

南 華 大 學

資訊管理學研究所

碩士論文

從顧客 RFM 資料及交易資料庫挖掘資料庫行銷之模糊
關聯法則

Mining fuzzy association rules from RFM data and
transaction data for database marketing

研 究 生：施 柏 魁

指 導 教 授：邱 宏 彬

中 華 民 國 九 十 五 年 六 月

南 華 大 學

資訊管理研究所

碩 士 學 位 論 文

從顧客 RFM 及交易資料庫挖掘資料庫行銷之模糊關聯法則

研究生：施柏魁

經考試合格特此證明

口試委員：陳清華
周志賢

指導教授：邱宏村

系主任(所長)： 吳光閔

口試日期：中華民國 95 年 6 月 3 日

誌 謝

二年資管所在職班的日子說長不長，說短不短！如果從先修班就開始算，很多同學：富一、順當、朝原、雅如、就是三年的同窗了。還有 204 的小同學：阿聰、靜瑜、婉容，以及 advisor 邱宏彬教授，都豐富我這一段生命的內容。

教書教 17 年了，工作、生活早已定型。早上起床忙著搭校車，進辦公室忙著公事與教學。但技職院校的學生本來就對數學比較恐懼又不用功，工作、教學上已很少有讓自己覺得有成就感的事物了！每個假日的進修雖然不輕鬆，但還是能聽老師上課、與同學切磋聊天，就覺得很值得。尤其能當邱老師的學生實在是一種福氣！

課堂上常聽到老師感嘆世俗道德日趨低下，要大家從本身做起，做好本份的工作外，更要熱心、積極地助人，即所謂的「利他為上」。老師除了說以外，更自己身體力行。在我撰寫論文過程中的指導，對簡報內容的指正。老師的循循善誘，不吝付出的態度，深深感動到一樣在大專院校教書的我。除了在學問上有所收穫外，在做人處事更有所啟發。

最後，還是一聲「老師，謝謝你！」。

施柏魁 謹誌

中華民國九十五年六月

從顧客 RFM 資料及交易資料庫挖掘資料庫行銷之模糊關聯法則

學生：施柏魁

指導教授：邱宏彬

南 華 大 學 資 訊 管 理 學 系 碩 士 班

摘 要

本研究討論資料探勘上兩個重要議題的關聯法則。一是交易資料庫的模糊關聯法則，可提供決策者制定生產決策時的依據；一是 RFM 資料間的模糊關聯法則，作為行銷策略的參考。

我們提出叢聚式關聯法則的架構整合這二個問題。首先是藉建立叢聚式模糊集表，找出模糊大項目組。可減少掃描資料庫的次數，增進執行效率。再者，以使用者自定的最小信賴度，建立交易資料庫的模糊大項目組與 RFM 資料客戶群的模糊大項目組間的關係，將之轉換成形如 R.L&F.H&M.M B.H 的關聯法則。

關鍵詞：資料探勘、模糊關聯法則、叢聚式模糊集表

A Fuzzy Data Mining Approach for Mining Fuzzy Association

Rules in Two Databases

Student : Yi-Tsung Tang

Advisor : Dr. Hung-Pin Chiu

Department of Information Management

The M.B.A. Program

Nan-Hua University

ABSTRACT

In this paper, two important issues of mining association rules are investigated. The first problem is the discovery of generalized fuzzy association rules in the transaction database. It's an important data-mining task, because more general and qualitative knowledge can be uncovered for decision making. However, few algorithms have been proposed in the literature. Moreover, the efficiency of these algorithms needs to be improved to handle real-world large datasets. The second problem is to discover association rules from the RFM data and the large itemsets identified in the transaction database. This kind of rules will be useful for marketing decision.

A cluster-based mining architecture is proposed to address the two problems. At first, an efficient fuzzy association rule miner, based on cluster-based fuzzy-sets tables, is presented to identify all the large fuzzy itemsets. This method requires less contrast to generate large itemsets. Next, a fuzzy rule discovery method is used to compute the confidence values for discovering the relationships between transaction database and RFM database. An illustrated example is given to demonstrate the effectiveness of the proposed methods .

Keywords: Data Mining, Fuzzy Association Rule, Cluster

目 錄

書名頁	i
博碩士論文授權書	ii
論文指導教授推薦書	iii
論文口試合格證明	iv
誌謝	v
中文摘要	vi
英文摘要	vii
目錄	viii
表目錄	ix
圖目錄	x
第一章 緒論	1
第二章 文獻探討	5
第一節 資料探勘	5
第二節 關聯法則挖掘法之相關研究	11
第三節 模糊邏輯與模糊切割	24
第三章 研究方法	32
第一節 研究架構	32
第二節 模糊 RFM Model	33
第三節 應用在交易資料庫的模糊叢聚式關聯法則	39
第四節 應用在 RFM 及交易資料庫間的模糊叢聚式關聯法則的挖掘	41
第四章 實例演練	43
第五章 結論與未來努力方向	64
參考書目	66

表 目 錄

表 2-1：資料探勘的演化步驟	5
表 2-2：食物項目分類以 Taxonomy 為例	19
表 2-3：多階層關聯法則演算法比較表	24
表 2-4：範例資料表	30
表 3-1：顧客分群表	36
表 4-1：顧客 RFM 資料表	43
表 4-2：顧客 RFM 資料表對應的模糊值表	44
表 4-3：1-itemsets RFM 資料方塊	47
表 4-4：2-itemsets RFM 資料方塊	48
表 4-5：3-itemsets RFM 資料方塊	48
表 4-6：RFM 高頻 1-itemsets (L_1)	49
表 4-7： C_2 的 FS	49
表 4-8： L_2 的 FS	50
表 4-9：RFM 關聯法則表	50
表 4-10：RFM 關聯法則表 (符合最小信賴度)	51
表 4-11：顧客交易資料表	52
表 4-12：叢聚式模糊集表	54
表 4-13：資料方塊對應的顧客表	58
表 4-14：大項目集及其對應的顧客群	59
表 4-15： L_{RFM_i} 及 L_j 所組成的關係矩陣	60
表 4-16：模糊 RFM 大項目與模糊大項目的關聯法則表	61

圖 目 錄

圖 2-1:知識挖掘過程示意圖	8
圖 2-2:產生大項目組之範例	13
圖 2-3:食物項目分類以 Taxonomy 為例	18
圖 2-4 : ML_T2L1 演算法	20
圖 2-5 : Cumulate 演算法	22
圖 2-6: 三角形歸屬度函數	27
圖 2-7: 以三角形歸屬度函數建立的 Simple fuzzy partition	28
圖 3-1 叢聚式模糊關聯法則架構圖	33
圖 3-2:模糊 RFM Model 流程架構圖	34
圖 3-3 :RFM 3 維資料方塊	35
圖 3-3 : CBFAR 挖掘演算法則	40
圖 4-1:R、F、M 歸屬函數	44
圖 4-2 : 商品項目分類樹	53
圖 4-3 : 商品項目的歸屬函數	53

第一章 緒論

在競爭激烈的環境下，如何創造其競爭優勢並提供優良的產品或最佳的服務來滿足顧客的需求，是現今許多企業所關注的議題。在許多研究報告中指出，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%，以及 80%的利潤來自 20%客戶【許智豪, 2000】，由此可見，保住既有客戶及提高其價值的重要性。因此利用資訊科技來強化與客戶之間的互動，並與客戶建立緊密且長久的關係，讓既有的顧客持續地在此企業進行消費行為將是企業創造利潤的重要關鍵，也因此顧客關係管理就成為當今企業最熱門的議題。

在顧客關係管理中，瞭解顧客的特性與消費行為是相當重要的工作。而瞭解顧客的特性與消費行為，最直接的方法就是從企業的交易資料庫中找出相關的資訊或知識。但由於交易資料庫的資料量快速地增加，使得現今資料庫相當的龐大。如何從龐大的資料中，找出它所隱含的知識並且加以利用，已成為一個企業所關心的議題。為了有效解決大量資料所產生的問題，資料探勘(Data Mining)技術已成為近年來愈受重視的焦點【Goebel and Gruenwald, 1999】並且藉由資料探勘技術可以為企業經營者快速地找到他(她)們所需要的資訊，進而協助他(她)們進行決策與分析【Agrawal, Srikant and Vu, 1997】。

在許多有關資料探勘技術的研究中，就以關聯法則最為著名。而在關聯法則演算法的研究中，如何從資料中找尋項目的相關性，成了大家所關注議題。從交易資料庫探討商品項目間的關係，找出哪些商品項目會被同時購買？業者可以更有效地預測商品需求量，決定進貨量或庫存

量；或是在店裡要如何鋪陳貨品，或用來評估店裡的促銷活動成效。【曾憲雄, 2004】例如：如果顧客買了筆記型電腦，則這個顧客同時購買隨身碟的機率是 80%，再例如：如果一個顧客買了全麥麵包以及低脂優酪乳，那麼這個顧客同時也買低脂牛奶的機率是 85%。

Agrawal and Srikant 曾提出著名的 Apriori 關聯法則演算法【Agrawal and Srikant, 1994】，重複掃描交易資料庫後，找出 1-itemset 的候選項目組後，以使用者制定的門檻值（最小支持度）產生 1-itemset 的大項目，再運用結合、刪除二個運算，逐層地產生 2-itemset、3-itemset... 的候選項目及大項目組。最後藉著不同的階層及使用者制定的門檻值（最小信賴度）產生關聯法則。為解決因重複掃描交易資料庫導致執行效率低落的缺點。有很多學者提出改進的方法，包括 DHP 演算法【Park et al., 1995】，Sampling 演算法【Toivnen, 1996】以及叢聚式關聯法則（cluster-based association rule）【Tsay and Chiang, 2005】。

在過去相當多的關聯法則，其支持度的計算方式是依據項目(組)在交易資料總筆數中其出現筆數所佔的比例，並以比例是否達到使用者所制定的門檻值來產生大項目組。若從對企業貢獻的角度來分析，這大項目組在交易資料庫中，其所出現的次數相當的高，因此這大項目組對企業貢獻而言是相當卓越的。然而這個說法並不盡然如此。原因是，有可能有的項目出現次數並不多但被購買量或是此商品項目的金額相當大。對於這些項目而言，其對企業的貢獻也是我們不可以忽視的。因此考慮項目所購買的數量或是購買的金額來作支持度計算是有其必要性的。故我們研究 RFM 屬性間的關係，探討 R、F、M 間的關聯性，除了能夠讓行

銷決策者更清楚了解顧客的行為外，也能更深入地從產品在市場的特性以及市場狀況進行較為全面的分析，以使企業能夠更有效地達成行銷決策。分析顧客過去的購買行為以判斷哪些顧客值得進一步的接觸式發展行銷策略上重要的議題。透過 RFM 的分析可以量化顧客消費行為並且衡量顧客忠誠度和貢獻度，以利顧客分群及目標客戶的鎖定。

近年來，模糊理論【Zadeh，1965】因它的簡易性及較符合人類的思考模式，已被廣泛地應用在人工智慧的研究工作上【Kandel，1992】。洪宗貝提出一個模糊關聯法則【Hong et al，1999】，處理單階層的數值資料。但是在實際的應用上，商品項目常被階層化組織起來因此挖掘多階層資料庫可得到很多重要的知識。【Hong et al，2003】

由於企業本身的資源有限，再加上顧客需求的多樣性。因此企業必須能夠確認不同客戶的目標市場，針對不同的客戶群提供客製化行銷策略。本研究從顧客交易資料庫之 RFM 資料與商品銷售紀錄資料庫萃取出此二資料庫的關聯法則，研究顧客消費行為與購物偏好，分析顧客過去的購買行為是屬於 RFM 分群中的哪一群顧客，即 RFM 客戶分群；及其會購買商品或商品種類，可以讓企業行銷部門在公司的預算內做最有效的廣告、促銷行為，以最低的行銷成本達到最好的廣告效益。如在 3C 大賣場的顧客交易資料庫，以一年的顧客 RFM 資料表與商品銷售紀錄，萃取出型如”若’ R.L 且 F.M 且 MH’ ，則’ 購買物品 D’ (conf=c)” 的關聯法則。如果用在廣告行銷的預算減少，我們可以調高最小支持度，使屬於’ R.L 且 F.M 且 MH’ 這個顧客群的人數變少；反之亦然。

本文其餘的章節架構如下：第二章相關文獻探討：介紹資料探勘、關聯法則、模糊理論模糊切割的基本概念，第三章研究方法：模糊 RFM Model 與模糊叢聚式挖掘法則，第四章實例演練及第五章結論及未來努力方向。

第二章 文獻探討

第一節 資料探勘

一、資料探勘的崛起

在 80 年代，由於資料庫技術的發達，許多企業積極地投入大量資源來建置資訊系統。這也使得企業所保存的資料變得龐大且複雜，但由於傳統統計侷限在小樣本的問題，使得傳統統計無法處理大量且複雜的資料。因此為了能獲得潛在資訊來幫助企業解決問題，遂發展出獨立之新領域－資料探勘。【M. S. Chen et al., 1996】整理得資料探勘的演化步驟如表 2-1：

表 2-1：資料探勘的演化步驟【M. S. Chen et al., 1996】

時間軸	→			
	1960s	1980s	1990s	迄今
演進	資料收集	資料處理	資料倉儲、決策技援	資料挖掘
企業問題	過去五年中公司的總收益多少？	在加州的分部去年三月的銷售額是多少？	在加州的分部去年三月的銷售額是多？舊金山據此可得出什麼結論？	下個月舊金山的銷售會如何？為什麼？
促成技術	電腦、磁帶、磁片	關聯式資料庫、SQL、ODBC	OLAP、資料倉儲、多維式資料庫	新演算法、多處理器電腦、大型資料庫
特性	提供歷史性的，靜態的資料	以記錄來提供歷史性的、動態資料	在各種層次上提供回溯的、動態的資料	提供預測性的資訊

二、資料探勘的定義與流程

資料探勘是近年來隨著人工智慧和資料庫技術發展的一門新興技術。它可從大量的資料中，萃取出隱含、過去不為人所知且可信與有效的知識。也可以說是依照使用者所設定的參數，在一群未經處理的資料中找到使用者感興趣的資訊，經過某些特殊的處理後，作為使用者決策判斷的參考依據。以下是常被引用的定義：

(一)【Fayyad, 1996】的定義則嚴格區分資料探勘與資料庫中之知識發掘(KDD)。其定義資料庫中知識發現為自資料中選取合適資料，進行資料處理、轉換、資料探勘至結果評估之一系列過程。而資料探勘為其中一步驟(圖 2-1)。

