亦可輸入預定促銷商品之項目，進而執行資料挖掘部分，以提供決策查詢。

二、改善模糊約略集合法的問題

在早期模糊約略集合理論之相關文獻中(王姿婷, 2000; Hong *et al.*, 2000)，雖然能夠有效解決明確集合之問題，並且提供信賴度及有效度兩個測量值來評估法則，但卻衍生出另一個影響資料探勘結果之因子。在進行模糊化過程中，若大多

數資料落在臨界值邊緣時，透過模糊約略集合理論去挖掘法則將會產生一些並非有用之法則。我們以例子說明之。假設資料探勘的目標為：「分析會員的年齡和工作經驗跟月收入之間的關聯性」。表 2 為一模糊化之資料庫：

表 2 模糊化之資料庫

Object	年齡 (A)	工作經驗 (J)	月收入 (M)
$Obj^{(1)}$	$(0.1/N+0.9/H)$	$(1/H)$	H
$Obj^{(2)}$	$(1/L)$	$(0.25/N+0.75/L)$	L
$Obj^{(3)}$	$(0.1/N+0.9/L)$	$(0.2/N+0.8/H)$	H

由該例子中，得知年齡為中之條件屬性完全包含於收入為高之決策屬性，因此我們可得到一下近似值規則如下：

IF 年齡 = 中 THEN 月收入 = 高
Confidence = 1.0；效度 = 0.1

然而我們可從資料庫中發現年齡為中之所有物件其歸屬度皆為 0.1，其值相當小，若用該資料進行推論，縱使其信賴度非常高，對於法則真正的效用度是沒意義的。因此，我們提出 α -cut 來克服該項問題。透過 α -cut 我們可將這些雜訊加以過濾，然後再根據過濾後的資料進行資料探勘。當我們將 α -cut 值設的越大，其所運算之資料可靠度越高，因此相對被篩選掉之資料越多。反之， α -cut 值設的越小，其所運算之資料可靠度越低，因此相對被篩選掉之資料越少。至於 α -cut 值要設多少，隨著資料不同而不同。以上例而言，

假設 $\alpha=0.3$ ，在過濾後之資料庫進行資料探勘，可發現”IF 年齡=中 THEN 月收入=高”該項法則已不存在，因此透過 α -cut 之觀念的確可有效挖掘出真正具有意義之法則。

