

南 華 大 學

資訊管理研究所

碩士論文

模糊邏輯與資料探勘技術為基礎在顧客關係管理上之研究與
應用

**Using Fuzzy Logic and Data Mining for Customer
Relationship Management**

研究生：蘇建源

指導教授：邱宏彬

中華民國九十三年六月十四日

南 華 大 學

碩 士 學 位 論 文

資 訊 管 理 學 系

模 糊 邏 輯 與 資 料 探 勘 技 術 在 顧 客 關 係 管 理 之 研 究 與 應 用

研 究 生：蘇 建 源

經 考 試 合 格 特 此 證 明

口 試 委 員：

楊 弘 章

謝 品 霖

指 導 教 授：

邱 宏 彬



所 長：

口 試 日 期：中 華 民 國 九 十 三 年 六 月 十 四 日

南華大學碩士班研究生
論文指導教授推薦函

資訊管理系碩士蘇建源君所提之論文
模糊邏輯與資料探勘技術為基礎在顧客關係
管理上之研究與應用

係由本人指導撰述，同意提付審查。

指導教授

邱宏彬

93年6月24日

誌 謝

輕鴻戲江潭，孤雁集洲沚，邂逅兩相親，緣念共無已。
風雨好東西，一隔頓萬里。追憶栖宿時，聲容滿心耳。
落日川渚寒，愁雲繞天起。短翮不能翔，徘徊煙霧裏。

兩年的研究生涯就如同浩瀚的交響樂即將劃下休止符。在離別時刻，要感謝這兩年來教導過我的資管所所有的老師，由於您們不倦的教誨，開闊我的視野、豐富我的專業領域知識也培養我解決問題的能力。在論文口試期間，感謝口試委員楊弘章老師、謝昆霖老師的建議與指教，使本論文能更臻完善。

我覺得我一直是個幸福的人，能得到眾多老師及朋友的幫忙，尤其是能夠加入 204 大家庭中。在這大家庭中，每天沈溺在溫馨快樂，與老師、師兄弟一起吃飯、玩樂、作研究，邱老師不僅在學術研究上給予很大的幫助，在無所適從的人生處事給予開釋才是受益無窮。204 實驗室是我的第二個家。邱老師是我們的大家長，高弘、阿牧、晉赫、文天、琪雯、揮明、洽堂、耀霖、書萍、熾薰就像是兄弟姊妹一般；怡甄、小貴、美倫、阿信、伊汝、雅雯、玟萃、佳偉、可愛的專題組員靜瑜、琬蓉、文楓、懿慧、建銘以及房東大姊英珍更是在我求學過程中，在心靈、課業及生活上不可或缺的伙伴。最後要感謝我的父母與家人，若不是您們長久以來的支持，不可能有我今天的小小成果。

我以為學如逆水行舟，如今輕舟已過萬重山。一路行來，點滴在心頭，雖然這不是一部完美的論文，但這部論文的完成，要感謝的人真的很多，在此無法一一詳列，僅以此文表達我的誠摯謝意。

蘇建源 僅誌於 嘉義
南華大學資訊管理學系
中華民國九十三年 六月

模糊邏輯與資料探勘技術在顧客關係管理之研究與應用

學生：蘇建源

指導教授：邱宏彬

南華大學資訊管理學系碩士班

摘 要

在現今競爭激烈與變化快速的環境下，瞭解與滿足顧客的需求是企業獲利的關鍵因素。因此，顧客關係管理就成為當今企業非常重視的議題。利用資料探勘技術可以幫助企業從龐大且複雜的顧客資料庫中找出隱含的資訊，其中又以關聯法則挖掘法最為著名。如何有效地推導出關聯法則，在過去已經有許多方法相繼被提出，但大部分的演算法對於(a)數值型資料的處理、(b)漸進式挖掘以及(c)線上挖掘等問題無法有效地同時加以處理。因此本研究提出以模糊切割與資料方塊為基礎之關聯法則演算法來建構單層次以及多階層等關聯法則演算法。並應用於顧客關係管理上的探討，來幫助企業在制定客製化行銷策略的決策上的一個有力的參考依據。