(二)【Berry, 29】則認為資料探勘是為挖掘有意義的特徵或法則，而必須從大量資料之中以自動或是半自動的方式來探索與分析資料。

(三)【Kleissner, 1998】認為資料探勘是一種新的且不斷循環的決策支援分析過程，它能夠從資料中，發現出隱藏價值的知識，以提供給企業專業人員參考。

在本研究，我們還是遵照 Fayyad *et al.* 的定義，將知識發掘與資料探勘分開看待，以免觀念上有所混亂。而知識挖掘可分為三個主要步驟，分別為前置階段、資料探勘階段、規則詮釋階段，而依照此三個階段又可以將知識挖掘的過程又分為以下四項：

(一) 資料的取得與淨化階段：

在進行資料探勘前，使用者必須確認需對何種領域(domain)的資料進行資料探勘的動作，例如零售業在資料探勘之前，必須確認挖掘的是客戶資料或是交易資料；醫院在對其內部資料庫進行挖掘時，也必需確認所要挖掘的資料是病人的病歷資料，或是各種疾病之間的相關性資料等。確認資料探勘時所要挖掘的資料來源後，才能有效的將所需要的資訊探勘出來。

在取得資料的來源後，亦必須確認資料本身的資料型態以及資料是否為有效的資料。例如日期型態的資料欄位內存放的是日期格式，年齡型態的欄位中存放的是數字，且數字的範圍是在年齡中是有效合理的範圍。此一過程即為資料的淨化階段。

一般來說，以上的兩個階段是資料探勘的前置工作，將資料取得與淨化之後，就可以交由電腦系統進行所需的挖掘動作，但此兩個階段所需要花費的人力與時間卻是整個資料探勘過程中最多的部分。

(二) 資料倉儲階段：

將資料取得與淨化後，必須選擇一個良好的資料庫系統，作為資料存放的位置。在此必須要考慮的是企業成本考量，資料庫的大小及效能，以及資料庫相關應用工具の種類與實用性。對一般大型企業來說，可以選擇大型的資料庫工具搭配中型的資料探勘的應用程式，建構一個完整的資料探勘的環境。而一般中小企業的使用者也可以利用各種資料庫系統作為資料倉儲的工具，自行開發資料探勘的應用程式，以達到資料探勘的目的。

(三) 知識擷取階段：

在此一階段中，首先確定資料倉儲中哪些欄位的資料是必要的。先將這些主要的欄位提出，再配合利用各種資料探勘的演算法，將使用者想要挖掘的資料從資料倉儲中取出並執行演算法後，並將相關的資訊從資料庫中擷取出來，成為有用的知識。此一階段為資料探勘階段，也是各種演算法在改進效能時所著重的階段。

(四) 規則詮釋階段：

以上的幾個階段中，由電腦程式中擷取出來的規則或項目組合，或許僅是幾個項目的關聯或是一些代碼，對於非相關領域的人看來或許只是幾組沒有意義的代碼或名詞。資料探勘至此必須經過此一領域相關的學者加以詮釋，並確認資訊的可用性，將新奇有用且讓使用者感興趣的資訊取出，用一般大眾都能了解的名詞解釋出來。此一階段就是規則詮釋的階段，而經過了一般化的資訊或是規則，就是一般所說的知識或是規則。

從資料庫系統中挖掘出有用的知識或是規則的整體過程，就是知識挖掘範疇了，整個資料挖掘的過程示意圖 2-1 如下所示：

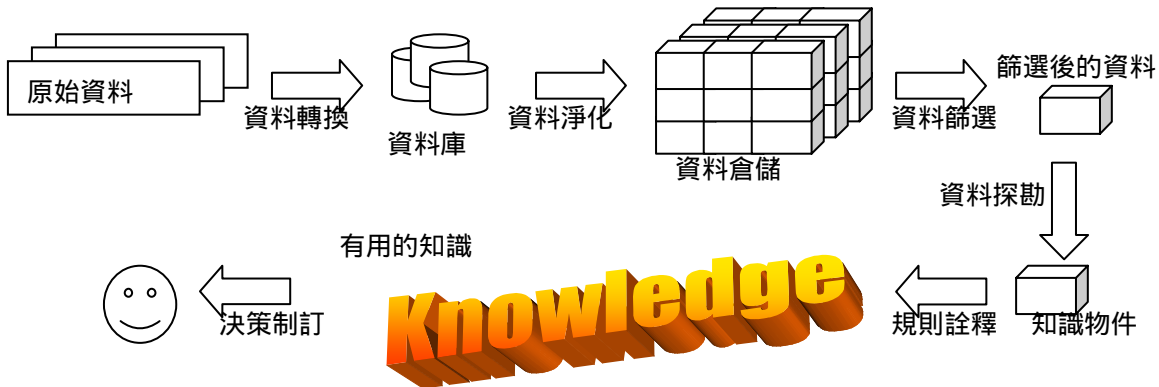


圖 2-1：知識挖掘過程示意圖

三、資料探勘模式

為瞭解決不同類型的領域中資料探勘的問題，資料探勘區分為許多不同的方法，例如關聯法則 (Association Rule)、資料分類 (Data

Classification)、資料分群(Data Clustering)、路徑尋訪樣式(Path Traversal Pattern)【M. S. Chen et al., 1996】……等不同的資料探勘模式。詳細介紹分別如下：

(一) 關聯法則：

由資料庫交易中找出項目(items)之間的關聯性，常使用的參數為支持度(support)和信賴度(confidence)，來評量一條關聯法則的發生頻率和強度。例如某大賣場發現購買麵包的顧客同時會購買牛奶，則該大賣場在行銷規劃可將牛奶與麵包放置同一架上，以增加連帶銷售目的。而關聯法則依層次來區分，可分為單層次關聯法則與多階層關聯法則；若把交易時間順序的因素來進行關聯法則的考量，則為稱為序列型樣【M. S. Chen, J. Han, and P. S. Yu, 1996】。關聯法則模式最常見的演算法如 Apriori 演算法【R. Agrawal and R. Srikant, 1995】、Sampling 演算法【H. Toivonen, 1996】等演算法。

(二) 資料分類：

根據一些資料的屬性來進行計算，從歷史資料中進行特徵及規則擷取，根據這些特徵來建立模式，最後針對其他未分類或是新的資料進行預測。例如，從以往信用卡核卡歷史資料中找出核準與否的特徵，建立分類模式，此分類模式便可依據新的客戶資料(年齡、職業、收入、教育程度、婚姻狀況…)推論是否核準此新客戶的開卡申請。其模式最常見的技術如決策樹……等。

(三) 資料分群：

是將資料分群，目的是找出同群集中資料的相似性，及各群集之間的差異性，使得同群中資料相似度最大，而各群之資料差異度最大。例如，銷售業者將客戶依其年齡、收入、居住地點、興趣…等的屬性進行分群，這樣市場區隔能讓行銷人員了解最適合行銷的客戶群，並提供最合適的產品及服務適當的顧客。其模式最常見的技術如 K-mean、Fuzzy C-mean、SOM……等。

(四) 路徑尋訪樣式：

在網際網路的環境中，此模式主要是擷取使用者瀏覽路徑存取特徵來瞭解使用者瀏覽網站之行為進而訂定出適合該使用者之個人化行銷環境。大部分的方法都是將網站使用的記錄檔轉換成路徑順序的格式，然後判斷特徵出現之頻率。

本研究在資料探勘模式的部分，主要是探討關聯法則並運用模糊邏輯來進行顧客關係管理之研究與應用。接下來第二節與第三節將分別介紹關聯法則以及模糊理論。

第二節 關聯法則挖掘法之相關研究

前一節，我們介紹了資料探勘的演化步驟、定義與流程以及模式。本節我們將針對資料探勘中最重要關聯法則作為探討對象，主要分為單層次關聯法則以及多階層關聯法則來進行探討。

一、單層次關聯法則介紹

(一) 關聯法則定義與相關名詞介紹：

關聯法則的描述如下：令 $I = \{i_1, i_2, \dots, i_m\}$ 是所有相異物品項目 (Items) 所成的集合， D 是所有交易記錄 (Transaction) T 的集合， T 是在 I 中任意物品項目的子集合。一個集合 $X \subseteq I$ 稱為項目組 (Itemset)，此項目組所包含的物品項目之個數稱為此項目組的長度，若其長度為 k ，則稱此項目組為 k -項目組 (k -itemset)。一筆交易記錄若可支持一個項目組 X ，則此交易記錄必包含此項目組的所有物品項目， $X \subseteq T$ ，記為 $\text{support}(X)$ 。

關聯法則的表現是相當明確且易懂的，一個關聯法則會包含兩個項目—前項 (antecedent) 和後項 (consequent)，它的表現形式可以定義為 $X \Rightarrow Y$ ， $X, Y \subseteq I$ 且 $X \cap Y = \phi$ ，其中 X 和 Y 分別表示在資料庫中不同之資料項目組【Lakshmanan, and Pang, 1998】。關聯法則靠支持度 (support) 及可信度 (confidence) 兩個參數來判斷此關聯法則是否有意義；支持度為資料庫中包含 $X \cup Y$ 的交易記錄所佔百分比，記為 $\text{support}(X \cup Y)$ ；而可信度則是定義此關聯法則可信的程度，也就是 X 出現的條件下， Y 也會跟著出現的條件機率，記為 $\text{support}(X \cup Y) / \text{support}(X)$ 。一個有效的關聯法則，其支持度及可信度必須要大於或等於使用者所定之最小門檻值 (Threshold)，只有滿足此條件限制的關聯法則，我們才認為其是有意義的。例如：購買麵包的會同時購買牛奶的顧客，它的關聯法則如下：

麵包 \Rightarrow 牛奶 [support = 40%, confidence = 50%]

一個相關項目組剛開始產生時，由於仍未開始搜尋資料庫以計算其支持度，我們並不知道此項目組是否大於或等於使用者所定之最小限制，此時的項目組我們稱之為候選項目組 (Candidate itemset)；而後經由搜尋資料庫以計算其支持度，若其滿足使用者

所定之最小限制，我們便稱此候選項目組為大項目組(Frequent itemset or Large itemset)，而我們在挖掘關聯法則的過程中，大項目組的搜尋是最重要的一環。

(二) 單層次關聯法則演算法介紹—Apriori 演算法：

Apriori 演算法首先於 1994 年由 Agrawal et al. 提出【R. Agrawal and R. Srikant,1994】。在大型資料庫中的銷售交易資料如何發現關聯法則中，Apriori 演算法是最具代表性的演算法，茲將其演算方式扼要說明如下：

Apriori 演算法中包含了兩個重要的步驟：

(1)反覆的產生候選項目組和搜尋整個資料庫，直到找出所有的大項目組。

(2)利用(1)所找出的大項目組，推導出所有的相關法則。

在步驟(1)產生大項目組的過程中，由掃描資料庫產生單一項目組(1-Itemset)開始逐層產生相關項目組(k-Itemset, $k>1$)。此過程主要分為兩個階段：第一個階段為產生新的項目組，若相關項目的長度為 k ，則稱為候選 k -項目組(candidate k -itemset)，記為 C_k ；第二階段為搜尋資料庫中 C_k 的支持度是否大於使用者最初設定的最小支持度門檻值的限制，符合條件的項目組 C_k 便稱為大項目組(Frequent k -itemset)，記為 L_k ，而不符合最小支持度限制的 C_k 項目組則刪除。

根據以上的步驟，而後再由 L_k 與 L_k 的聯集產生下一層的新候選項目組 C_{k+1} ，並再搜尋資料庫以產生 L_{k+1} 。如此反覆遞迴產生下一層級的 C_{k+1} 與 L_{k+1} ，直到無法產生長度更長的大項目組時停止。

以上所描述的步驟，我們舉一範例說明其流程。假設使用者所定

之最小支持度限制為 50%，其推導之過程如圖 2-2 所示：

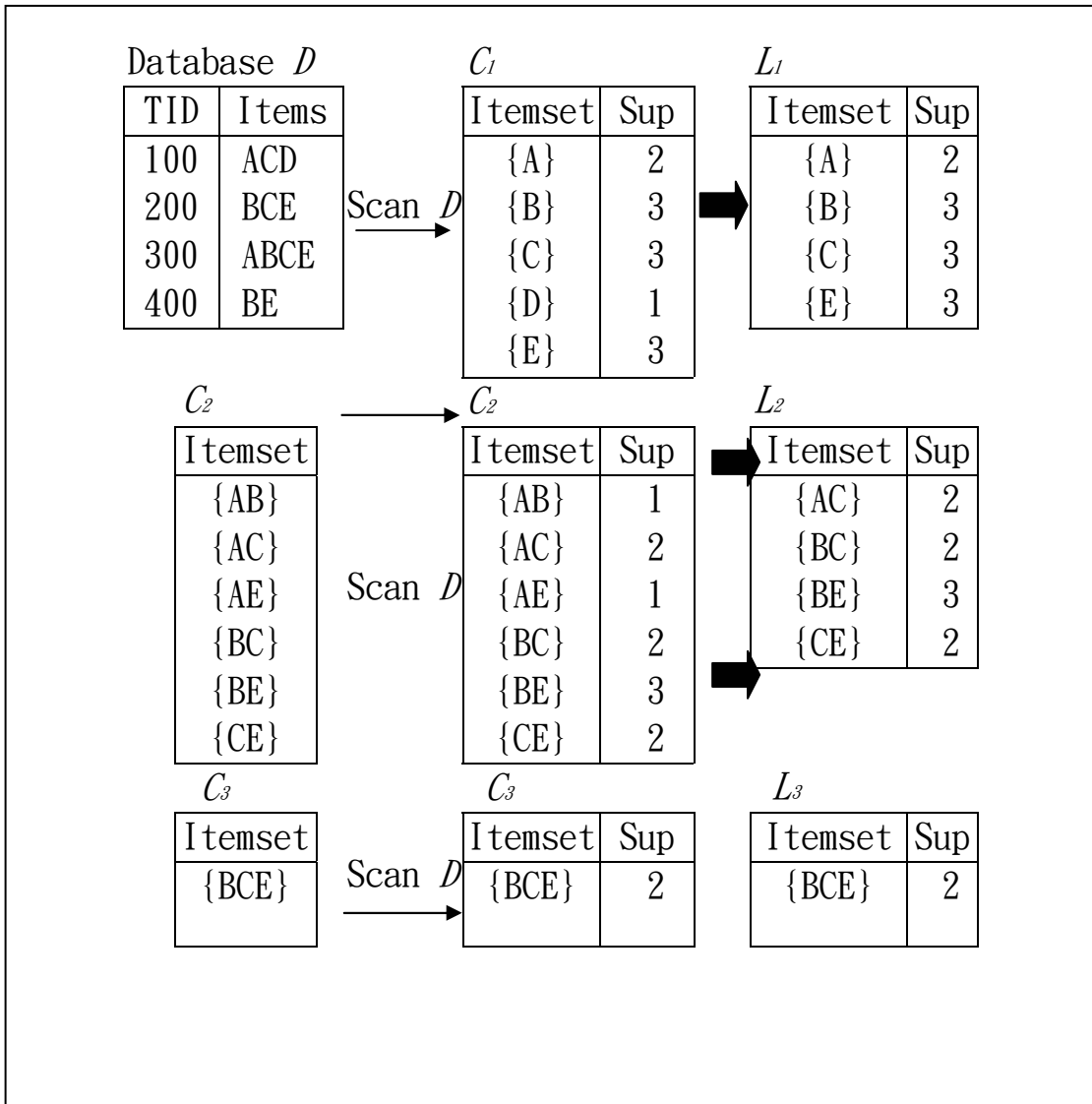


圖 2-2：產生大項目組之範例

(三) 關聯法則演算法相關衍生問題之介紹：

1. 漸進式關聯法則挖掘

現實情況中，資料庫一定絕大部份都是呈現動態狀態，在進行資料探勘後，不能因為後續資料遞增，必須重新挖掘一次，會太過於費時費力，已有很多研究來解決此問題。FUP 演算法便是一個很典型的例子。