三、系統功能與演算流程

在系統功能部分，主要可分為三大部分：會員資料之模糊化、商品項目組合之產生以及商品項目組合與會員型態關聯法則之分析。其詳細執行步驟如下所示：

(一) 會員資料之模糊化

Input：會員資料庫、每個欲轉換屬性之歸屬函數。

Output：模糊化會員資料庫。

步驟 1：會員資料庫之前置處理，先選擇目標之資料屬性，選完後再將一些有缺漏值、異常值等資料，進行淨化以完成此步驟。

步驟 2：管理者選定及輸入歸屬函數，將目標資料進行模糊化。

(二) 商品購買樣式之產生

Input：會員交易資料庫、最小支持度。

Output：商品項組合目。

步驟 1：交易資料庫之前置處理，將以交易時間編號為鍵值之交易資料庫，轉為以會員編號為鍵值之資料庫，方可進行資料探勘。

步驟 2：由管理者選定欲促銷之商品項目。

步驟 3：刪除會員交易資料庫中非促銷之商品項目，以整合出促銷目標資料表。

步驟 4：利用 Apriori 演算法，找出高頻項目組，進而產生商品購買樣式。

(三) 商品購買樣式與會員特性之關聯法則

Input：會員資料庫、會員交易資料庫、商品項目組合、 α -cut、信賴度。

Output：模糊約略法則。

步驟 1：法則分析之前置處理：

步驟 1.1：將高頻項目集之類別進行編號。

步驟 1.2：將會員購買商品組合對應至高頻項目集類別。

步驟 1.3：為了避免某一會員之特定消費習慣而影響整體所挖掘出之法則，因此必須將具有相同之會員編號以及高頻項目集類別的紀錄留下一筆，其餘刪除之。

步驟 1.4：合併會員個人資料與高頻項目集類別。

步驟 2：利用模糊約略集合理論產生法則：

步驟 2.1：設定 α -cut 值。

步驟 2.2：產生模糊決策等價類別 X_D 。

步驟 2.3：產生所有模糊條件等價類別 Y_C 。

步驟 2.4：設 $l=0$ ， l 為模糊決策等價類別正被處理之變數。

步驟 2.5： $l=l+1$ 。

步驟 2.6：設 $q=0$ ， q 為正被處理屬性個數之變數。

步驟 2.7： $q=q+1$ 。

步驟 2.8：計算 B 子集之所有 q 屬性與 X_l 之模糊下近似空間。

步驟 2.9：產生 B 子集之模糊下近似值規則

步驟 2.10：計算 B 子集之所有 q 屬性與 X_l 之模糊上近似空間。

步驟 2.11：產生 B 子集之模糊上近似值規則，並利用公式(5)計算其信賴度。

步驟 2.12：設定信賴度門檻值為 β ，當 $p(B_k(x)) \geq \beta$ 時，則產生 B 子集之模糊上近似值規則。

步驟 2.13：由於兩個或兩個以上之條件屬性仍與決策屬性存在著關聯性，因此我們必須進一步考慮兩個或兩個以上之屬性去產生關聯法則，稱

之為組合法則(Combinatorial rules)。

為了避免組合法則與單一屬性法則所產生之法則重複，例如：某一法則為 $A \rightarrow B$ ，而另一法則為 $A \cap C \rightarrow B$ ，當 A 已確定決定 B 時， A 且 C 亦能決定 B 的資訊已無意義了。因此，必須將訓練資料和類別 l 中移除模糊下近似值之物件。也就是，將 $B_k(x)$ 和 X_l 中之 $B'_k(x)$ 元素移除，其中 $B'_k(x)$ 是指完全包含於 X_l 之元素。