關鍵字：顧客關係管理、資料探勘、關聯法則、漸進式挖掘、線上挖掘、模糊切割、資料方塊、單層次關聯法則、多階層關聯法則

Using Fuzzy Logic and Data Mining for Customer Relationship Management

Student : Chien-Yuan Su

Advisor : Dr. Hung-Pin Chiu

Department of Information Management

The M.B.A Program

Nan-Hua University

ABSTRACT

In the era of great competition, understanding and satisfying customers' requirements are the critical tasks for a company to make a profit such that customer relationship management becomes the important business issue at present. With the help of the data mining techniques, the manager can explore and analyze from a great quantity of data to discover meaningful patterns and rules. Mining association rules from transaction databases is most commonly seen in data mining. However, most conventional algorithms can not simultaneously and effectively satisfy the following requirements: (a) the relationships among transactions with numeric values, (b) incremental mining, and (c) on-line mining. In this thesis, we integrate the data cube and fuzzy partition techniques to propose a single-level association rule miner and a multi-level association rule miner. This mined knowledge can be applied in customer relationship management to help decision maker make correct business decisions for marketing strategies.

Keyword : Customer Relationship Management 、 Data Mining 、 Association Rules 、 Incremental Mining 、 On-Line Mining 、 Fuzzy Partition 、 Data Cube 、 Single-Level Association Rule 、 Multi-Level Association Rule

目 錄

書名頁	I
論文指導教授推薦書	II
博碩士論文授權書	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	XI
第一章 緒論	1
第一節 研究背景	1
第二節 研究動機與目的	2
第三節 研究方法	4
第四節 研究貢獻	5
第五節 論文架構	6
第二章 文獻探討	7
第一節 資料探勘	7
第二節 關聯法則挖掘法之相關研究	13
第三節 模糊邏輯與模糊切割	26
第三章 模糊切割在關聯法則之研究與應用	34
第一節 模糊切割於單層次關聯法則之建構	34
第二節 傳統 RFM Model	42
第三節 Fuzzy RFM Model	47
第四節 實驗與探討	61
第五節 結論	86
第四章 模糊切割在多階層關聯法則之研究	88
第一節 模糊切割於多階層關聯法則之建構	88
第二節 模糊切割之漸進式與線上多階層關聯法則	95
第三節 討論與結論	104
第五章 結論與未來展望	105
參考文獻	108

表 目 錄

表 2-1	資料探勘的演化步驟.....	7
表 2-2	食物項目分類以 Taxonomy 為例.....	21
表 2-3	多階層關聯法則演算法比較表.....	26
表 2-4	範例資料表.....	31
表 3-1	交易資料表.....	37
表 3-2	交易資料轉換表.....	38
表 3-3	長度為 1 之候選項目.....	38
表 3-4	A.L*B.L 範例表.....	39
表 3-5	長度為 2 之候選項目.....	39
表 3-6	Large itemsets.....	40
表 3-7	(B.L, D.H) 為例之候選法則.....	41
表 3-8	符合最小模糊信賴度之法則.....	41
表 3-9	傳統 RFM Model 的缺點比較表.....	44
表 3-10	顧客 RFM 資料表.....	50
表 3-11	1-itemsets RFM 資料方塊.....	51
表 3-12	2-itemsets RFM 資料方塊.....	51
表 3-13	3-itemsets RFM 資料方塊.....	52
表 3-14	顧客分群表.....	52
表 3-15	RFM 權重表.....	55
表 3-16	RFM 1-itemsets 的 FS.....	57
表 3-17	RFM 高頻 1- itemsets (L_1).....	57
表 3-18	C_2 的 FS.....	58
表 3-19	L_2 的 FS.....	58
表 3-20	C_3 的 FS.....	58
表 3-21	RFM 關聯法則表.....	59
表 3-22	RFM 關聯法則表(符合最小信賴度).....	59
表 3-23	α -cut 值對於不同顧客擁有群數之總人數影響 (1997 年 RFM 資料表).....	64
表 3-24	α -cut 值對於不同顧客擁有群數之總人數影響 (1998 年 RFM 資料表).....	64
表 3-25	1997 年顧客五等分 (第 1 群) 在 Fuzzy RFM Model 分群上之狀態.....	68
表 3-26	1998 年顧客五等分 (第 1 群) 在 Fuzzy RFM Model 分群上之狀態.....	68
表 3-27	1997 年行為五等分 (第 51 群) 在 Fuzzy RFM Model 分	