Cheung *et al.* 為了改善 Apriori 演算法無法對動態資料庫做一有效的挖掘，提出了一種名為 FUP(Fast UPdate)的演算法【D.W. Cheung, J. Han, V. Ng, and C.Y. Wang,1997】，此一演算法以 Apriori 演算法為基礎，批次計算某一段時間所產生的交易資料，而主要的處理方式為將挖掘過的資料儲存至資料庫中，當新的交易資料進入時，再搜尋新增的資料庫找出其中的大項目組，只要在兩者其中的任一資料庫中為大項目組，則其有可能為更新後的大項目組，而若在兩資料庫中皆為非大項目組，則其更新後的結果也必為非大項目組。利用此新的大項目組與原本資料庫中的大項目組結合處理，便可得到資料庫更新後的關聯法則。

FUP 演算法的處理過程，其處理分為兩個過程，第一的過程為處理原始資料庫，求出各項目組的大項目組，此時處理方式與使用 Apriori 演算方式並無二致，但在處理完後，必須將各項大項目組的計數值加以儲存。第二個過程是當新增資料進入 FUP 演算法時，FUP 演算就新增資料加以處理，也是利用 Apriori 演算方式一樣，將新增項目中的大項目組取出，此時會發生下列四種情形：

- (1) 舊有資料庫中為大項目，新增資料中亦為大項目。
- (2) 舊有資料庫中為大項目，新增資料中為非大項目。
- (3) 舊有資料庫中為非大項目，新增資料中為大項目。
- (4) 舊有資料庫中為非大項目，新增資料中為非大項目。

FUP 演算法最主要的處理，就是在第二與第三種狀況。

FUP 演算法的優勢在於預先記錄大項目組的資訊，當新增資料進入計算時，僅需計算新增資料的部分，原始大項目組可以跟新增資料的 count 值加總，合併計算。如此可以避免重新搜尋整體資料庫所花費的時間及處理大量資料時所耗費的系統資源。

但 FUP 演算法在批次處理新增資料的時候，可以視為 Apriori 演算法，所以記憶體及 CPU 所消耗的系統資源與 Apriori 演算法一樣，會隨著資料量的增大而變多。再者，若項目組在新增資料中為大項目組而原始資料中為非大項目組，即必須再次搜尋整體的資料庫，而從原始大項目與新增項目合併計算後若成為非大項目，亦必須移出大項目組。若有一項目在大與非大項目組的門檻值間來回的改變，FUP 演算法就必須反覆的計算此一項目組，並將此項目組移入移出，造成系統的負擔。

2. 線上探勘

要能有效率的作線上探勘之關聯法則，必須考慮以下問題：

- (1) 要提供使用者線上調整門檻值，則不管使用者設定的門檻值為何，都必須能產生相對應的規則，因此儲存的資訊越詳細越好。

- (2)要達到線上探勘時，則必須在有限的時間之內產生所有的規則，因此在挖掘過程中較花費時間的掃描資料庫及計算工作必須在挖掘之前先完成。
- (3)在有限的記憶體空間中，要預先計算並儲存所有的項目集是不可能的。
- (4)若儲存的資訊越詳細，調整時就越能減少必須到資料庫中作確認的情況。

二、多階層關聯法則介紹

(一) 多階層關聯法則的由來與優勢：

多階層關聯法則的由來與其優勢，我們歸類下列四點來進行探討：

1. 商品項目多樣化所造成項目稀少性之問題

近年來，隨著市場競爭的激烈，為了滿足不同顧客的需求，商品項目漸趨於多樣化。若依據上節介紹的單層次關聯法則，很難很快達到我們所定的最小支持度，所找出來的法則數目可能會變少，甚至因此產生許多隱藏的知識無法被挖掘出來的問題。

2. 很難去找出具有強烈關聯性(strong association)的關聯法則

在資料挖掘的應用中，由於單層次關聯法則所搜尋的資料是屬於較低層次的(low level or primitive level)，有時候

會很難去找出具有強烈關聯性(strong association)的關聯法則，在實際的交易紀錄中，可能導致無法滿足我們所定的最小支持度。相較於搜尋層面較高的資料，我們可以比較容易找出較多的關聯法則。

3. 決策者需要大方向性的資訊來進行及時決策

由於現今企業在講求即時決策的環境下，當決策者在制定決策時，反而比較需要大方向性的資訊而非詳細的資訊。而這些較高層次的關聯法則可以提供決策者大方向的資訊與知識，讓決策者能夠即時了解大環境的趨勢，以下達決策。

4. 不同管理階層之決策者所需要的資訊是不同的

在企業組織中是由許多不同階層的管理者或員工所組成的。而這些管理者在每天的工作上也會遇到許多大小不同的決策。而不同階層的管理者在制定決策時，其所需要的資訊也有所差異。通常愈高層的管理者所需要的資訊是比較大方向性的資訊，反之，愈低層的管理者所需要的資訊是愈詳細的。而多階層關聯法則能夠同時滿足不同管理階層之決策者所需要資訊。

(二) 多階層關聯法則演算法介紹：

在眾多多階層關聯法則的研究中，所提出來的演算法，大致上可區分為利用單層次關聯法則演算法來進行多階層關聯法則探勘以及利用多層次方式來進行多階層關聯法則探勘【武家慶,2001】。前者的優勢是能夠進行跨階層的探勘，因此我們在這演算法介紹的部份，主要是介紹前者有名的演算法。

在介紹演算法之前，我們必須先介紹多階層關聯法則的前置處理作業。在進行多階層關聯法則之前，我們必須先將商品項目進行分類。而分類的方法可區分為兩種方式，一為透過GID編碼的方式來對項目進行分類，而另外一種是利用Taxonomy來進行分類。而不同的前處理方式也產生了許多不同的演算法。接下來，我們舉一個例子來說明這兩種分類方法。

假設以食物為例，若以Taxonomy來進行分類如下圖2-3

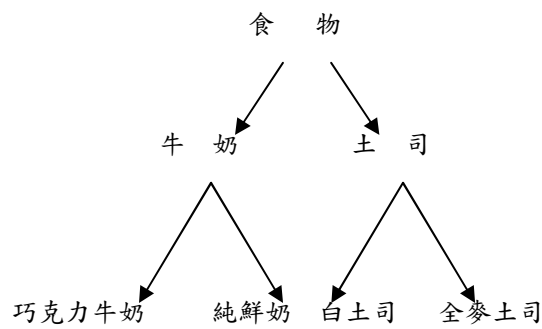


圖 2-3：食物項目分類以 Taxonomy 為例

若以GID編碼的方式其結果如下表2-2

表2-2:食物項目分類以Taxonomy為例商品項目	GID碼
食物	1**
牛奶	11*
土司	12*

巧克力牛奶	111
純鮮奶	112
白土司	121
全麥土司	122

介紹完不同分類的方法之後，接下來介紹利用單層次關聯法則進行多階層關聯法則演算法。著名的有Yongjian Fu et al(1995). 所提出的ML_T2L1演算法【Jiawei Han and Youngjian Fu,1995】與Ramakrishnan Srikant.et al(1997). 所提出的Cumulate演算法、Stratify演算法與EstMerge演算法【R. Srikant and R. Agrawal,1997】

1. ML_T2L1演算法

ML_T2L1演算法，首先，先對於商品項目進行GID編碼，然後再利用Apriori演算法來找出每一階層的大項目組。其演算法如下圖2-4

```

for (level=1;L[level,1]!= 0 and level < max-level;level++){
if  level = 1 {
L[level,1]=get_frequent1itemsets(T[1],level);
T[2]=get_frequent_t_table(T([1],L[1,1]));
}
else L[level,1]=get_frequent1itemsets(T[2],level);
for (k = 2;L[level, k-1] != 0;k++){
C[k]=apriori_gen(L[level,k-1]);
for each transaction t in T[2] {
C[t]:=get_subsets(C[k],t);
for each candidate c in C[t] do c.support++; }
L[level,k]={c in C[k]|c.support >= minsup[level]}}
LL[level]=Union(L[level,k]);
}

```

圖 2-4：ML_T2L1 演算法

對於 ML_T2L1 演算法，其限制為項目之間若要成為候選項目組的話，兩個項目彼此間不能有親子的關係存在。如此將產生大量的候選項目組，按照以往的做法，又將比對資料庫後才能找出大項目組，在效能上將大打折扣。所以，雖能提供跨階層關聯法則的求取，但是並不方便。

2. Cumulate演算法

在1997年，Ramakrishnan Srikant. *et al.* 提出了Cumulate演算法。Cumulate演算法是透過Taxonomy架構中的商品項目，來求得關聯法則。至此，對於求取多層次關聯法則的方式，不再侷限於使用傳統的方式。Srikant認為找出歸納後的商品項目之間的關聯法則，所會遇見的問題，可以分為三個部分：

- (1) 找出所有的大項目組。
- (2) 使用大項目組來產生關聯法則。
- (3) 去除不感興趣的關聯法則。

為了解決上述三個問題，來找出滿足使用者的關聯法則，Srikant提出了兩種類型的演算法。第一種偏向傳統演算法的方式，主要是改善其搜尋的資料量；第二種是利用階層的概念，

若上層不滿足最小支持度，其下層亦不滿足最小支持度，所以可以忽略不去計算，也是減少搜尋的資料量。

Cumulate演算法是屬於第一類型，主要是將傳統的演算法，透過以下三點最佳化的方式來加以改善：

(1)只將出現在候選商品項目組中的祖先加入交易紀錄中，而非全部商品項目組的祖先都加入。

(2)先找出每個商品項目的祖先。透過所有商品項目的祖先，便不需再循者Taxonomy架構來找出其祖先。

(3)去除同時包含商品項目及其祖先的商品項目組。

其演算法如下圖2-5：

```
Compute T*, the set of ancestors of each item, from T.
L1 := {frequent 1-itemsets};
k := 2;
while (Lk-1 ≠ ∅) do
  begin
    Ck := New candidates of size k generated from Lk-1.
    if (k = 2) then
      Delete any candidate in C2 that consists of an item and its ancestor.
      Delete any ancestors in T* that are not present in any of the candidates in Ck.
    forall transactions t ∈ D do
      begin
        foreach item x ∈ t do
          Add all ancestors of x in T* to t.
          Remove any duplicates from t.
          Increment the count of all candidates in Ck that are contained in t.
        end
      Lk := All candidates in Ck with minimum support.
      k := k+1;
    end
  Answer := ∪k Lk;
```

圖 2-5：Cumulate 演算法

Cumulate演算法雖然是透過三個最佳化的方式來減少搜尋的資料量，但是仍需要事先準備祖先的資料表。隨著商品項目組的長度增加，使用三個最佳化的方式雖能減少許多不必要的計算時間，但仍需反覆的計算各長度的大項目組。而使用 Taxonomy 架構的好處，就是當找出{牛奶，麵包}、{純鮮奶，麵包}及{低指純鮮奶，麵包}三個長度為2之候選商品項目組時，若{牛奶，麵包}無法滿足最小支持度時，則就不需再去比對{純鮮奶，麵包}及{低指純鮮奶，麵包}。相對的，若{牛奶，麵包}滿足最小支持度時，再去比對{純鮮奶，麵包}。若{純鮮奶，麵包}滿足最小支持度時，再去比對{低指純鮮奶，麵包}。透過這樣的觀念，R. Srikant 又提出幾種演算法，都是屬於第二類型的演算法。在此僅簡單介紹Stratify及EstMerge這兩種演算法。

3. Stratify演算法

Stratify演算法也是透過Taxonomy架構，首先計算第0層的候選商品項目。將滿足最小支持度的商品項目用來計算第1層的候選商品項目。而第0層中，不滿足最小支持度的商品項目，就不再計算之後的階層。如此反覆下去，直到找不出商品項目為止。這種做法將浪費許多時間在商品項目組的計算與比對資料庫。

4. EstMerge演算法

EstMerge演算法是先找出高頻1商品項目組，再使用隨機取樣的方式，對資料庫取出一部分進行候選商品項目組的比對。若高頻1商品項目於隨機的資料集中能滿足最小支持度，則下一層商品項目便被視為候選商品項目組。祖先及其子孫都再經由搜尋真實資料庫，以證實其是否真正的滿足最小支持度。

綜合上述此四種多階層關聯法則的特點與缺失整理於表2-3：

表2-3：多階層關聯法則演算法比較表

	ML_T2L1 演算法	Cumulate演算法	Stratify演算法	EstMerge演算法
結構	透過GID碼結構來進行多階層挖掘	透過 Taxonomy 架構來進行多階層挖掘	透過 Taxonomy 架構來進行多階層挖掘	透過 Taxonomy 架構來進行多階層挖掘
大項目組產生方式	以傳統演算法 (Apriori) 的方式	以傳統演算法 (Apriori) 的方式並結合三點改善方式進行	利用階層的概念，若上層不滿足最小支持度，其下層亦不滿足最小支持度	1. 利用階層的概念，若上層不滿足最小支持度，其下層亦不滿足最小支持度 2. 採用隨機取樣方式
是否能夠進行跨階層探討	○：但效率與方便性較差	○：效率與方便性較佳	○：效率與方便性較佳	○：效率與方便性較佳
是否能夠解決商品多樣性所導致法則稀少	○	○	○	○