步驟 2.14：假如 $X_l = \emptyset$ ，則執行步驟 2.5，否則執行下一步驟。

步驟 2.15：產生模糊組合規則：重複步驟 2.7 ~ 2.14，直到 $q = m$ ，則執行步驟 2.16， m 為所有屬性之個數。

步驟 2.16：重複步驟 2.5 ~ 2.15，直到 $l = c$ ， c 為所有決策等價類別之個數。

步驟 2.17：產生所有法則。

步驟 3：將所產生之法則儲存至知識庫其完整執行流程如下圖 2 所示：

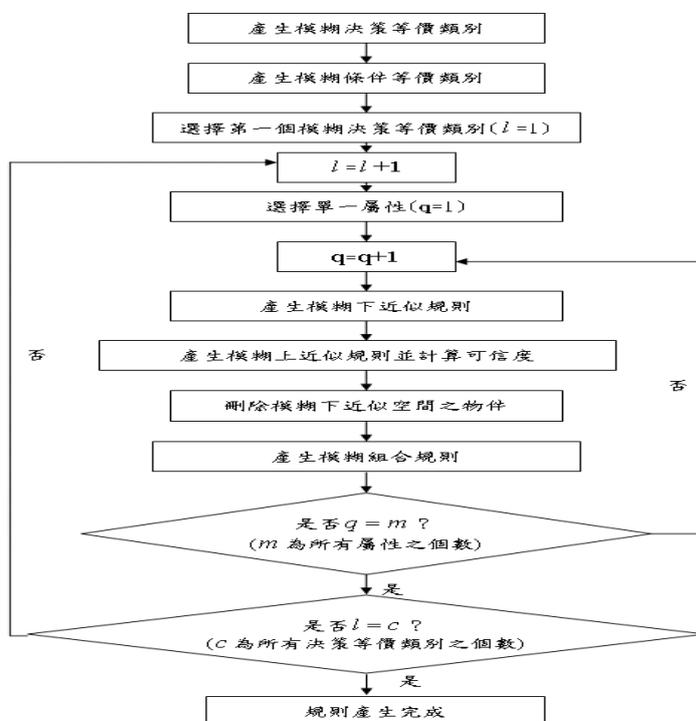


圖 2 模糊約略集合法執行流程圖

肆、系統範例說明

在本節中，利用一個簡單的範例來說明整個系統流程。資料探勘的目標定為：分析會員的『年齡』及『月收入』兩者與『商品項目組合』之間的關聯性。

一、會員資料之模糊化

表 3 為篩選出年齡及月收入兩屬性之會員資料庫。接著，使用圖 3 和圖 4 之梯形模糊歸屬函數，將會員資料庫轉換為模糊化會員資料庫，如表 4 所示。

表 3 會員資料庫

會員編號	年齡	月收入
M1	45	70000
M2	20	20000
M3	37	53000
M4	33	28000
M5	65	65000

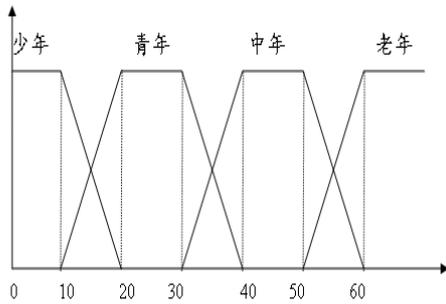


圖 3 會員年齡屬性欄位之歸屬函數

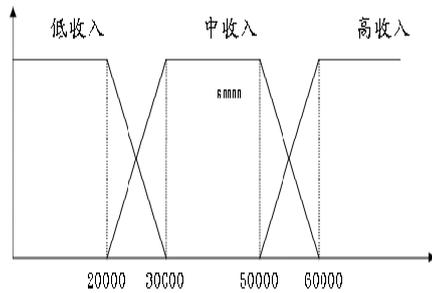


圖 4 會員月收入屬性欄位之歸屬函數

二、商品購買樣式之產生

將原本以交易時間編號為鍵值之交易資料庫，轉換為以交易編號為鍵值之資料庫。然後，由管理者選定欲促銷之商品項目。假設預定促銷之商品項目編號 PID 為： $\{001、002、003、004、005\}$ ，將會員交易資料庫中非促銷之商品項目給予刪除，得到表 5 之促銷目標資料表。

接著，利用 Apriori 演算法，找出高頻項目組，進而產生商品項目組合。最後在支持度為 2 所產生之高頻項目組為： $\{001,003\}$ 、 $\{002,003\}$ 、 $\{002,004\}$ 、

$\{002,005\}$ 、 $\{003,005\}$ 、 $\{002,003,005\}$ 。由於 $\{002,003,005\}$ 包含於 $\{002,003\}$ 、 $\{002,005\}$ 、 $\{003,005\}$ ，因此最後得到之最長高頻項目組為： $\{001, 003\}$ 、 $\{002, 004\}$ 、 $\{002, 003, 005\}$ ，分別以類別 T1、T2、T3 加以編號。

表 4 模糊化會員資料庫

會員編號	年齡	月收入
M1	(1.0/中年)	(1.0/高收入)
M2	(1.0/青年)	(1.0/低收入)
M3	(0.7/中年+0.3/青年)	(0.3/高收入+0.7/中收入)
M4	(0.3/中年+0.7/青年)	(0.8/中收入+0.2/低收入)
M5	(1.0/老年)	(1.0/高收入)

表 5 促銷目標資料表

TID	會員編號	購買商品項目
T001	M3	002,003
T002	M4	002,004,005
T003	M3	001,003
T004	M5	001,003
T005	M2	002,003,005
T006	M1	002,003,004
T007	M2	002,004
T008	M4	004
T011	M3	002,003,005
T013	M1	003,005
T015	M5	001,003,005
T016	M1	001
T017	M3	001

三、商品購買樣式與會員型態關聯法則之分析

(一) 法則分析之前置處理

- 1、將高頻項目集類別對應至會員編號。同時，將具有相同會員編號

以及高頻項目集類別的紀錄留下一筆，其餘刪除之。結果之資料表如表 6 所示。

ID	會員編號	年齡	月收入	高頻項目集類別
1	M4	(0.3/中年+0.7/青年)	(0.8/中收入+0.2/低收入)	T2
2	M3	(0.7/中年+0.3/青年)	(0.3/高收入+0.7/中收入)	T1
3	M5	(1.0/老年)	(1.0/高收入)	T1
4	M2	(1.0/青年)	(1.0/低收入)	T3
5	M1	(1.0/中年)	(1.0/高收入)	T2
6	M2	(1.0/青年)	(1.0/低收入)	T2
7	M3	(0.7/中年+0.3/青年)	(0.3/高收入+0.7/中收入)	T3