群上之狀態.....	68
表 3-28 1998 年行為五等分 (第 56 群) 在 Fuzzy RFM Model 分 群上之狀態.....	69
表 3-29 1997 年行為五等分 (第 56 群) 在 Fuzzy RFM Model 分 群上之狀態.....	69
表 3-30 1998 年行為五等分 (第 51 群) 在 Fuzzy RFM Model 分 群上之狀態.....	69
表 3-31 1997 年顧客五等分各個分數之顧客人數表.....	72
表 3-32 1998 年顧客五等分各個分數之顧客人數表.....	72
表 3-33 1997 年行為五等分各個分數之顧客人數表.....	73
表 3-34 1998 年行為五等分各個分數之顧客人數表.....	73
表 3-35 RFM Model 分數綜合比較表.....	75
表 3-36 同分且意義不同之發生次數比較表.....	82
表 3-37 1997 年顧客五等分分數之不同組合人數.....	83
表 3-38 1998 年顧客五等分分數之不同組合人數.....	84
表 3-39 1997 年行為五等分分數之不同組合人數.....	84
表 3-40 1998 年行為五等分分數之不同組合人數.....	85
表 3-41 Stone RFM 分數之不同組合人數.....	85
表 4-1 交易資料表.....	90
表 4-2 含有祖先項目的交易表.....	91
表 4-3 長度為 1 之項目.....	92
表 4-4 A.L*B.L 範例表.....	93
表 4-5 長度為 2 之項目.....	93
表 4-6 Large 1-itemsets.....	94
表 4-7 Large 2-itemsets.....	94
表 4-8 候選法則.....	95
表 4-9 符合最小模糊信賴度之法則.....	95
表 4-10 新增資料表.....	96
表 4-11 新增部分的支持度.....	96
表 4-12 與第 5 筆資料相關之項目的歸屬值.....	97
表 4-13 資料筆數漸增示意表.....	101
表 4-14 動態調整門檻值示意表.....	103

圖 目 錄

圖 1-1	研究流程圖.....	4
圖 2-1	知識挖掘過程圖.....	10
圖 2-2	產生大項目組織範例.....	15
圖 2-3	食物項目分類以 Taxonomy 為例.....	20
圖 2-4	ML_T2L1 演算法.....	22
圖 2-5	Cumulate 演算法.....	24
圖 2-6	三角形歸屬度函數.....	29
圖 2-7	以三角形歸屬度函數建立的 Simple fuzzy partition.....	29
圖 3-1	模糊切割之單層次關聯法則流程圖.....	37
圖 3-2	模糊模糊歸屬函數.....	37
圖 3-3	模糊 RFM Model 流程架構圖.....	47
圖 3-4	RFM 3 維資料方塊.....	48
圖 3-5	RFM 三角形歸屬度函數.....	50
圖 3-6	RFM Model Systems 與 RFM 資料表.....	62
圖 3-7	1997 年顧客五等分之每一群之顧客人數.....	66
圖 3-8	1998 年顧客五等分之每一群之顧客人數.....	66
圖 3-9	1997 年行為五等分之每一群之顧客人數.....	66
圖 3-10	1998 年行為五等分之每一群之顧客人數.....	67
圖 3-11	1997 年顧客五等分各個分數之顧客人數狀況圖.....	72
圖 3-12	1998 年顧客五等分各個分數之顧客人數狀況圖.....	72
圖 3-13	1997 年顧客五等分各個分數之顧客人數狀況圖.....	73
圖 3-14	1998 年顧客五等分各個分數之顧客人數狀況圖.....	73
圖 3-15	1997 Stone RFM 各個分數之顧客人數狀況圖.....	74
圖 3-16	1998 Stone RFM 各個分數之顧客人數狀況圖.....	74
圖 3-17	1997 Fuzzy RFM 各個分數之顧客人數狀況圖.....	74
圖 3-18	1998 Fuzzy RFM 各個分數之顧客人數狀況圖.....	75
圖 3-19	1997 年 Stone RFM 次多人數之分數之顧客其各 R 值人數 直條圖.....	77
圖 3-20	1997 年 Stone RFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 R 值人數直條圖.....	77
圖 3-21	1997 年 Stone RFM 次多人數之分數之顧客其各 F 值人數 直條圖.....	78
圖 3-22	1997 年 Stone RFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 F 值人數直條圖.....	78
圖 3-23	1997 年 Stone RFM 次多人數之分數之顧客其各 M 值人數	