問題或是無法產生法則等問題				
支援大方向性的決策	○	○	○	○
動態資料庫挖掘	×	×	×	×
線上挖掘	×	×	×	×

第三節 模糊邏輯與模糊切割

一、模糊邏輯之定義及其運算

模糊邏輯又稱為多元邏輯，由 Zadeh 在 1965 年所提出的【Zadeh,1965】。在傳統明確邏輯而言，多具有明確、精準及確定的特性，所謂明確是指的是二分法（非真即假或是元素必然歸屬於或不屬於某集合兩種中其中一種）；精確指的是建模所使用的參數都能很精確地表達；確定則是建模的結構與參數都能被明確的知道，不論是數值或是存在性都沒有任何疑問。相較於傳統的邏輯，模糊邏輯是屬於多元邏輯，也就是除了是真或是假外還允許許多漸進的值。舉個例子來說，假設現在有 3 個 15 歲、3 個 20 歲、3 個 29 歲、2 個 36 歲和 2 個 45 歲的人們，請問有幾個青年人？這個答案就很難確定，因為每個人的認知都有所不同，因此，在這個例子中，傳統明確的邏輯概念就不適合，而模糊理論就是專門處理這方面的問題。

透過上述的模糊理論的概念來對傳統集合的觀念做個延伸而產生了模糊集合，而模糊集合也是模糊理論當中最重要，因此本節將介紹模糊集合的一些基本概念。模糊集合的表示定義如下：

定義 1：當 x 在 U 的範圍內的元素，會對應到一個介於 0 至 1 之間的

數值 $\mu_A(x)$ 。所有 x 所成的集合就是一個模糊集合。(U 為論域、A 為一模糊集合、 $\mu_A(\cdot)$ 為歸屬度函數) 其表示如下：

$$A = \{(x, \mu_A(x)) \mid x \in U\}$$

由於模糊集合可以視為傳統集合的擴充，因此模糊集合也具有傳統集合的運算性質，例如包含、相等、補集、交集、聯集……等。在這裡我們主要是運用到交集、聯集這兩個運算，所以就介紹這兩種運算。首先，令 A、B 為宇集合 U 的兩個模糊集合，而歸屬度函數分別為 μ_A 、 μ_B 且 $A = \{(u_i, \mu_A(u_i)) \mid u_i \in U\}$ ， $B = \{(u_i, \mu_B(u_i)) \mid u_i \in U\}$ ，則兩模糊集合的交集與聯集運算定義如下：

定義 2：模糊集合 A 交集 B， $A \cap B$ ，其標準交集定義如下：

$$\mu_{A \cap B}(u_i) = \min(\mu_A, \mu_B) \quad u_i \in U \dots\dots\dots (2-1)$$

模糊理論的標準運算並非唯一的，還有其他常見的，像利用代數積方法如下：

$$\mu_{A \cap B}(u_i) = \mu_A \otimes \mu_B = \mu_A \times \mu_B \quad u_i \in U \dots\dots\dots (2-2)$$

定義 3：模糊集合 A 聯集 B， $A \cup B$ ，其標準聯集定義如下：

$$\mu_{A \cup B} = \max(\mu_A, \mu_B) \quad u_i \in U \dots\dots\dots (2-3)$$

利用代數和方法如下：

$$\mu_{A \cup B} = \mu_A + \mu_B - (\mu_A \times \mu_B) \quad u_i \in U \dots\dots\dots (2-4)$$

二、模糊邏輯之歸屬度函數介紹

在模糊理論中，歸屬度函數可以說是相當重要且基本的概念，透過

歸屬度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【李允中、王小璠、蘇木春,1993】。根據文獻我們可以發現歸屬度函數有相當多種，最常被應用的有三角形、梯形、L-R、指數函數、S函數等【李允中、王小璠、蘇木春,1993】。

如何找出一個適當的歸屬度函數是目前模糊理論研究中相當實際且重要的問題，一般來說並沒有通用的定理或公式，通常是依據經驗或統計方法來加以確定，很難具客觀性。許多研究學者希望能夠透過系統化的方式來找出比較客觀的歸屬度函數，最常見的做法是先建立粗略的歸屬度函數，然後藉由學習與不斷地實驗，逐步進行修正和調整使得歸屬度函數更加客觀。簡單三角形歸屬函數可使系統具有較佳的執行效率，且已成功應用在許多問題上【Yen John and Reza Langari,1999】，因此，在不失一般性的前提下，本研究先利用三角形歸屬度函數進行探討，其定義如下所示：

$$\mu_A(a,b,c;x) = \begin{cases} 0, & x < a \\ (x-a)/(b-a), & a \leq x \leq b \\ (c-x)/(c-b), & b \leq x \leq c \\ 0, & x > c \end{cases} \quad (2-5)$$

其中， a, b, c 分別為定義三角模糊歸屬度函數的三個參數，其圖形如下圖 2-6 所示：

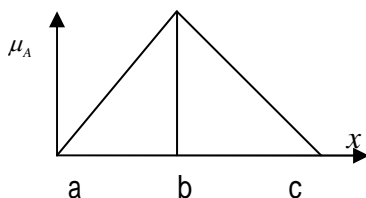


圖 2-6: 三角形歸屬度函數

三、模糊切割介紹

在許多模糊理論的應用中，Fuzzy if-then rules 可說是最成功且最

廣的。而模糊切割 (Fuzzy partition) 是在 Fuzzy if-then rules 中相當重要的概念。Ishibuchi 在 1992 年【Ishibuchi and Tanaka,1992】與 1995 年【Ishibuchi and Tanaka,1995】分別提出了 Simple fuzzy partition 與 Multiple fuzzy partition 的概念。本研究主要是利用 Simple fuzzy partition 的概念來建立關聯法則，因此我們在此主要是介紹 Simple fuzzy partition 的概念與用法。

假設在維度 D 上建立 K 個模糊集合，如圖 2-7 所示，則第 i 個集合的模糊歸屬函數定義如下 (2-6)。

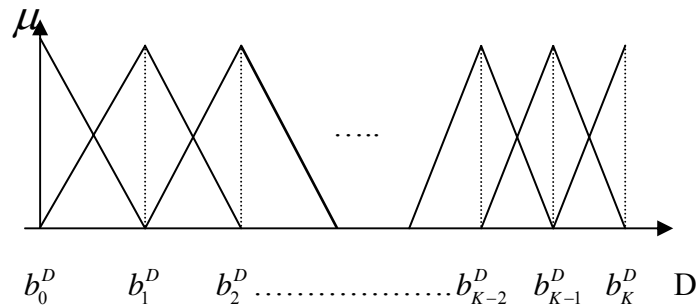


圖 2-7: 以三角形歸屬度函數建立的 Simple fuzzy partition

$$\mu_i^D(x) = \begin{cases} 0, & x < a_i \\ (x - a_i) / (b_i - a_i), & a_i \leq x \leq b_i \\ (c_i - x) / (c_i - b_i), & b_i \leq x \leq c_i \\ 0, & x > c_i \end{cases} \quad \text{其中 } 1 \leq i \leq K \dots\dots\dots (2-6)$$

一般而言，均分且對稱的三角形模糊歸屬函數即可有不錯的效果。

因此我們可以求得 a, b, c 的間距 $S = \frac{(x_{\max}^D - x_{\min}^D)}{(K-1)}$ 其中 x_{\max}^D 與 x_{\min}^D 分別

為維度 D 資料的最大值與最小值，則第 i 個模糊歸屬函數的參數分別為

$$\begin{aligned} b_i &= x_{\min}^D + S \cdot (i-1) \\ a_i &= b_i - S \quad \dots\dots\dots (2-7) \\ c_i &= b_i + S \end{aligned}$$

接下來我們舉個例子說明之：

範例 1：

資料中的年齡分別為 (10, 20, 15, 0, 20, 10, 60, 35, 50, 40) 年紀最大者為 60 歲；最小則為 0 歲。我們把年齡這十筆資料分成三個集合分別為年輕人 (A_1^{Age})、中年人 (A_2^{Age})、老年人 (A_3^{Age}) 則

$S = \frac{(60-0)}{(3-1)} = 30$ ，而可計算得模糊歸屬函數之參數為：

$$\begin{aligned} (a_1, b_1, c_1) &= (-30, 0, 30) \\ (a_2, b_2, c_2) &= (0, 30, 60) \\ (a_3, b_3, c_3) &= (30, 60, 90) \end{aligned}$$

因此我們可分別計算出 10 歲屬於年輕人、中年人、老年人的歸屬度為：

$$\begin{aligned} \mu_1^{Age}(10) &= \frac{(30-10)}{30} = 0.667 \\ \mu_2^{Age}(10) &= \frac{(10-0)}{30} = 0.337 \\ \mu_3^{Age}(10) &= 0 \end{aligned}$$

依此類推可計算出 10 筆年齡的歸屬度。



接下來，我們將各維度所分割之模糊集合的相關參數記錄起來做為後續處理使用。最後，我們可由一維模糊切割進一步建立多維的模糊切割，以便進行模糊關聯法則的挖掘【Hu,Chen, and Gwo-Tzeng,2003】。假設各

維分割的模糊集合(Fuzzy sets)數目分別為 K_1, K_2, \dots, K_n 且 $x = (x_1, x_2, \dots, x_n)$ 表示一筆 n 維的資料，則第 i 維資料值在其第 t_i 個模糊集合的歸屬度為 $\mu_{t_i}^{D_i}(x_i)$ 其中 $1 \leq t_i \leq K_i$ 如此，我們可以計算 x 在模糊集合 t_1, t_2, \dots, t_n 所組合之關係的歸屬度為：

$$\mu_{t_1 \times t_2 \times \dots \times t_i \times \dots \times t_n}^{D_1 \times D_2 \times \dots \times D_i \times \dots \times D_n}(x_1, x_2, \dots, x_i, \dots, x_n) = \mu_{t_1}^{D_1}(x_1) \otimes \mu_{t_2}^{D_2}(x_2) \otimes \dots \otimes \mu_{t_i}^{D_i}(x_i) \otimes \dots \otimes \mu_{t_n}^{D_n}(x_n) \quad \dots \dots$$

(2-8)

其中 \otimes 為 Fuzzy conjunction operator，若取代數積(Product operation) 作為 Fuzzy conjunction operator 則上式可寫成

$$\mu_{t_1 \times t_2 \times \dots \times t_i \times \dots \times t_n}^{D_1 \times D_2 \times \dots \times D_i \times \dots \times D_n}(x_1, x_2, \dots, x_i, \dots, x_n) = \mu_{t_1}^{D_1}(x_1) \times \mu_{t_2}^{D_2}(x_2) \times \dots \times \mu_{t_i}^{D_i}(x_i) \times \dots \times \mu_{t_n}^{D_n}(x_n) \quad \dots \dots$$

(2-9)

若取最小運算 (Min operation) 則 (2-8) 式可寫成

$$\mu_{t_1 \times t_2 \times \dots \times t_i \times \dots \times t_n}^{D_1 \times D_2 \times \dots \times D_i \times \dots \times D_n}(x_1, x_2, \dots, x_i, \dots, x_n) = \min(\mu_{t_1}^{D_1}(x_1), \mu_{t_2}^{D_2}(x_2), \dots, \mu_{t_i}^{D_i}(x_i), \dots, \mu_{t_n}^{D_n}(x_n))$$

(2-10)

接下來我們舉個例子說明之：

範例二：

資料有年齡和薪資這兩個維度，共 8 筆資料分別如下表 2-4：

表 2-4： 範例資料表

	年齡	薪資
r_1	10	1000
r_2	20	15000
r_3	15	5000
r_4	0	0
r_5	20	10000

r_6	10	2000
r_7	60	40000
r_8	35	50000

我們在年齡與薪資上各定義三個集合與五個集合，分別為年輕人 (A_1^{Age})、中年人 (A_2^{Age})、老年人 (A_3^{Age})、最低收入 (A_1^{Salary})、低收入 (A_2^{Salary})、中收入 (A_3^{Salary})、高收入 (A_4^{Salary})、最高收入 (A_5^{Salary})，則 $S^{age} = \frac{(60-0)}{(3-1)} = 30$ 和 $S^{salary} = \frac{(50000-0)}{(5-1)} = 12500$ ，計算得年齡的三角模糊歸屬函數之參數

分別為：

$$(a_1, b_1, c_1) = (-30, 0, 30)$$

$$(a_2, b_2, c_2) = (0, 30, 60)$$

$$(a_3, b_3, c_3) = (30, 60, 90)$$

而薪資的三角模糊歸屬函數之參數分別為：

$$(a_1, b_1, c_1) = (-12500, 0, 12500)$$

$$(a_2, b_2, c_2) = (0, 12500, 25000)$$

$$(a_3, b_3, c_3) = (12500, 25000, 37500)$$

$$(a_4, b_4, c_4) = (25000, 37500, 50000)$$

$$\text{及 } (a_5, b_5, c_5) = (37500, 50000, 62500)。$$

以第一筆為例，年齡值為 10，其歸屬度計算為：

$$\mu_1^{Age}(10) = 0.667, \mu_2^{Age}(10) = 0.333, \mu_3^{Age}(10) = 0$$

而薪資 1000 元之歸屬度計算為：

$$\mu_1^{salary}(1000) = 0.92, \mu_2^{salary}(1000) = 0.08, \mu_3^{salary}(1000) = 0, \mu_4^{salary}(1000) = 0, \text{ 及 } \mu_5^{salary}(1000) = 0。$$

因此，可以算出第一筆紀錄在二維模糊分割的每一個子方塊 (subcube) 的歸屬度分別為：

$$\mu_{1 \times 1}^{Age \times Salary}(10, 1000) = \mu_1^{Age}(10) \times \mu_1^{Salary}(1000) = 0.667 \times 0.92 = 0.614,$$

$$\mu_{1 \times 2}^{Age \times Salary}(10, 1000) = \mu_1^{Age}(10) \times \mu_2^{Salary}(1000) = 0.053,$$

$$\mu_{2 \times 1}^{Age \times Salary}(10, 1000) = \mu_2^{Age}(10) \times \mu_1^{Salary}(1000) = 0.306,$$

$$\mu_{2 \times 2}^{Age \times Salary}(10, 1000) = \mu_2^{Age}(10) \times \mu_2^{Salary}(1000) = 0.027,$$

其他的歸屬度皆為 0，依此類推可算出其他 7 筆的歸屬度。

第三章 研究方法

第一節 研究架構

首先，從顧客 RFM 資料，以 fuzzy partition 建立 3-dimensional fuzzy RFM DataCube，由使用者自訂的最小支持度過濾出模糊大項目組。利用這 Model 可將做顧客分群及挖掘出 RFM 值間的模糊關聯法則。再者，整理顧客交易資料庫的商品項目，建立商品分類樹 (Taxmony)，並依商品被購買的次數計算模糊值，儲存在叢聚式模糊集表後，過濾出模糊大項目組，挖掘出商品項目間的模糊關聯法則。最後，掃描 RFM 資料庫三維 RFM 顧客分群，計算各分群的支持度，過濾出 RFM 顧客分群的模糊大項目組 (支持度大於最小支持度) 及 fuzzy large itemsets 建立兩者間的關係矩陣，並將大於最小信賴度的關係轉換成型如 R. L&F. H&M. H→D. Low 的關聯法則。