表 6 高頻項目集類別對應會員編號資料表

2、合併會員資料庫與高頻項目集類別換成如表 7。

TID	會員編號	高頻項目集類別
T002	M4	T2
T003	M3	T1
T004	M5	T1
T005	M2	T3
T006	M1	T2
T007	M2	T2
T011	M3	T3

表 7 合併會員資料庫與高頻項目集類別

(二) 利用模糊約略集合理論產生法則其資訊系統以數學式表示如下：

$$U = \{Obj^{(1)}, Obj^{(2)}, Obj^{(3)}, Obj^{(4)}, Obj^{(5)}, Obj^{(6)}, Obj^{(7)}\}$$

$$B = \{\text{年齡, 月收入, 高頻項目集類別}\}$$

$$V = \{\text{青年, 中年, 老年; 低收入, 中收入, 高收入; T1, T2, T3}\}$$

2.1：設定 α -cut 值為 0.3，過濾歸屬度低於 0.3 的雜訊。

2.2：產生模糊決策等價類別，如表 8。

2.3：產生模糊條件等價類別，如表 9。

表 8 模糊決策等價類別

l	高頻項目集類別	X_l
1	T1	$\{Obj^{(2)}, Obj^{(3)}\}$
2	T2	$\{Obj^{(1)}, Obj^{(5)}, Obj^{(6)}\}$
3	T3	$\{Obj^{(4)}, Obj^{(7)}\}$

表 9 模糊條件等價類別

A_{ij}		Y_{ij}
$i=A$	$j=Y$	$Y_{AY} = \{(Obj^{(1)}, Obj^{(2)}, Obj^{(4)}, Obj^{(6)}, Obj^{(7)}), 0.3\}$
$i=A$	$j=M$	$Y_{AM} = \{(Obj^{(1)}, Obj^{(2)}, Obj^{(5)}, Obj^{(7)}), 0.3\}$
$i=A$	$j=O$	$Y_{AO} = \{(Obj^{(3)}), 1.0\}$
$i=C$	$j=L$	$Y_{CL} = \{(Obj^{(4)}, Obj^{(6)}), 0.2\}$
$i=C$	$j=M$	$Y_{CM} = \{(Obj^{(1)}, Obj^{(2)}, Obj^{(7)}), 0.7\}$
$i=C$	$j=H$	$Y_{CH} = \{(Obj^{(2)}, Obj^{(3)}, Obj^{(5)}, Obj^{(7)}), 0.3\}$

註： i 為屬性， j 為語意值；其中屬性A表示年齡，其語意值Y表青年，M表中年，O表老年；屬性C表示月收入，其語意值L表低收入，M表中收入，H表高收入。其表示式如下：
 $Y_{ij} = \{(Obj^{(x)}, Obj^{(y)})\}$
 $\min(f_{jk}^{(i)} \cap f_{jk}^{(r)})$

2.4：設 $l=0$ ， l 為模糊決策等價類別正被處理之變數。

2.5： $l=l+1$ 。即 $X_1 = \{\text{Obj}^{(2)}, \text{Obj}^{(3)}\}$ 。

2.6：設 $q=0$ ， q 為正被處理屬性個數之變數。

2.7： $q=q+1$ 。即處理條件屬性為單一。

2.8：計算單一屬性 X_1 之模糊下近似規則，我們可發現此資料庫中，年齡為老之模糊條件等價類別 Y_{AO} 完全包含於 X_1 ，因此產生一個模糊下近似規則。

$$B_{*AO}(X_1) = Y_{AO} \cap X_1 = \{(\text{Obj}^{(3)}, 1.0)\}$$

2.9：產生 X_1 之模糊下近似值規則
Certain Rule：

IF 年齡 = 老年 THEN
T1 ; Confidence = 1.0，有效度 = 1.0。

2.10：計算單一屬性 X_1 之模糊上近似規則

從資料庫中可發現 Y_{AY} 、 Y_{AM} 、 Y_{CM} 、 Y_{CH} 四個條件等價類別部分包含於 X_1 。

2.11：產生 X_1 模糊上近似規則並計算可信度，其信賴度計算如下：

$$B_{*AY}(X_1) = \{(\text{Obj}^{(1)}, \text{Obj}^{(2)}, \text{Obj}^{(4)}, \text{Obj}^{(6)},$$