直條圖.....	78
圖 3-24 1997 年 Stone RFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 M 值人數直條圖.....	79
圖 3-25 1998 年 Stone RFM 次多人數之分數之顧客其各 R 值人數直條圖.....	79
圖 3-26 1998 年 Stone RFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 R 值人數直條圖.....	79
圖 3-27 1998 年 Stone RFM 次多人數之分數之顧客其各 F 值人數直條圖.....	80
圖 3-28 1998 年 Stone RFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 F 值人數直條圖.....	80
圖 3-29 1998 年 Stone RFM 次多人數之分數之顧客其各 M 值人數直條圖.....	80
圖 3-30 1998 年 Stone RFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 M 值人數直條圖.....	81
圖 4-1 模糊切割之多階層關聯法則流程圖.....	90
圖 4-2 商品項目分類圖.....	91
圖 4-3 模糊歸屬函數.....	91
圖 4-4 商品項目與祖先項目關係圖.....	100
圖 4-5 資料筆數漸增示意圖.....	101
圖 4-6 動態調整門檻值示意圖.....	103

第一章 緒論

第一節 研究背景

隨著網際網路興起、國際市場的開放以及顧客至上的時代，對現今企業產生了許多新的挑戰以及新的經營策略。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 [8,11,12,14]，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[30]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[36]。

最近幾年以來，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。而關聯法則之資料探勘已成為學術界與企業界相當熱門的領域，相關研究的學者過去在此領域提出了許多的演算法，企圖解決資料探勘時所遇到的問題[23]。

第二節 研究動機與目的

在關聯法則的研究中，雖然過去有許多學者相繼提出相當多的方法。然而過去這些關聯法則演算法，遇到以下三點問題將無法同時有效地處理：

一、資料庫數值型資料的處理

在過去相當多的關聯法則演算法，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，而依據此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此項目的交易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。

二、資料庫經常性的變動

在現實生活中，企業的交易是隨時進行的，而資料庫中的資料也必須隨著交易的新增而動態的記錄新的資料，因此產生了在動態資料庫進行漸進式探勘的需求。然而過去關聯法則演算法主要是產生大項目組，而非大項目組的部份並不會保留在記憶體或是硬碟中，也因此當資料庫的資料有所變更時，就必須重新掃描更動後的整個資料庫而不能只考慮新增記錄集。但這對於現今龐大的資料庫來說，重新執行關聯法則是相當費時且費力的過程。也因此，如何不必重新掃描更動後的整個資料庫又能得到資料變動後所造成得結果，這是目前漸進式探勘所要研究的方向。

三、線上關聯法則探勘的問題

使用者常利用網路或在時間有限的情形下，尤其當使用者隨意改變門檻值時，系統必須能很快速地提供正確的法則；在這種情形下，已不適用在傳統多次掃描資料庫來產生高頻項目組的作法了。

以上三點問題，在數值型的資料處理上[18,41,48,50]、資料庫經常性變動的問題[1,2,5,9,10,11,19,43]以及線上關聯法則探勘[1,5,10,11,16]，其實過去有一些研究已經提出許多的方法，但這些方法只針對三點其中某一點或兩點問題來進行解決。若要同時解決這三個問題的探討，過去到現今的研究中並不多。因此本研究想提出一個方法能夠同時有效地解決這三個問題。

第三節 研究方法

在此論文研究中，我們先廣讀有關於資料挖探勘的各種參考文獻，尤其是關聯法則演算法分別在這三個問題上之探討。分別針對這三個問題進行研究，期望能設計出一套適性化的整合性方法，將這三個問題一併解決；因此我們設計出了 Fuzzy Partition Base Association Rules Algorithms(FPBARA)。研究流程如下圖 1-1