此模糊規則挖掘法如圖 3-1 可分解成上述的三大步驟：

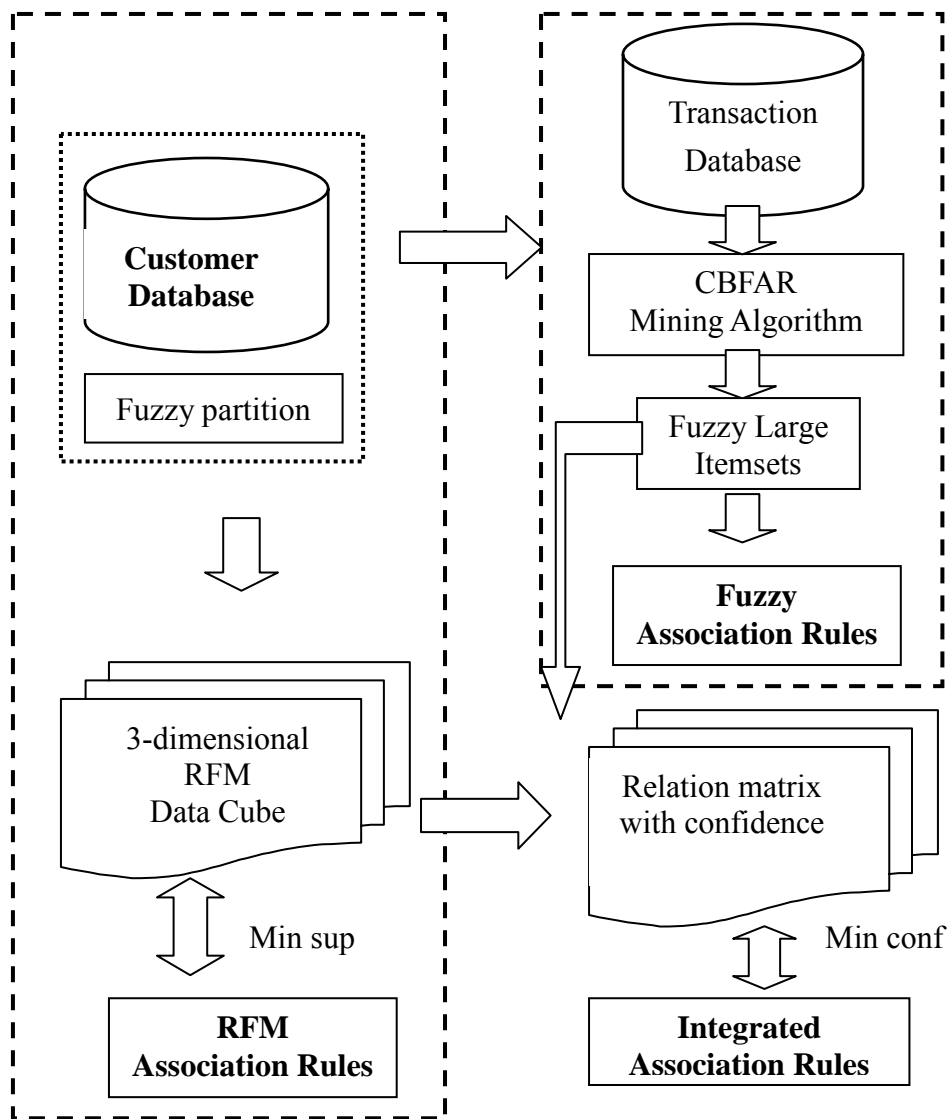


圖 3-1 叢聚式模糊關聯法則架構圖

第二節 模糊 RFM Model

一、Fuzzy RFM Model 流程架構：

我們利用模糊切割 (Fuzzy partition) 的觀念將 RFM 值量化並表示成 3 維的資料方塊 (Data Cube) 以建構一個多用途的模糊 RFM Model，

如圖 3-2 所示。利用該 Model 可獲得顧客特性分析、顧客分群以及 RFM 值變化的關聯法則等知識，使得相較於傳統的 RFM Model 能夠以客觀且彈性的方法提供企業最有效的資訊來制定決策以便因應如此變化快速的環境。

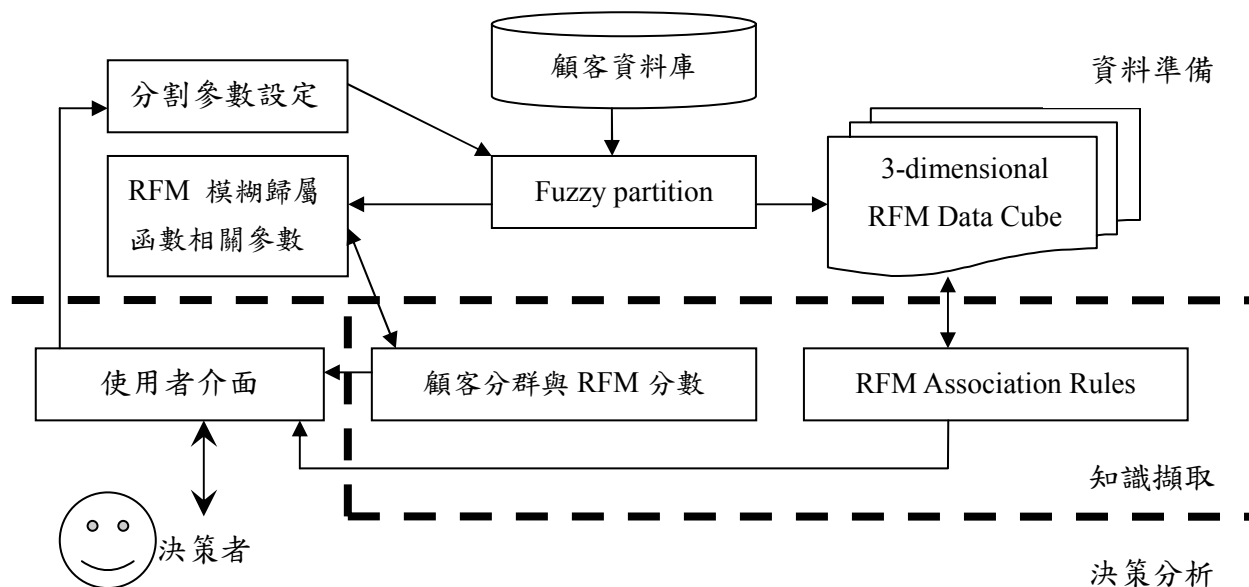


圖 3-2：模糊 RFM Model 流程架構圖

根據 Fuzzy RFM Model 的流程，我們可分為三個部分。分別為資料準備、知識擷取、以及最後結果的決策分析。在資料準備的部分，Fuzzy partition 已在參考文獻（第二章第三節）中詳細說明了，因此，在此部分，我們主要將探討 R、F、M 這三個參數如何在資料方塊中建構；在知識擷取的部分，主要分為顧客分群以及 RFM Association Rules，也是 Fuzzy RFM Model 最主要的功能部分；最後，決策分析的部分，在此我們並不深入探討此部分，理由是每個產業的環境以及每個經營者分析的角度與其需求不同而有所不同的分析方式。

二、RFM 值的 3 維資料方塊之建構：

當我們在 RFM 三個維度進行模糊分割而建立各維度的模糊集合之後，可以將各模糊集合視為一項目，如此可以利用 Apriori-like 演算法擷取 R、F、M 項目之間的模糊關聯法則【楊昇樺,2002】。但當資料庫中的資料量很龐大時，Apriori 演算法【Agrawal and Srikant,1994】是相當費時的。因為 RFM 只有 3 維，因此我們可以事先計算這些項目組的模糊支持度值並儲存在資料方塊之中，如此不必掃描資料庫即可快速求出關聯法則，達成線上挖掘的功能，而當新增紀錄時，亦可即時加以處理並更新資料方塊中對應項目組的支持度，以達到漸進式挖掘的目的。RFM 值的 3 維資料方塊的圖示如下圖 3-3:

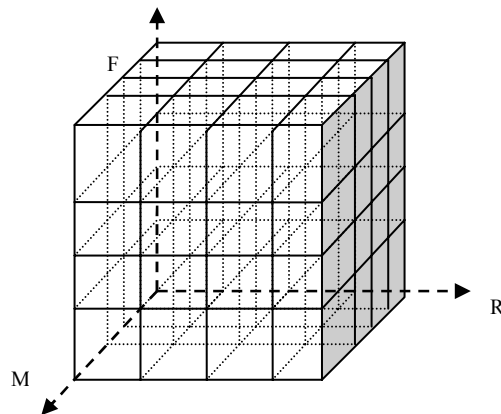


圖 3-3：RFM 3 維資料方塊

假設 R 值、F 值、M 值分割成 K_R, K_F, K_M 個模糊集合，則資料方塊的大小為 $(K_R + 1)(K_F + 1)(K_M + 1)$ 。換句話說其 Subcube 的座標可由 $(0,0,0)$ 到 (K_R, K_F, K_M) 。令 $1 \leq r \leq K_R, 1 \leq f \leq K_F, 1 \leq m \leq K_M$ 則座標 $(r,0,0), (0,f,0), (0,0,m)$ 儲存的是 1-itemset 的模糊支持度值；座標 $(r,f,0), (r,0,m), (0,f,m)$ 儲存的是 2-itemset 的模糊支持度值；而座標 (r,f,m) 儲存的是 3-itemset 的模糊支持度值。三維 RFM Data cube 的建構方式如下所示：

步驟 1 從顧客交易資料庫擷取出顧客 RFM 資料表

步驟 2 使用者設定參數 K_R, K_F, K_M 以便系統由 RFM 資料表求所有歸屬函數的三個參數

步驟 3 利用公式 (9) 或公式 (10) 計算每一筆紀錄在 K-itemsets 的歸屬度值，並利用公式 (11) 算出 K-itemsets 的模糊支持度至資料方塊所對應之 Subcube 中

三、Fuzzy RFM Model 之知識擷取：

(一) 顧客分群：

顧客分群是行銷策略成功的一個重要前置步驟，假設利用模糊切割將 R、F、M 分別定義 L、M、H 三個模糊集合，則我們可得到如表 3-1 的顧客分群表【李健興 2000】。我們計算出每一筆 R、F、M 的歸屬程度後，根據所算出來的數值以及企業所制定的顧客分群表可找出每一個顧客所屬的群和類型，以作為企業行銷決策的依據。

表 3-1：顧客分群表

群	R	F	M	顧客類型
1	L	H	H	忠誠的老顧客(高消費非理性型)
2	L	H	M	忠誠的老顧客(中低消費非理性)
3	L	H	L	忠誠的老顧客(中低消費非理性)
4	L	M	H	忠誠的長期顧客(高消費半理性型)
5	L	M	M	潛在型顧客(中低消費半理性)

6	L	M	L	想到才來的顧客
7	L	L	H	充滿好奇的顧客
8	L	L	M	新顧客
9	L	L	L	過客
10	M	H	H	有一段時間高忠誠顧客(高消費非理性型)
11	M	H	M	有一段時間潛在型顧客或即將流失的顧客 (中低消費非理性)
12	M	H	L	有一段時間低忠誠顧客即將流失的顧客(中 低消費非理性)
13	M	M	H	有一段時間高忠誠顧客(高消費半理性型)
14	M	M	M	有一段時間潛在型顧客或即將流失的顧客 (中低消費半理性)
15	M	M	L	有一段時間低忠誠顧客或即將流失的顧客
16	M	L	H	有一段時間低忠誠顧客或潛在型顧客(高消 費理性型)
17	M	L	M	有一段時間低忠誠顧客或潛在型顧客(中低 消費理性)
18	M	L	L	有一段時間低忠誠顧客

19	H	H	H	離開很久的高忠誠顧客(高消費非理性型)
20	H	H	M	離開很久的潛在型顧客或即將流失的顧客 (中低消費非理性)
21	H	H	L	離開很久的低忠誠顧客即將流失的顧客(中 低消費非理性)
22	H	M	H	離開很久的高忠誠顧客(高消費半理性型)
23	H	M	M	離開很久的潛在型顧客或即將流失的顧客
24	H	M	L	離開很久的低忠誠顧客或即將流失的顧客
25	H	L	H	離開很久的低忠誠顧客或潛在型顧客(高消 費理性型)
26	H	L	M	離開很久的低忠誠顧客或潛在型顧客(中低 消費理性)
27	H	L	L	離開很久的低忠誠顧客

(二) RFM Association Rules 的挖掘：

因為所有 itemsets 的模糊支持度值已經存在資料方塊中，而且 Data cube 並不大，所以我們可以快速掃描資料方塊並與門檻值（最小模糊支持度）比對來產生大項目組（K-itemsets）再由大項目組推導出關聯法則，並與門檻值（最小模糊信賴度）比對產生符合使用者的法則。

假定在 RFM 上分割之模糊集合的個數分別為 K_R, K_F , 與 K_M ; 使用者制定的門檻值分別為：最小模糊支持度 minfs 與最小模糊信賴度 minfc , 則挖掘 Fuzzy RFM Association Rule 的步驟與演算法如下所示：

步驟 1 掃描 3 維 RFM 資料方塊中，對應 1-itemsets 的 Subcubes，若其儲存的模糊支持度大於 minfs ，則加入 Large RFM

1-itemsets 的集合 L

步驟 2 產生高頻 RFM 2-itemsets，有兩種方式可用：

- A. 合併 L_1 產生 C_2 ，再到 Data cube 中檢查 C_2 中每一個候選 2-itemsets 的模糊支持度，若大於 minfs ，則加入集合 L_2 中。
- B. 直接掃描 Data cube 中，對應 2-itemsets 的 Subcubes 之模糊支持度，若大於 minfs ，則加入 L_2 中。

步驟 3 同理可產生 L_3

步驟 4 由 L_1, L_2 與 L_3 產生模糊 RFM 關聯法則。若 $FC(R) \geq \text{minfc}$ ，則 R 為所要之有效的關聯法則。

第三節 應用在交易資料庫的模糊叢聚式關聯法則

(Cluster-Based Fuzzy Association Rules)

本研究提出的 CBFAR 挖掘演算法則為從交易資料庫的商品項目階層關係及建立叢聚式模糊集表 (cluster-based fuzzy sets tables) 之模糊關聯法則挖掘技巧。