$\text{Obj}^{(7)}) , 0.3\}$ ，其 Confidence 等於

$$\frac{0.3(\text{Obj}^{(2)})}{0.7+0.3+1.0+1.0+0.3(\text{Obj}^{(1)}, \text{Obj}^{(2)}, \text{Obj}^{(4)}, \text{Obj}^{(6)}, \text{Obj}^{(7)})} = 0.09$$

而 $B_{*AM}(X_1)$ 、 $B_{*CM}(X_1)$ 、 $B_{*CH}(X_1)$ 的

Confidence 分別為 0.26、0.32 和 0.5。

2.12：設定信賴度 $\beta=0.3$ ，因此，我們可得到以下之模糊上近似規則；

Possibly Rules：

1. IF 收入 = 中收入 THEN
T1 ; Confidence = 0.32

2. IF 收入 = 高收入 THEN
T1 ; Confidence = 0.5

2.13：從訓練集 U 和決策類別 X_1 中刪除 $\text{Obj}^{(3)}$ ；因此 $X_1 = (\text{Obj}^{(2)})$ 。

2.14：由於 $X_1 \neq \emptyset$ ，因此執行下一步驟。

2.15：產生 X_1 模糊組合規則

產生年齡與收入跟高頻項目組 T1 之關聯法則，此時 $X_1 = (\text{Obj}^{(2)})$ 。

a. $q=q+1$ ，因此 $q=2$ 。

b. 產生 X_1 模糊下近似組合規則

$$B_{*A \cap C}(X_1) = \emptyset，因此，$$

並未得到任何模糊下近似組合法則。

c. 產生 X_1 模糊上近似組合規則並計算信賴度

從資料庫中發現 $Y_{AY,CM}$ 、 $Y_{AY,CH}$ 、 $Y_{AM,CM}$ 、 $Y_{AM,CH}$ 四個條件等價類別部分包含於

$$X_1 \circ B_{AY,CM}^*(X_1) \cdot B_{AY,CH}^*(X_1) \\ \cdot B_{AM,CH}^*(X_1) \cdot B_{AM,CM}^*(X_1)$$

之信賴度分別為 0.23、0.5、0.41、0.19。因此，我們可得到信賴度大於 $\beta=0.3$ 之模糊上近似組合規則：

Possibly Rules :

1. IF 年齡 = 青年 AND 收入 = 高收入 THEN T1 ; Confidence = 0.5
2. IF 年齡 = 中年 AND 收入 = 中收入 THEN T1 ; Confidence = 0.41

d. 由於 $m = 2$ 且 $q = 2$ ，因此結束模糊決策等價類別 X_1 。

2.16：重複步驟 2.4 ~ 2.13，直到 $l = 3$ 。

2.17：產生所有之法則

伍、系統實作與實驗結果分析

一、系統實作

本雛型系統從 Microsoft SQL Server 2000 中所附之 FoodMart 零售市場交易資料庫中獲取所需之資料表，其中會員資料表所包含之會員人數為 10,284 人，並且其交易資料表中發現，在 1997 年，計有 87,599 筆交易；在 1998 年，計有 164,558 筆交易。我們將會員資料進行前置處理，

轉換成可進行資料探勘之型態，如生日轉換成年齡，其系統介面如圖 5。接著，透過管理者或是專家訂定歸屬函數，進行顧客資料的模糊化，同時可設定 α -cut 值來過濾歸屬度雜訊，介面如圖 6 所示。