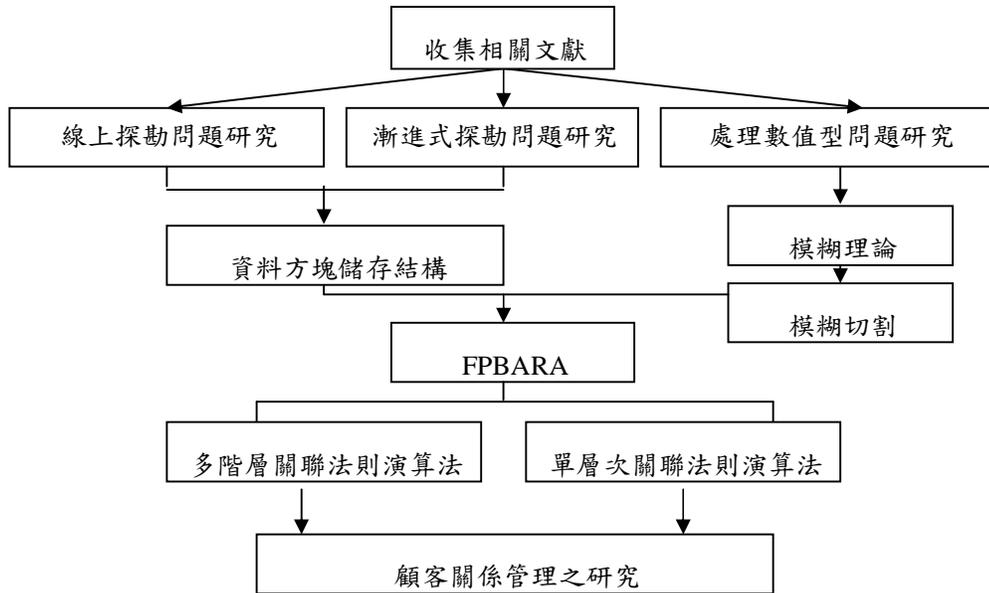


圖 1-1：研究流程圖

根據我們所蒐集的文獻，在處理數值型資料的問題之研究，大部分學者是結合模糊理論的概念來進行處理[18,41,48,50]，而在漸進式探勘與線上探勘的問題上，有的學者採用資料方塊的概念來進行建構[1,5,10,11]。因此本研究利用模糊理論中的模糊切割技術以及資料方塊之儲存結構，來建構本研究的方法—FPBARA。利用 FPBARA 方法來建構模糊單層次關聯法則演算法以及模糊多階層關聯法則演

算法，並應用於顧客關係管理上的探討，來幫助企業有效地做好顧客關係管理，並為企業創造獲利與競爭力的提升。

第四節 研究貢獻

本研究主要研究貢獻有以下兩點如下：

一、能夠同時有效處理數值型資料、漸進式探勘與線上探勘的問題：

我們提出一個能夠同時解決處理數值型資料的問題、漸進式探勘的問題以及線上探勘的問題之方法來建構與改良單層次關聯法則演算法以及多階層關聯法則演算法。

二、顧客關係管理之應用：

利用我們所建構與改良的兩種關聯法則演算法，其在顧客關係管理之應用。其應用有：

(一) 利用我們所提的單層次關聯法則演算法來進行顧客消費

行為之探討—RFM Model 的改良與應用。

(二) 利用我們所提的多階層關聯法則演算法在商品交易上的

商品項目之探勘以及能夠滿足不同階層管理者所需要的

決策資訊。

第五節 論文架構

這篇論文的其餘章節組織如下，第二章我們將詳細探討相關的研究，主要介紹資料探勘的基本概念；以及單層次關聯法則與多階層關聯法則；最後介紹模糊邏輯以及模糊切割。第三章將介紹如何利用 FPBARA 方法來進行單階層關聯法則，並介紹其如何應用於顧客關係管理來幫助企業經營者進行行銷決策。第四章將介紹如何利用 FPBARA 方法來進行多階層關聯法則。最後，第五章我們將簡短地為這篇論文做總結和描述未來的研究方向。

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.[32]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 [32]

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[47]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[17]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的种类與實用性。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：

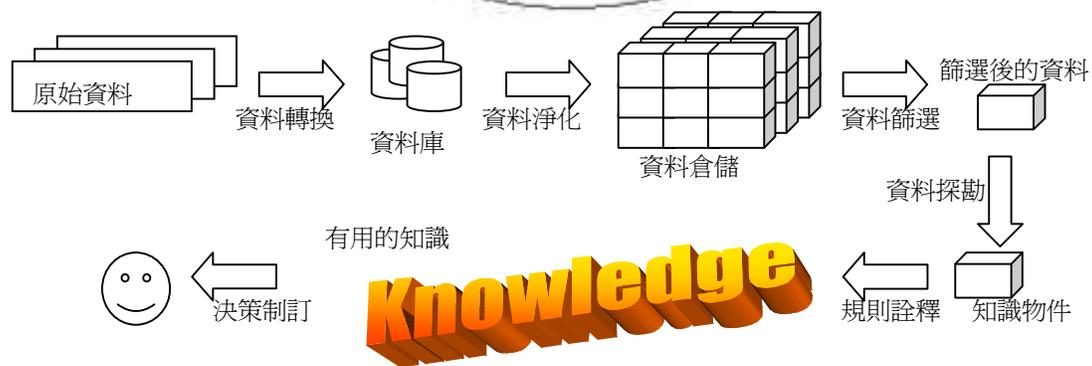


圖 2-1：知識挖掘過程示意圖

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) [33]……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[35]。關聯法則模式最常見的演算法如 Apriori 演算法[34]、Sampling 演算法[22]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 $X \cap Y = \emptyset$ ，其中 X 和 Y 分別表示在資料庫中不同之資料項目組 [38,39]。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 $support(X \cup Y)/support(X)$ 。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 [34]。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} ，直到無法再產生長度更長的大項目組則停止。