如圖 3-4 所示，可將此演算法分成二步驟：

步驟一：產生模糊大項目集

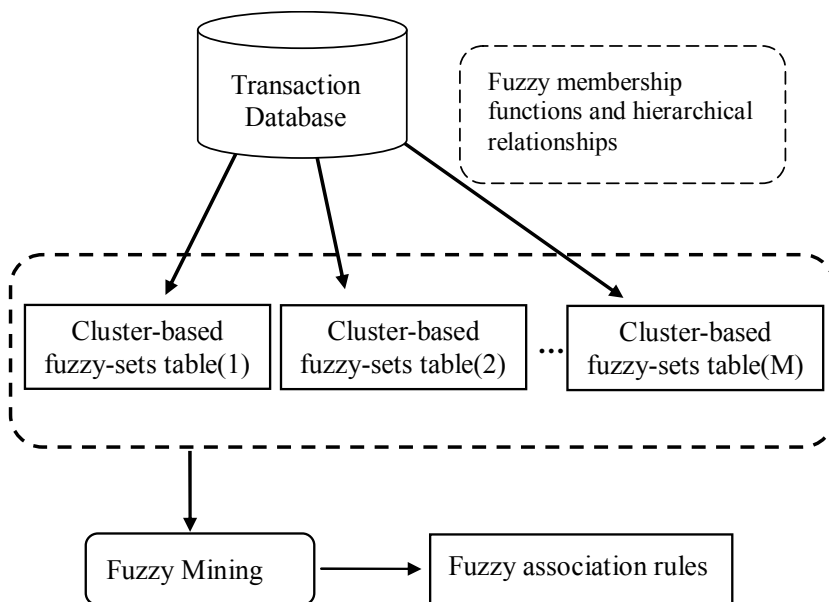


圖 3-4 CBFAR 挖掘演算法則

(1) 掃瞄資料庫以產生叢聚式模糊集表 (cluster-based fuzzy sets tables)。首先依分類樹 (taxonomy) 的分類，將項目的“祖先”加入，形成含祖先的擴充型交易項目集。交易項目數一樣多的就形成一個叢聚式模糊集 (cluster-based fuzzy set)。

(2) 再將含祖先的擴充型交易項目的數值型資料代入商品項目的歸屬函數，計算其模糊值，並儲存在叢聚式模糊集表對應的位置。

(3) 由 L_{n-1} 的項目集自我結合 (self-join) 成候選項目集 C_n 長度為 k 的候選項目集，其支持度是由計算叢聚式模糊集 (k) 至叢聚式模糊集 (M) 模糊值的和所得。假如 C_n 的支持度大於等於最小支持度 α ，候選項目集

就變成大項目集，將 C_n 置入大項目集 L_n 。否則與下一個叢聚式模糊集表 (k+1) 做比較。當大於等於最小支持度或到達叢聚式模糊集 (M) 的盡頭，則停止做比較的動作。

步驟二： 挖掘模糊關聯法則

使用由 CBFAR 方法所產生的模糊大項目集及最小信賴度挖掘商品項目間的模糊關聯法則。

第四節 應用在 RFM 及交易資料庫間的模糊叢聚式關聯法則的挖掘

(Cluster-Based Fuzzy Association Rules)

1. 使用 Fuzzy RFM model，找出三維 RFM 客戶分群的大項目集（支持度大於最小支持度）：給定一個最小支持度，可掃描模糊 RFM data cube 過濾出三維的模糊大項目集 $LRFM_i$ ，而此 $LRFM_i$ 對應的顧客集可記為 $LRFM_i_CID$ 。
2. CBFAR 挖掘演算法則，找出商品項目的大項目集：給定一個最小支持度，可掃描如第二節所用的叢聚式模糊集表過濾出每個模糊大項目集 L_j ，而此對應的顧客集可記為 L_j_CID 。

3. 由 $LRFM_i$ 及 L_j 所組合成的關係矩陣 R ，每個元素 $R(LRFM_i, L_j)$ 儲存 $LRFM_i$ 及 L_j 的信賴度。假如 $(LRFM_i_CID\ sets) - (L_j_CID\ sets) = \phi$ ，則它的信賴度 $\beta = 1$ ；否則，信賴度 $\beta = n((LRFM_i \cap L_j)_CID\ sets) / n(LRFM_i_CID\ sets)$ ，其中 $n(S)$ 表集合 S 的大小。
4. 假設 $(LRFM_i, L_j)$ 的信賴度大於等於最小信賴度，將這些關係轉換成型如 If-then 的規則。

第四章 實例演練

在本章我們舉三個例子分別說明如何挖掘顧客 RFM 資料庫模糊關聯法則、交易資料模糊叢聚式法則與顧客 RFM 及交易資料庫間的模糊關聯法則。

例 1. 顧客 RFM 資料庫模糊關聯法則

假設有 30 個顧客，其 RFM 資料表如表 4-1 所示

表 4-1 顧客 RFM 資料表

顧客編碼	R	F	M
1	256	17	22122
2	194	1	1145
3	210	20	32673
4	105	1	1145
5	109	2	1361
6	282	7	26787
7	258	10	18753
8	16	1	1145
9	150	2	1373
10	313	1	1145
11	287	4	3995
12	317	9	19762
13	345	6	3576
14	132	11	20367
15	191	2	2065
16	19	4	12139
17	108	5	10139
18	226	9	20536
19	235	15	32063
20	96	13	25745

21	101	30	92685
22	302	7	3740
23	300	13	31722
24	214	9	18587
25	331	12	32872
26	82	16	23288
27	253	10	19025
28	88	5	12520
29	194	3	7461
30	38	9	14466

R、F、M 歸屬函數如圖 4-1 所定義，

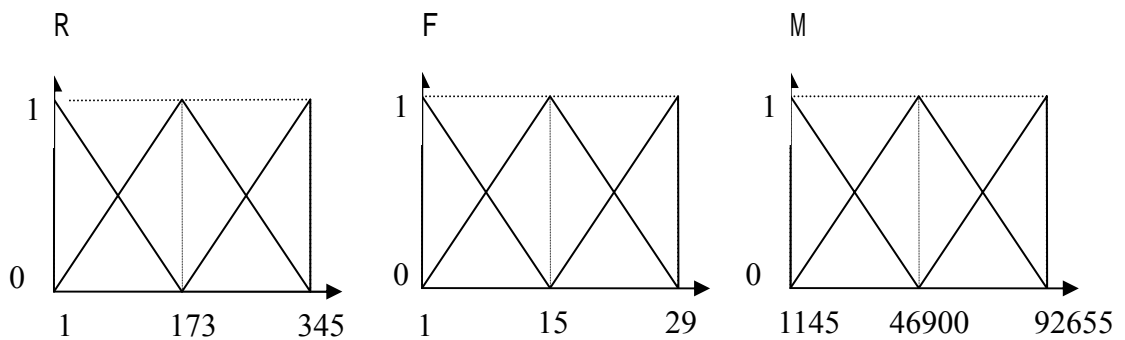


圖 4-1 R、F、M 歸屬函數

再將 30 筆資料的 R、F、M 值個別代入圖 4-1 R、F、M 歸屬函數，可得到顧客 RFM 資料表對應的模糊值表，如表 4-2 所示。

表 4-2 顧客 RFM 資料表對應的模糊值表

	R			F			M		
1	256	rl	0	17	fh	0.14	22122	mh	0
		rm	0.52		fm	0.86		mm	0.46

		rh	0.48		fl	0		ml	0.54
2	194	rl	0	1	fh	0	1145	mh	0
		rm	0.89		fm	0		mm	0
		rh	0.11		fl	1		ml	1
3	210	rl	0	20	fh	0.36	32673	mh	0
		rm	0.79		fm	0.64		mm	0.69
		rh	0.21		fl	0		ml	0.31
4	105	rl	0.41	1	fh	0	1145	mh	0
		rm	0.59		fm	0		mm	0
		rh	0		fl	1		ml	1
5	109	rl	0.39	2	fh	0	1361	mh	0
		rm	0.61		fm	0.07		mm	0.02
		rh	0		fl	0.93		ml	0.98
6	282	rl	0	7	fh	0	26787	mh	0
		rm	0.37		fm	0.43		mm	0.56
		rh	0.63		fl	0.57		ml	0.44
7	258	rl	0	10	fh	0	18753	mh	0
		rm	0.51		fm	0.64		mm	0.38
		rh	0.49		fl	0.36		ml	0.62
8	16	rl	0.94	1	fh	0	1145	mh	0
		rm	0.06		fm	0		mm	0
		rh	0		fl	1		ml	1
9	150	rl	0.15	2	fh	0	1373	mh	0
		rm	0.85		fm	0.07		mm	0.05
		rh	0		fl	0.93		ml	0.95
10	313	rl	0	1	fh	0	1145	mh	0
		rm	0.19		fm	0		mm	0
		rh	0.81		fl	1		ml	1
11	287	rl	0	4	fh	0	3995	mh	0
		rm	0.34		fm	0.21		mm	0.06
		rh	0.66		fl	0.79		ml	0.94
12	317	rl	0	9	fh	0	19762	mh	0
		rm	0.16		fm	0.57		mm	0.41
		rh	0.84		fl	0.43		ml	0.59
13	345	rl	0	6	fh	0	3576	mh	0
		rm	0		fm	0.36		mm	0.05

		rh	1		fl	0.64		ml	0.95
14	132	rl	0.25	11	fh	0	20367	mh	0
		rm	0.75		fm	0.71		mm	0.42
		rh	0		fl	0.29		ml	0.58
15	191	rl	0	2	fh	0	2065	mh	0
		rm	0.91		fm	0.07		mm	0.02
		rh	0.09		fl	0.93		ml	0.98
16	19	rl	0.92	4	fh	0	12139	mh	0
		rm	0.08		fm	0.21		mm	0.24
		rh	0		fl	0.79		ml	0.76
17	108	rl	0.39	5	fh	0	10139	mh	0
		rm	0.61		fm	0.28		mm	0.19
		rh	0		fl	0.72		ml	0.91
18	226	rl	0	9	fh	0	20536	mh	0
		rm	0.7		fm	0.57		mm	0.42
		rh	0.3		fl	0.43		ml	0.58
19	235	rl	0	15	fh	0	32063	mh	0
		rm	0.65		fm	1		mm	0.68
		rh	0.35		fl	0		ml	0.32
20	96	rl	0.46	13	fh	0	25745	mh	0
		rm	0.54		fm	0.86		mm	0.54
		rh	0		fl	0.14		ml	0.46
21	101	rl	0.44	30	fh	1	92685	mh	1
		rm	0.56		fm	0		mm	0
		rh	0		fl	0		ml	0
22	302	rl	0	7	fh	0	3740	mh	0
		rm	0.25		fm	0.43		mm	0.06
		rh	0.75		fl	0.57		ml	0.94
23	300	rl	0	13	fh	0	31722	mh	0
		rm	0.26		fm	0.86		mm	0.67
		rh	0.74		fl	0.14		ml	0.33
24	214	rl	0	9	fh	0	18587	mh	0
		rm	0.77		fm	0.57		mm	0.38
		rh	0.23		fl	0.43		ml	0.62
25	331	rl	0	12	fh	0	32872	mh	0
		rm	0.08		fm	0.79		mm	0.69

		rh	0.92		fl	0.21		ml	0.31
26	82	rl	0.55	16	fh	0.07	23288	mh	0
		rm	0.45		fm	0.93		mm	0.48
		rh	0		fl	0		ml	0.52
27	253	rl	0	10	fh	0	19025	mh	0
		rm	0.54		fm	0.64		mm	0.39
		rh	0.46		fl	0.36		ml	0.61
28	88	rl	0.51	5	fh	0	12520	mh	0
		rm	0.49		fm	0.28		mm	0.25
		rh	0		fl	0.72		ml	0.75
29	194	rl	0	3	fh	0	7461	mh	0
		rm	0.89		fm	0.14		mm	0.14
		rh	0.11		fl	0.86		ml	0.86
30	38	rl	0.81	9	fh	0	14466	mh	0
		rm	0.19		fm	0.57		mm	0.29
		rh	0		fl	0.43		ml	0.71

計算每一筆紀錄在 K-itemsets 的歸屬度值，算出 K-itemsets 的模糊支持度至資料方塊所對應之 Subcube 中：

如 1-itemsets 的計算就以 R 為例。就第一筆而言， $\mu_1^R(256)=0$ ，其餘 29 筆的歸屬度分別為 0，0，0，0.41，0.39，...，0，0，如此可求得 R.L 的模糊支持度值為 0.2073，存入 Data cube (1, 0, 0) 之中，所有的 1-itemsets 的模糊支持度值如下表 4-3 所示：

表 4-3：1-itemsets RFM 資料方塊

	1	2	3
R	0.2073	0.4867	0.306

	1	2	3
F	0.0523	0.4253	0.5223

	1	2	3
M	0.0333	0.2847	0.6853

而 2-itemsets 的計算就以 R.LxF.H 為例。就第一筆而言， $\mu_{1 \times 1}^{R \times F}(256,17) = \mu_1^R(256) \times \mu_1^F(17) = 0 \times 0.14 = 0$ ，其餘 29 筆的歸屬度分別為 0，0，0，...，0.44，...，0，0，如此可求得 R.LxF.L 的模糊支持度值為 0.016，

存入 Data cube (1,1,0) 之中。所有的 2-itemsets 的模糊支持度值如下表 4-4 所示：

表 4-4：2-itemsets RFM 資料方塊

M = 0				F = 0				R = 0			
F	1	2	3	M	1	2	3	M	1	2	3
R				R				F			
1	0.016	0.068	0.124	1	0.015	0.043	0.151	1	0.033	0.012	0.007
2	0.032	0.256	0.256	2	0.019	0.334	0.334	2	0.000	0.228	0.228
3	0.005	0.159	0.142	3	0.000	0.105	0.201	3	0.000	0.101	0.443

依此類推計算出 3-itemsets 的模糊支持度值，如下表 4-5 所示：

表 4-5：3-itemsets RFM 資料方塊

R = 1				R = 2				R = 3			
M	1	2	3	M	1	2	3	M	1	2	3
F				F				F			
1	0.014	0.0006	0.0007	1	0.018	0.0082	0.005	1	0.0028	0.002	0.000
2	0.000	0.026	0.042	2	0.000	0.091	0.109	2	0.073	0.085	0.000
3	0.000	0.017	0.108	3	0.000	0.037	0.220	3	0.000	0.029	0.113

1. FUZZY RFM Association Rules 的挖掘：

因為所有 itemsets 的模糊支持度值已經存在資料方塊中，而且 Data cube 並不大，所以我們可以快速掃描資料方塊並與門檻值（最小模糊支持度）比對來產生大項目組（K-itemsets）再由大項目組推導出關聯法則，並與門檻值（最小模糊信賴度）比對產生符合使用者的法則。假定

在 RFM 上分割之模糊集合的個數分別為 K_R, K_F , 與 K_M ; 使用者制定的門檻值分別為: 最小模糊支持度 \minfs 與最小模糊信賴度 \minfc , 則挖掘 Fuzzy RFM Association Rule 的步驟與演算法如下所示:

步驟 1 掃描 3 維 RFM 資料方塊中, 對應 1-itemsets 的 Subcubes, 若其儲存的模糊支持度大於 \minfs , 則加入 Large RFM 1-itemsets 的集合 L_1 。