圖 5 會員資料前處理系統畫面

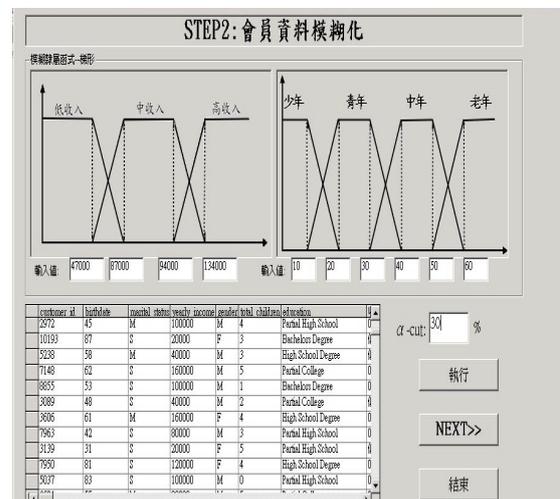


圖 6 會員資料模糊化系統畫面

管理者可輸入欲促銷之商品編號，如圖 7 所示。輸入完成後，接下來以 Apriori 演算法挖掘出商品之購買樣式，以做組合式行銷，如圖 8。之後，運用模糊約略集合法挖掘顧客特性與商品購買樣式之模糊約略法則，如圖 9 所示。



圖 7 欲促銷商品設定之系統畫面



圖 8 產生商品組合之系統畫面



圖 9 產生模糊約略法則之系統畫面

二、實驗結果分析

我們以 Microsoft SQL server 2000 的 FoodMart 資料庫內所包含之 sales_fact_1998 資料表，共有 164558 交易紀錄，來探討與比較模糊約略集合法與傳統約略集合法所產生法則之差異性，此外，我們亦將比較 α -cut 值對於模糊約略集合法所產生之法則的影響性。真實環境的大量分析資料並不容易取得，FoodMart 資料庫是由真實世界資料所建立的資料庫，因此，其實驗結果可望作為真實環境資料的分析參考。另外，我們僅針對法則數目的差異進行探討，未來，可進一步分析在獲利上的影響。

在進行實驗前，首先透過系統隨機選出 5 個商品項目，接下來透過 Apriori 演算法進行商品項目間關聯性之探勘。在實驗中將支持度設為 1000 筆，共可探勘出 8 個高頻項目組，因此，在以下實驗中之所用決策屬性共有 8 項。

(一) 模糊約略集合法與傳統約略集合法之法則比較

我們利用所設定之歸屬函數計算出會員收入及年齡之歸屬度，接下來分別利用本研究所撰寫之模糊約略集程式及傳統約略集程式進行實驗測試。實驗結

方法	信賴度	0	0.1	0.2	0.3	0.4	0.5
Fuzzy Rough set		472	280	66	60	32	8
Rough set		430	248	10	8	8	8

果如表 10 所示。

表 10 模糊約略集合與傳統約略集合法則比較表 (單位：筆)

由實驗中可發現，利用模糊約略集合理論所產生之法則數在信賴度為 0、0.1、0.2、0.3、0.4 都比約略集合理論之法則數多，由此可知利用模糊理論之觀念可將顧客屬性值完整的表達，這樣顧客之屬性值可能涵蓋一個以上之值，因此所挖掘出的法則也會較全面，所以其相對產生的法則數會較多。

(二) α -cut 值對於模糊約略集合法則之影響

在此實驗中將探討 α -cut 值對法則產生之影響性，是否可有效過濾不必要之法則。實驗結果如表 11 所示。由實驗中可發現當 α -cut 值設越大時，在整體上幾乎所產生之法則數目相對減少。此外，當 α -cut 值為 0.3 時，可發現信賴度大於 0.5 之法則數目皆為 0，由此可見在沒透過 α -cut 值處理的模糊約略集合理論所產生

幾筆信賴度大於 0.5 之法則，可能是有大多數資料落在臨界值邊緣，也就是其組成之資料多為異端值，這樣狀態下之法則實為無效之法則。例如，實驗中，由法則庫發現某一筆法則為：IF 收入=低收入 AND 年齡=少年 AND 性別=男性 THEN 商品項目組為 7，其信賴度為 0.94，當 α -cut 值為 0.2 時已不見該法則。因此，由本實驗中得知運用 α -cut 可有效的過濾不必要之法則，探勘出真正有意義之法則。

信賴度 α -cut 值	0	0.1	0.2	0.3	0.4	0.5
α -cut=0	472	280	66	60	32	8
α -cut=0.1	472	280	66	60	32	8
α -cut=0.2	454	290	56	49	25	7
α -cut=0.3	399	253	45	45	16	0
α -cut=0.4	376	230	40	34	12	0
α -cut=0.5	376	232	40	34	8	0