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：

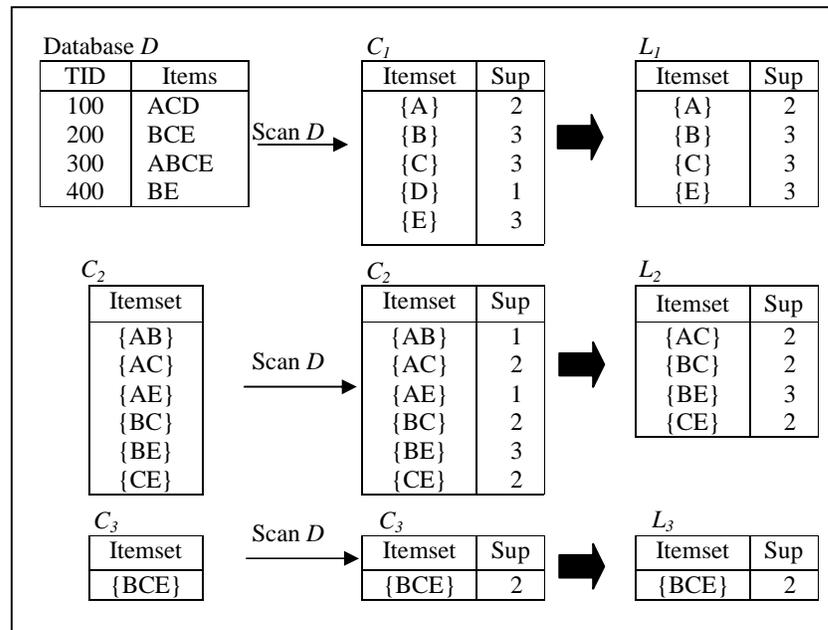


圖 2-2：產生大項目組之範例

(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[19]，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 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，我們以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許多隱藏的知識便無法被挖掘出來的問題產生。

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[7]。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

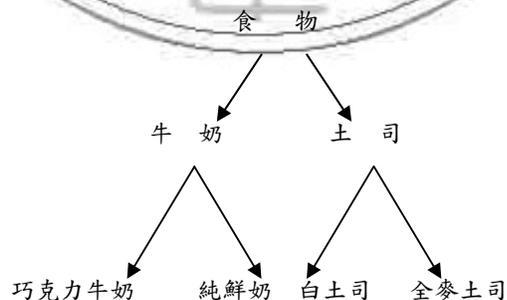


圖 2-3：食物項目分類以 Taxonomy 為例

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 [24] 與 Ramakrishnan Srikant.et al(1997).所提出的Cumulate演算法、Stratify演算法與 EstMerge演算法[42]