經過最小模糊支持度門檻值 (0.30) 的比對後得到 L_1 , 如下表 4-6:

表 4-6: RFM 高頻 1- itemsets (L_1)

	R.M	R.H	F.M	F.L	M.L
FS	0.4867	0.306	0.4253	0.5223	0.6853

步驟 2 產生高頻 RFM 2-itemsets, 有兩種方式可用:

- A. 合併 L_1 產生 C_2 , 再到 Data cube 中檢查 C_2 中每一個候選 2-itemsets 的模糊支持度, 若大於 \minfs , 則加入集合 L_2 中。
- B. 直接掃描 Data cube 中, 對應 2-itemsets 的 Subcubes 之模糊支持度, 若大於 \minfs , 則加入 L_2 中。

假設進行合併以產生 RFM 候選 2-itemsets 的集合 C_2 , 其對應的模糊支持度列表 4-7 如下:

表 4-7: C_2 的 FS

	R.MxF.M	R.MxF.L	R.MxM.L	R.HxF.M
FS	0.256	0.256	0.151	0.159
	R.HxF.L	R.HxM.L	F.MxM.L	F.LxM.L
FS	0.142	0.201	0.228	0.443

若按照之前設定的門檻值 (0.3) 來比對, 則只有一個法則產

生，因此，為了方便說明，我們在此階段調整門檻值，假設為 0.25，則產生了 L_2 ，如下表 4-8

表 4-8： L_2 的 FS

	R.MxF.M	R.MxF.L	F.LxM.L
FS	0.256	0.256	0.443

步驟 3 同理可產生 L_3 R.MxF.LxM.L=0.220

步驟 4 由 L_1 、 L_2 與 L_3 產生模糊 RFM 關聯法則。若 $FC(R) \geq \minfc$ ，則 R 為所要之有效的關聯法則。

接下來，從這些高頻項目組可產生 6 條規則，其中模糊信賴度的計算，以” R is 短 => F is 少 ” 為例，計算如下：

$$\frac{FS(A_1^R \times A_1^F)}{FS(A_1^R)} = \frac{0.256}{0.4867} = 0.526$$

依此計算出其他法則的結果，如下表 3-21 所示：

表 4-9： RFM 關聯法則表

法則	FC
R is 普通=> F is 普通	0.526
R is 普通=> F is 少	0.526
F is 普通 => R is 普通	0.601
F is 少=> M is 少	0.848
F is 少=> R is 普通	0.490
M is 少=> F is 少	0.646

計算這 6 條法則的信賴度是否符合所制定的最小模糊信賴度(0.6)，

如下表 4-10：

表 4-10 RFM 關聯法則表（符合最小信賴度）

法則	FC
F is 普通 => R is 普通	0.601
F is 少=> M is 少	0.848
M is 少=> F is 少	0.646

總共產生了 3 條有效的法則，供行銷經理人去作分析並且提供制定決策的依據：

1. 以顧客行為的角度來分析可以發現最近購買時間都不會太短而且購買次數少或普通，這表示此企業的顧客大部分都是老顧客居多，因此作為企業的行銷經理在做行銷決策時，重心應該擺在如何讓這些老顧客有很好的購物經驗或加強廣告行銷以提升顧客的購買次數來採購或採用企業的產品或服務。
2. 以產品的特性角度來分析，此企業有可能是較專業且單價不會太高的產品，像專業的書籍等。若以專業書籍為例，行銷經理應該對顧客所閱讀的專業領域來分群，然後以 e-mail 或郵寄的方式來宣傳符合顧客所讀的專業領域的最新書籍資訊來刺激顧客購買次數與金額。
3. 以市場的角度來分析，此企業所屬的產業正面臨經濟的不景氣而企業營運欠佳。若是如此，企業經營者可能要以促銷方式或其他方法來刺激買氣。

RFM 的關聯法則在行銷的策略與行銷分析上除了上述之外，其實還可以有相當多的方面可以進行探討，因此探討 RFM 三者的關聯是企業的行銷經理在做決策前相當重要的參考的依據。

例 2. 交易資料庫 FUZZY Association Rules 的挖掘：

假設一顧客交易資料表如表 4-11、商品項目分類樹如圖 4-2 及

商品項目的歸屬函數如圖 4-3：

表 4-11 顧客交易資料表

顧客編碼	商品項目及其在某段時間內被顧客購買的件數
1	(A,2)(B,2)(C,4)(D,2)(E,2)(F,7)
2	(A,1)
3	(A,1)(C,6)(D,4)(E,6)(F,3)
4	(A,1)
5	(A,1)(C,1)
6	(A,1)(B,1)(D,3)(E,1)(F,1)
7	(A,1)(B,1)(C,3)(D,4)(E,1)
8	(A,1)
9	(A,1)(F,1)
10	(A,1)
11	(A,1)(B,1)(E,1)(F,1)
12	(A,1)(C,4)(D,1)(F,3)
13	(A,1)(B,1)(D,2)(E,1)(F,1)
14	(A,2)(B,1)(C,3)(D,1)(E,2)(F,2)
15	(A,1)(C,1)
16	(A,1)(C,1)(D,1)(E,1)
17	(D,3)(F,2)
18	(A,1)(C,2)(D,1)(E,3)(F,2)
19	(C,4)(D,7)(E,3)(F,1)
20	(C,6)(D,2)(E,4)(F,1)
21	(B,2)(C,11)(D,5)(E,6)(F,6)
22	(C,3)(D,1)(E,2)(F,1)
23	(A,1)(C,1)(D,5)(E,3)(F,3)
24	(C,4)(D,2)(F,3)
25	(C,1)(D,5)(E,2)(F,1)
26	(B,1)(C,2)(D,6)(E,3)(F,4)
27	(C,4)(D,2)(E,1)(F,3)

28	(B,1)(D,2)(E,1)(F,2)
29	(D,1)(F,2)
30	(C,2)(D,2)(E,4)(F,1)

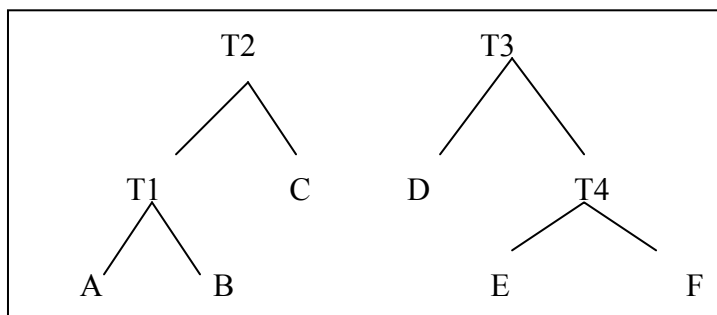


圖 4-2 商品項目分類樹

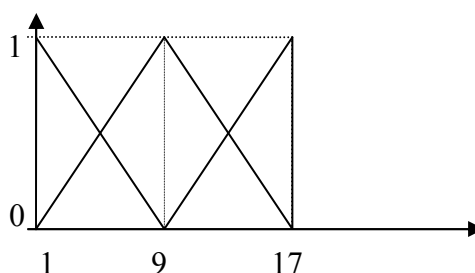


圖 4-3 商品項目的歸屬函數

步驟 1：

掃瞄資料庫以產生叢聚式模糊集表 (cluster-based fuzzy sets tables)。首先依分類樹 (taxonomy) 的分類，將項目的“祖先”加入，形成含祖先的擴充型交易項目集。交易項目數一樣多的就形成一個聚落，再將含祖先的擴充型交易項目的數值型資料代入圖 4-3 商品項目的歸屬函數，計算其模糊值，並儲存在如表 4-12 叢聚式模糊集表對應的位

置。例如交易 2 僅有 (A, 1)，但同時屬於類別 T1、T2，則其擴充型交易項目集為 { (A, 1), (T1, 1), (T2, 1) }，故其屬於叢聚式模糊集 (3)；又計算模糊值，得到 {L1, L1, L1}，再儲存在表 4-12 叢聚式模糊集表對應的位置。

表 4-12 叢聚式模糊集表

交易	A	B	C	D	E	F	T1	T2	T3	T4
----	---	---	---	---	---	---	----	----	----	----

叢聚式模糊集 (1)

null

叢聚式模糊集 (2)

null

叢聚式模糊集 (3)

2	L,1						L,1	L,1		
4	L,1						L,1	L,1		
8	L,1						L,1	L,1		
10	L,1						L,1	L,1		

叢聚式模糊集 (4)

5	L,1		L,1				L,1	L,0.875 M,0.125		
15	L,1		L,1				L,1	L,0.875 M,0.125		
17				L,0.75 M,0.25		L,0.875 M,0.1			L,0.5 M,0.5	L,0.875 M,0.125

						25				
29				L,1		L,0.8 75 M,0.1 25			L,0.75 M,0.25	L,0.875 M,0.125

叢聚式模糊集 (5)

null

叢聚式模糊集 (6)

9	L,1					L,1	L,1	L,1	L,1	L,1
24			L,0.625 M,0.375	L,0.875 M,0.125		L,0.75 M,0.25		L,0.625 M,0.375	L,0.5 M,0.5	L,0.75 M,0.25

叢聚式模糊集 (7)

19			L,0.625 M,0.375	L0.25 M0.75	L0.75 M0.25	L1		L,0.625 M,0.375	M,0.75 H,0.25	L,0.625 M,0.375
20			L,0.375 M,0.625	L,0.875 M,0.125	L,0.625 M,0.375	L,1		L,0.375 M,0.625	L,0.25 M,0.75	L,0.5 M,0.5
22			L,0.75 M,0.25	L,1	L,0.875 M,0.125	L,1		L,1	L,0.125 M,0.875	L,0.75 M,0.25
25			L,1	L,0.5 M,0.5	L,0.875 M,0.125	L,1		L,0.75 M,0.25	L,0.625 M,0.375	L,0.75 M,0.25
27			L,0.625 M,0.375	L,0.875 M,0.125	L,1	L,0.75 M,0.25		L0.625 M0.375	L0.25 M0.75	L,0.5 M,0.5
30			L,0.875 M,0.125	L,0.875 M,0.125	L,0.625 M,0.375	L,1		L,0.875 M,0.125	L,0.375 M,0.625	L,0.375 M,0.625

叢聚式模糊集 (8)

11	L,1	L,1			L,1	L,1	L,0.875 M,0.125	L,0.875 M,0.125	L,0.875 M,0.125	L,0.875 M,0.125
12	L,1		L,0.625	L,1		L,0.75	L,1	L,0.5	L,0.625	L,0.75

			M,0.375			M,0.25		M,0.5	M,0.375	M,0.25
16	L,1		L,1	L,1	L,1		L,1	L,0.875 M,0.125	L,0.875 M,0.125	L,1
28		L,1		L,0.875 M,0.125	L,1	L,0.875 M,0.125	L,1	L,1	L,0.5 M,0.5	L,0.75 M,0.25

叢聚式模糊集 (9)

3			L,0.375 M,0.625	L,0.625 M,0.375	L,0.375 M,0.625	L,0.75 M,0.25	L,0.25 M,0.25		L,0.5 M,0.5	M,1
6		L,1		M,0.75 H,0.25	L,1	L,1	L,0.875 M,0.125	L,0.875 M,0.125	L,0.5 M,0.5	L,0.875 M,0.125
7	L,1	L,1	L,0.75 M,0.25	L,0.625 M,0.375	L,1		L,0.875 M,0.125	L,0.5 M,0.5	L,0.5 M,0.5	L,1
13	L,1	L,1		L,0.875 M,0.125	L,1	L,1	L,0.875 M,0.125	L,0.875 M,0.125	L,0.625 M,0.375	L,0.875 M,0.125
18	L,1		L,0.875 M,0.125	L,1	L,0.75 M,0.25	L,0.875 M,0.125	L,1	L,0.75 M,0.25	L,0.375 M,0.625	L,0.5 M,0.5
21		L,0.875 M,0.125	M,0.75 H,0.25	L,0.5 M,0.5	L,0.375 M,0.625	L,0.375 M,0.625	L,0.875 M,0.125	M,0.5 H,0.5	H,1	M,0.625 H,0.375
23	L,1		L,1	L,0.5 M,0.5	L,0.75 M,0.25	L,0.75 M,0.25	L,1	L,0.875 M,0.125	M,0.75 H,0.27	L,0.375 M,0.625
26		L,1	L,0.875 M,0.125	L,0.375 M,0.625	L,0.75 M,0.25	L,0.625 M,0.375	L,1	L,0.75 M,0.25	M,0.5 H,0.5	L,0.25 M,0.75

叢聚式模糊集 (10)

1	L,0.875 M,0.125	L,0.875 M,0.125	L,0.625 M,0.375	L,0.875 M,0.125	L,0.875 M,0.125	L,0.25 M,0.75	L,0.625 M,0.375	L,0.125 M,0.875	L,0.75 M,0.25	M,1
14	L,0.875 M,0.125	L,1	L,0.75 M,0.25	L,1	L,0.875 M,0.125	L,0.875 M,0.125	L,0.75 M,0.25	L,0.375 M,0.625	L,0.5 M,0.5	L,0.625 M,0.375

步驟二：

由 L_{n-1} 的項目集自我結合 (self-joint) 成候選項目集 C_n 長度為 k 的候選項目集，其支持度是由計算叢聚式模糊集表 (k) 至叢聚式模糊集表 (M) 模糊值的和所得。假如 C_n 的支持度大於等於最小支持度 α ，候選項目集就變成大項目集，將 C_n 置入大項目集 L_n 。否則與下一個叢聚式模糊集表 ($k+1$) 做比較。當大於等於最小支持度或到達叢聚式模糊集 (M) 的盡頭時，則停止做比較的動作。

我們整理了叢聚式模糊集表表 4-12，設定最小支持度 15。得到 $L_1 = \{A. Low(17.75), D. Low(17), E. Low(15.5), F. Low(18.38), T1. Low(19.75), T2. Low(20.25), T4. Low(15.13)\}$ 及 $L_2 = \{(A. Low, T1. Low)(16.88)\}$

步驟三：

利用最小信賴度找出模糊關聯法則。

假如候選模糊關聯法則的信賴度大於等於最小信賴度就將其放入規則資料庫。 $A.Low \Rightarrow T1.Low$ 的信賴度為 $\frac{16.88}{17.75} = 0.94$ ， $T1.Low \Rightarrow A.Low$ 的信賴度為 $\frac{16.88}{19.75} = 0.81$ 。如果最小信賴度 = 0.9，就將 $A.Low \Rightarrow T1.Low$ 放入規則資料庫。