表 11 α -cut 值對法則數之比較 (單位：筆)

陸、結論與未來研究

資料庫行銷已成為企業有效探勘出消費者行為與制定行銷策略的工具。本研究中，我們運用資料探勘技術建構了一套決策支援系統，來輔助管理者擬定相關策略。以下為本研究的主要貢獻：

- (1) 找出顧客特徵與商品購買樣式間之關聯法則，並且克服資料轉換誤差的問題。

- (2) 克服在資料進行模糊化過程中，若多數資料落在臨界值邊緣時，可能會讓模糊約略集合法探勘出一些並非有用法則的問題。
- (3) 設計決策支援雛型系統，協助企業辨認目標顧客群，以找出最合適的行銷模式。

我們提出下列幾點未來可能的研究方向：

- (1) 本研究中並未利用約略集中屬性精簡 (Reduct) 之技術 (Walczak and Massart, 1999)，因此，在未來研究中，可先找出對決策屬性有影響之條件屬性，然後，再進行資料探勘方可更精簡找出顧客特性與其消費樣式。
- (2) 商品通常存在多階層 (Multi-Level) 之架構 (Han and Kamber, 2007)，例如：可樂可被分為飲料類，從多階層的角度去探討，結果可提供不同層面的策略。
- (3) 結合其他技術使本系統架構更完善，例如，利用 RFM 之觀念找出目標顧客群 (邱宏彬、蘇建源，2004)，再進行目錄發送以及其他行銷方式，減少行銷成本。
- (4) 取得真實環境的大量分析資料，進一步分析本方法在獲利上的影響，以驗證其有效性。

參考文獻

- [1] 王姿婷，「利用約略集論於數量形資料之知識擷取」，義守大學資訊工程研究所碩士論文，民國 89 年。
- [2] 謝綺蓉譯，「80/20 法則(理查.科克原著)」，大塊文化公司，民國 87 年。
- [3] 邱宏彬、蘇建源，「一個可彈性支援顧客關係管理與資料庫行銷之模糊 RFM Model」，電子商務學報，第 6 卷第 2 期，民國 93 年，頁 149-173。
- [4] 蔡玉娟、張簡雅文、黃彥文，「快速反向關聯法則與調整緊密規則—促銷商品之應用」，資訊管理學報，第 10 卷第 1 期，民國 90 年，頁 181-204。
- [5] R. Agrawal and R. Srikant (1994), "Fast algorithms for mining association rules", in *Proc. of the 20th Conf. on Very Large Databases*, 487-499.
- [6] S. W. Changchien and T.C. Lu (2001), "Mining association rules procedure to support on-line recommendation", *Expert Systems with Applications*, 20, 325-335.
- [7] J. Han and M. Kamber (2007), *Data Mining : Concepts and Techniques*, Morgan Kaufmann Publishers, second Edition.

- [8] S.H. Ha, and S.C. Park (1998), "Application of Data Mining Tools to Hotel Data Mart on the Intranet for Database Marketing", *Expert Systems With Applications*, 15, 1-31.
- [9] T.P. Hong, T.T. Wang, S.L. Wang and B.C. Chien (2000), "Knowledge Learning a coverage set of maximally general fuzzy rules by rough sets", *Expert Systems with Application*, 19, 97-103.
- [10] Z. Pawlak (1982), "Rough Sets", *International Journal of Information and Computer Sciences*, 11, 341-356.
- [11] Z. Pawlak (2002), "Rough Sets and Intelligent Data Analysis," *Information Sciences*, 1-12.
- [12] P.S.M. Tsai and C.M. Chen (2004), "Mining interesting association rules from customer databases and transaction databases", *Information Systems*, 29, 685-696.
- [13] B. Walczak and D.L. Massart (1999), "Tutorial: Rough Sets Theory," *Chemometrics and Intelligent Laboratory Systems* , 47, 1-16.
- [14] J. Yen and L. Reza (1999), *Fuzzy Logic: Intelligence, Control, and Information*, Prentice-Hall, Inc.
- [15] L.A. Zadeh (1965), "Fuzzy Sets" , *Information Control*, 8, 338-353.