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_frequentItemsets(T[1],level);
T[2]=get_frequent_t_table(T[1],L[1,1]);
}
else L[level,1]=get_frequentItemsets(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後選項目組的話，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  $T$ .
 $L_1 := \{\text{frequent 1-itemsets}\}$ ;
 $k := 2$ ;
while ( $L_{k-1} \neq \emptyset$ ) do
  begin
     $C_k :=$  New candidates of size  $k$  generated from  $L_{k-1}$ .
    if ( $k = 2$ ) then
      Delete any candidate in  $C_2$  that consists of an item and its ancestor.
      Delete any ancestors in  $T^*$  that are not present in any of the candidates in  $C_k$ .
    forall transactions  $t \in D$  do
      begin
        foreach item  $x \in t$  do
          Add all ancestors of  $x$  in  $T^*$  to  $t$ .
          Remove any duplicates from  $t$ .
          Increment the count of all candidates in  $C_k$  that are contained in  $t$ .
        end
       $L_k :=$  All candidates in  $C_k$  with minimum support.
       $k := k+1$ ;
    end
  Answer :=  $\cup_k L_k$ ;

```

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}。透過這樣的概念，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，則其的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。

綜合上述介紹這四種多階層關聯法則的特點與缺失，如下表

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[27]。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 $m_A(x)$ 。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、 $m_A(\cdot)$ 為歸屬度函數) 其表示如下：

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

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 m_A 、 m_B 且 $A = \{(u_i, m_A(u_i)) \mid u_i \in U\}$ ， $B = \{(u_i, m_B(u_i)) \mid u_i \in U\}$ ，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：

$$m_{A \cap B}(u_i) = \min(m_A, m_B) \quad u_i \in U \dots\dots\dots (公式 1)$$

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：

$$m_{A \cap B}(u_i) = m_A \otimes m_B = m_A \times m_B \quad u_i \in U \dots\dots\dots (公式 2)$$

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

$$m_{A \cup B} = \max(m_A, m_B) \quad u_i \in U \dots\dots\dots (公式 3)$$

利用代數和方法如下：

$$m_{A \cup B} = m_A + m_B - (m_A \times m_B) \quad u_i \in U \dots\dots\dots (公式 4)$$

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[3,20,21,27,28,49,50]。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[3,20,21,27,28,49,50]。

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[49]，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：

$$m_A(a, b, c; x) = \begin{cases} 0, x < a \\ (x-a)/(b-a), a \leq x \leq b \\ (c-x)/(c-b), b \leq x \leq c \\ 0, x > c \end{cases} \dots\dots\dots (公式 5)$$

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：

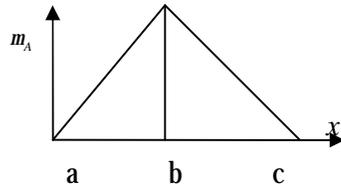


圖 2-6: 三角形歸屬度函數

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[20]與 1995 年[21]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 個集合的模糊歸屬度函數定義如下（公式 6）。

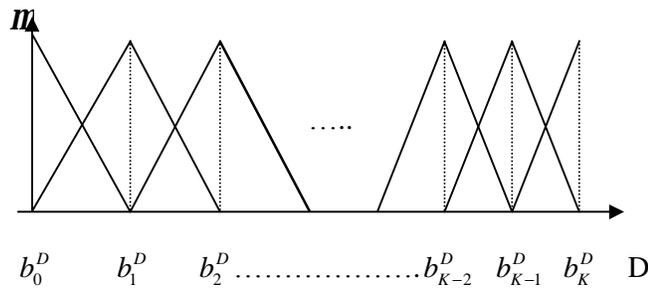


圖 2-7: 以三角形歸屬度函數建立的 Simple fuzzy partition

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

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[48]。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

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

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：

$$\begin{aligned} m_1^{Age}(10) &= \frac{(30-10)}{30} = 0.667 \\ m_2^{Age}(10) &= \frac{(10-0)}{30} = 0.337 \\ m_3^{Age}(10) &= 0 \end{aligned}$$

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[50]。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 $m_i^{D_i}(x_i)$ 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：

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

..... (公式 8)