例 3. 顧客 RFM 及交易資料庫間的模糊關聯法則的挖掘

模糊規則挖掘法可分解成四個步驟：

1. 給定一個最小支持度 (0.02)，我們可掃描模糊 RFM data cube 過濾出三維的模糊大項目集 $LRFM_i$ ，而此 $LRFM_i$ 對應的顧客集可記為 $LRFM_i_CID$ 。如我們在表 4-2 顧客 RFM 資料表對應的模糊值表中 (1, 2, 2) 的模糊支持度為 0.026 (30 筆資料之 (1, 2, 2) 模糊值的算術平均數)，且顧客 14, 16, 17, 20, 26, 28, 30 之 (1, 2, 2) 模糊值分別為 0.075, 0.046, 0.021, 0.214, 0.246, 0.036, 0.134，所以，(1, 2, 2) 的 $LRFM_i_CID = \{14, 16, 17, 20, 26, 28, 30\}$ 。

表 4-13：資料方塊對應的顧客表

$LRFM_i$	對應的顧客集 $LRFM_i_CID$ (最小支持度=0.02)	對應的顧客集 $LRFM_i_CID$ (最小支持度=0.04)
(1,2,2)	14,16,17,20,26,28,30	14, 16,17,20,26,30
(1,2,3)	5,14, 16,17,20,26,28,30	14, 16,17,20,26,28,30
(1,3,3)	4,8,9,14,16,17,20,28,30	4,8,9,14,16,17,28,30
(2,2,2)	1,3,6,7,12,14,17,18,19,20,23,24, 25,26,27,28,30	1,3,7,14,18,19,20,23,24,25,26,27
(2,2,3)	1,3,5,6,7,9,11,12,14,15,17,18,19, 20,22,23,24,25,26,27,28,29,30	1,3, 5,6,7,9,11,12,14,15,17,18,19, 20,22,23,24,26,27,28,29,30
(2,3,2)	4,6,7,9,12,14,15,17,18,20,23,24, 27,28,29,30,	4,6,7,9,14,15,17,18,20,24,27,28,29
(2,3,3)	2,4,5,6,7,8,9,10,11,12,14,15,16, 17,18,20,22,24,27,28,29,30	2,4,5,6,7,8,9,10,11,12,14,15,16,17,18,2 2,24,27,28,29,30
(3,2,2)	1,3,6,7,12,19,23,24,25,27	1,3,6,7,12,19,23,24,25,27
(3,2,3)	1,3,6,7,11,12,13,18,19,22,23,24,	1,3,6,7,11,12,13,18,19,22,23,24,

	25,27	25,27
(3,3,2)	6,7,11,12,13,18,22,23,24,25,27	6,7,12,18,23,25,27
(3,3,3)	2,6,7,10,11,12,13,15,18,22,23,24,25,27,29	2,6,7,10,11,12,13,15,18,22,24,25,27,29

2. 給定一個最小支持度，吾人可掃描叢聚式模糊集過濾出每個大模糊項目集 L_j ，而此對應的顧客集可記為 L_{j_CID} 。

表 4-14 大項目集及其對應的顧客群

L_j	模糊支持度	大項目集對應的顧客 L_{j_CID}
A.Low	17.75	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,18,23
D.Low	17	1,3,6,7,12,13,14,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30
E.Low	15.5	1,3,6,7,11,13,14,16,18,19,20,21,22,23,25,26,27,28,30
F.Low	18.38	1,3,6,9,11,12,13,14,17,18,19,20,21,22,23,24,25,26,27,28,29,30
T1.Low	19.75	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,18,21,23,26,28
T2.Low	20.25	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,18,19,20,22,23,24,25,26,27,28,30
T4.Low	15.13	6,7,9,11,12,13,14,16,17,18,19,20,22,23,24,25,26,27,28,29,30

3. 由 $LRFM_i$ 及 L_j 所組成的關係矩陣 R ，每個元素 $R(RFM_i, L_j)$ 儲存 $LRFM_i$ 及 L_j 的信賴度。假如 $(LRFM_i_CID\ sets) - (L_j_CID\ sets) = \phi$ 則它的信賴度 $\beta = 1$ ；否則，信賴度 $\beta = n((LRFM_i \cap L_j)_CID\ sets) / n(LRFM_i_CID\ sets)$ ，其 $n(S)$ 表集合 S 的大小。

由表 4-13：資料方塊對應的顧客表及表 4-14 大項目集及其對應的顧客群我們可以步驟 3 的方法，計算出 $LRFM_i$ 及 L_j 的信賴度並將之整理成表 4-15。如 (1) $R. L\&F. M\&M. M-D. Low = \phi$ ，信賴度 $\beta = 1$ (2) $R. L\&F. M\&M. M - E. Low = \{17\}$ ，信賴度 $\beta = 6/7$ ，...

表 4-15 $LRFM_i$ 及 L_j 所組成的關係矩陣 R

信賴度 β	A.Low	D.Low	E.Low	F.Low	T1.Low	T2.Low	T4.Low
(1,2,2)	2/7	1	6/7	6/7	4/7	6/7	1
(1,2,3)	3/8	7/8	6/8	6/8	5/8	7/8	7/8
(1,3,3)	5/9	6/9	5/9	6/9	5/9	7/9	7/9
(2,2,2)	8/17	1	14/17	16/17	10/17	16/17	14/17
(2,2,3)	12/23	19/23	16/23	20/23	13/23	21/23	19/23
(2,3,2)	9/16	13/16	9/16	13/16	10/16	14/16	15/16
(2,3,3)	14/22	14/22	11/22	14/22	15/22	20/22	15/22
(3,2,2)	6/10	1	8/10	9/10	6/10	1	8/10

(3,2,3)	<i>9/14</i>	<i>13/14</i>	<i>12/14</i>	<i>13/14</i>	<i>10/14</i>	<i>1</i>	<i>12/14</i>
(3,3,2)	<i>7/11</i>	<i>10/11</i>	<i>9/11</i>	<i>10/11</i>	<i>7/11</i>	<i>1</i>	<i>1</i>
(3,3,3)	<i>10/15</i>	<i>11/15</i>	<i>9/15</i>	<i>12/15</i>	<i>10/15</i>	<i>14/15</i>	<i>12/15</i>

4. 假設 $(LRFM_i, L_j)$ 的信賴度大於等於最小信賴度，這關係可轉換成型如 If-then 的規則。如在表 4-16 中+或++的欄位其信賴度大於等於最小信賴度，可轉換成模糊 RFM 大項目與模糊大項目間的關聯法則。如：
 $RL \& FM \& MM \quad D.Low ; RL \& FM \& MM \quad T4.Low$ 。

表 4-16 模糊 RFM 大項目與模糊大項目的關聯法則表

信賴度 β	A.Low	D.Low	E.Low	F.Low	T1.Low	T2.Low	T4.Low
RL.FM.MM		+	+	+		+	++
RL.FM.ML		+	+	+	+	+	++
RL.FL.ML		+		+		+	++
RM.FM.MM		+	+	+		+	++
RM.FM.ML		+	+	+		+	++
RM.FL.MM		+		+	+	+	++
RM.FL.MML	+	+		+	+	++	++

RH.FM.MM	+	+	+	+	+	++	++
RH.FM.ML	+	+	+	+	+	++	++
RH.FL.MM	+	+	+	+	+	++	++
RH.FL.ML	+	+	+	+	+	++	++

5. 因在分類樹中 A、B 包含於 T1； T1、C 包含於 T2 及 E、F 包含於 T4；
T4、D 包含於 T3，我們將表 4-16 欄位中有++的欄位刪除，剩餘的規則就是所謂的”有趣”的關聯規則(interesting rules)。如表 4-17：

表 4-17 模糊 RFM 大項目與模糊大項目的”有趣”關聯法則表

信賴度 β	A.Low	D.Low	E.Low	F.Low	T1.Low	T2.Low	T4.Low
RL.FM.MM		+	+	+		+	
RL.FM.ML		+	+	+	+	+	
RL.FL.ML		+		+		+	
RM.FM.MM		+	+	+		+	
RM.FM.ML		+	+	+		+	
RM.FL.MM		+		+	+	+	
RM.FL.MML	+	+		+	+		
RH.FM.MM	+	+	+	+	+		
RH.FM.ML	+	+	+	+	+		

RH.FL.MM	+	+	+	+	+		
RH.FL.ML	+	+	+	+	+		

另外在表 4-13，我們調高最小支持度從 0.02 至 0.04 時，RL.FM.MM (= (1, 2, 2)) 的目標客戶由 {14, 16, 17, 20, 26, 28, 30} 減少為 {14, 16, 17, 20, 26, 30}。故若公司的行銷預算減少時，我們只要對主要客戶 14, 16, 17, 20, 26, 30 做相關的促銷活動。

第五章 結論及未來努力方向

隨著資訊科技的快速發展，以及市場競爭激烈的環境下，企業導入相當多的資訊科技來幫助經營者提升其競爭優勢，尤其是顧客關係管理系統。透過顧客關係管理系統，企業能夠深入瞭解顧客的需求，進而針對不同顧客的需求來量身設計產品與服務，來提升客戶的忠誠度，以及企業的獲利能力與競爭力。但如何有效地進行顧客關係管理，是目前企業所關注的焦點，其中，又以資料探勘最為著名。

本研究主要是針對資料探勘中最熱門的研究－關聯法則來進行探討。著名的 Apriori 關聯法則以重複掃描資料庫及結合、刪除，產生階層大項目組 (1-item sets, 2-item sets, 3-item sets, …)，再萃取出大於最小信賴度的關聯法則。為改善重複掃描資料庫的缺點，我們利用模糊 RFM Model，以模糊分割建立模糊資料方塊，只要掃描資料方塊；模糊叢聚式關聯法則，以分類樹將商品項目分類，形成含祖先的擴充型交易項目，再建立叢聚式模糊集表，只要掃描資料庫一次。並找出 (1) 商品間的關聯性 (2) R、F、M 間的關聯性 (3) 顧客交易資料庫之 RFM 資料表與商品銷售紀錄資料庫萃取出此二資料庫的關聯法則。藉著不同的關聯法則，企業可以瞭解顧客的消費行為可以更有效地決定進貨量或庫存量；或是在店裡要如何鋪陳貨品，或用來評估店裡的促銷活動成效。深

入地從產品在市場的特性以及市場狀況進行較為全面的分析，以使企業能夠更有效地達成行銷決策。更可以讓企業行銷部門在公司的預算內做最有效的廣告、促銷行為，以最低的行銷成本達成最好的廣告效益。

參 考 文 獻

一、中文部分：

- [1]曾憲雄、蔡秀滿、蘇東興、曾秋蓉、王慶堯，資料探勘台北：旗標出版社，民國 93 年。
- [2]李允中、王小璠、蘇木春，模糊理論及其應用，台北：全華科技圖書，民國 92 年。
- [3]李健興、賴鼎宇、崔殷豪、黃啟瑞、陳偉昇、郭耀煌，「基於 XML 之智慧型個人化服務實驗性網站」，成功大學/華新麗華數位生活科技研究中心研究報告，民國 89 年。
- [4]武家慶，「適用於挖掘多層次聯結規則之架構」，逢甲大學資訊工程學系碩士論文，民國 90 年
- [5]許智豪，「在動態資料庫中作動態挖掘關聯式法則」，國立中興大學資訊科學研究所碩士論文，民國 89 年。
- [6]楊昇樺、毛立人、邱宏彬，「關聯法則之多層更新挖掘法及其應用」，第三屆產業資訊管理學術暨新興科技實務研討會，pp.549~557，台北，民國 91 年。

二、英文部分：

- [1] A. Kandel, *Fuzzy Expert Systems*, CRC Press, Boca Raton, FL, 1992 pp.8-19.
- [2] C. Kleissner, “Data mining for the enterprise,” *Proc. of the Thirty-First Hawaii International Conference*, Vol. 7, pp. 295-304,1998.
- [3]D.W. Cheung, J. Han, V. Ng, and C.Y. Wang, “A general incremental technique for maintaining discovered association rules,” *Proc. of Conf. Sys.*, pp.185-19,1997.
- [4] H. Ishibuchi, K. Nozaki, and H. Tanaka, “Distributed representation of fuzzy rules and its application to pattern classification”,*Journal of Fuzzy Sets and Systems*,Vol.52(1),pp.21-32,1992.

- [5] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Selecting fuzzy if-then rules for classification problems using genetic algorithms", *IEEE Transaction on Fuzzy Systems*, Vol.3(3), pp.260-270,1995.
- [6] H. Toivonen, "Sampling large databases for association rules," *Proc. of the 22nd Conf. on VLDB*, pp.134-145,1996.
- [7] J.S. Park, M. S. Chen, and P. S. Yu, "An Effective Hash Based Algorithm for Mining Association Rules", *Proceedings of ACM SIGMOD*, 1995 pp.175-186.
- [8] L. A. Zadeh, "Fuzzy sets" ,*Information Control*, Vol.8(3), pp.338-353,1965.
- [9] M. Berry and G. Linoff, "Data Mining Techniques for marketing, sales, and Customer Support," *New York. Wiley Computer Publishing*,1997.
- [10] M. Goebel and L. Gruenwald, "A survey of data mining and knowledge discovery software tools," *ACM SIGKDD*, Val. 1, pp. 20- 33,1999.
- [11] M.S. Chen, J. Han, and P.S. Yu, "Data Mining: An overview from a database perspective," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, pp.866-883,1996.
- [12] M. S. Chen, J. S. Park, and P. S. Yu, "Efficient data mining for path traversal patterns," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 10, No. 2, pp.209-221,1998.
- [13] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," *Proc. of the 20th Conf. on Very Large Databases*, pp.487-499,1994.
- [14] R. Agrawal and R. Srikant, "Mining Sequential Patterns," *Proc. of the 11th Conf. on Data Engineering*, pp.3-14,1995.
- [15] R. Agrawal, R. Srikant, and Q.Vu, "Mining association rules with item constraints," *Proc. 3rd Int. Conf. Knowledge Discovery and Data Mining*, Newport Beach, California,1997.
- [16] R. Srikant and R. Agrawal, "Mining generalized association rules,"*Proc. of the 21st VLDB*, pp. 407-419,1997.
- [17] T.P. Hong, C.S. Kuo, S.C. Chi, "A data mining algorithm for transaction data with quantitative values", *Intelligence Data Analysis*, 1999 pp.363-376.
- [18] T.P. Hong, K.Y. Lin, S.L. Wang, "Fuzzy data mining for interesting generalized association rules", *Fuzzy sets and systems*, 2003 pp.255-269.
- [19] U. M. Fayyad, "Data mining and knowledge discovery: making sense out of

data,”*IEEE Expert*, Vol. 11, No. 5, pp.20-25,1996.

[20] Yen John and Reza Langari, *Fuzzy Logic Intelligence, Control, and Information*, Prentice-Hall, Inc., 1999.

[21] Yuh-Jiuan Tsay, Jiunn-Yann Chiang, “CBAR: an efficient method for mining association rules”, *Knowledge-Based Systems*, 2005 pp.99-105