⊗ 為 Fuzzy conjunction operator 若取代數積 (Product operation) 作為 Fuzzy conjunction operator 則

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

..... (公式 9)

若取最小運算 (Min operation) 則

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

..... (公式 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：

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

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

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

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

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

以第一筆為例，年齡的值為 10，其歸屬度計算為：

$$m_1^{Age}(10) = 0.667, m_2^{Age}(10) = 0.333, m_3^{Age}(10) = 0$$

而薪資歸屬度計算為：

$$m_1^{salary}(1000) = 0.92, m_2^{salary}(1000) = 0.08, m_3^{salary}(1000) = 0, m_4^{salary}(1000) = 0, \text{及 } m_5^{salary}(1000) = 0 \text{。}$$

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：

$$m_{1 \times 1}^{Age \times Salary}(10, 1000) = m_1^{Age}(10) \times m_1^{Salary}(1000) = 0.667 \times 0.92 = 0.614,$$

$$m_{1 \times 2}^{Age \times Salary}(10, 1000) = m_1^{Age}(10) \times m_2^{Salary}(1000) = 0.053,$$

$$m_{2 \times 1}^{Age \times Salary}(10, 1000) = m_2^{Age}(10) \times m_1^{Salary}(1000) = 0.306,$$

$$m_{2 \times 2}^{Age \times Salary}(10, 1000) = m_2^{Age}(10) \times m_2^{Salary}(1000) = 0.027,$$

其他的歸屬度皆為 0，依此可類推算出其他 7 筆的歸屬度。

□

第三章 模糊切割在關聯法則之研究與應用

本章我們將介紹如何利用模糊切割的概念來建構單層次關聯法則，我們利用這方法應用於顧客關係管理中的 RFM Model 的改良—Fuzzy RFM Model。首先，在第一節部分，我們將介紹這方法的流程與演算法，並用一範例說明之。在第二節部分，我們將介紹傳統 RFM Model 的方法以及這些方法的缺失。在第三節部分，介紹 Fuzzy RFM Model 的流程與演算法，並用一範例說明之。在第四節部分，介紹 RFM Model 的實驗，並探討其結果。最後，此章結論於第五節部分。

第一節 模糊切割於單層次關聯法則之建構

一、問題描述

在單層次關聯法則的研究中，過去雖然有許多學者相繼提出相當多的演算法，然而這些演算法在數值型資料、動態資料庫的漸進式挖掘以及線上挖掘等問題上都有其限制。雖然近幾年，有學者針對這三個問題的一兩點問題來進行解決，但能夠完全處理這三問題的單層次關聯法則演算法並不多見。因此，我們將利用模糊切割來處理數值型資料，並運用資料方塊的結構優勢來處理漸進式挖掘與線上挖掘等問題。當我們在商品項目進行模糊切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（Item），如此可以利用 Apriori-like 演算法擷取商品項目之間的模糊關聯法則[1,11]。但當資料庫中的資料量很龐大時，Apriori 演算法[34,37]是相當費時的。我們可以事先計算這些項目組的模糊支持度（Fuzzy support）值並儲

存在資料方塊 (Data cube) 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 (On-line mining) 的功能[1,11]，而當新增紀錄時，亦可即時加以處理並更新資料方塊 (Data cube) 中對應項目組的支持度，以達到漸進式挖掘(Incremental mining)的目的 [1,11]。

二、模糊切割之單層次關聯法則演算法

假定我們在每個商品項目上分割之模糊集合個數為K；使用者所設定的門檻值分別為最小模糊支持度 α 與最小模糊信賴度 β ，則模糊切割之單層次關聯法則的步驟說明與步驟流程圖3-1如下：

步驟 1 從交易表格中取出我們要的資訊 (交易代號、交易商品項目以及每個商品項目的數值，如購買的量)，並建構一個交易轉換表。

步驟 2 在每一筆交易紀錄中，所有商品項目資訊 (如數量) 透過模糊切割而轉換成商品之模糊項。並透過模糊歸屬函數計算每一個商品資訊在各個模糊項的歸屬度值，最後加總為每一個商品之模糊集合 (長度為1的候選項目組) 的模糊支持度 (Fuzzy support) (如下公式11) 並儲存至資料方塊 (Data cube) 中。

$$FS(A_{i_1}^{D_1} \times A_{i_2}^{D_2} \times \dots \times A_{i_k}^{D_k}) = \sum_{i=1}^n m_{i_1 \times i_2 \times \dots \times i_k}^{D_1 \times D_2 \times \dots \times D_k}(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = \sum_{i=1}^n m_{i_1}^{D_1}(x_{i_1}) \times m_{i_2}^{D_2}(x_{i_2}) \times \dots \times m_{i_k}^{D_k}(x_{i_k})$$

其中 x_{i_k} 為第 i 筆資料之維度 D_k 的值。

..... (公式 11)

步驟 3 依據每一個長度為 1 之項目 C_1 進行合併，成為長度為 2 之項目組 C_2 。

步驟 4 利用公式 9 或公式 10 來計算長度為 2 之項目組在每一筆的歸屬度值，並加總歸屬度值為長度為 2 之項目組的支持度 ($FS(C_2)$)，最後儲存至資料方塊中。

步驟 5 反覆步驟 3 與步驟 4 的方式產生其他長度之項目組以及其支持度，直到產生項目長度為 k 之項目組 C_k ，而無法合併產生項目長度為 $(k+1)$ 之項目組為止。

步驟 6 從資料方塊中每一個 C_1 的模糊支持度到每一個 C_k 的模糊支持度與最小模糊支持度 α 進行比對，符合 ($FS(C_t) \geq \alpha \mid t=1 \text{ to } k$) 的項目組為大項目組。

步驟 7 由大項目組產生候選法則。

步驟 8 計算候選法則的模糊信賴度 (如下公式 12) 並與最小模糊信賴度 β 進行比對，符合 ($FC(R) \geq \beta$) 的法則為最終符合我們需求的法則。

Rule R : $A_{t_1}^{D_1} \times A_{t_2}^{D_2} \times \dots \times A_{t_i}^{D_i} \Rightarrow A_{t_{i+1}}^{D_{i+1}} \times \dots \times A_{t_k}^{D_k}$ 則

$$FC(R) = \frac{FS(A_{t_1}^{D_1} \times A_{t_2}^{D_2} \times \dots \times A_{t_k}^{D_k})}{FS(A_{t_1}^{D_1} \times A_{t_2}^{D_2} \times \dots \times A_{t_i}^{D_i})} \dots\dots\dots (公式 12)$$

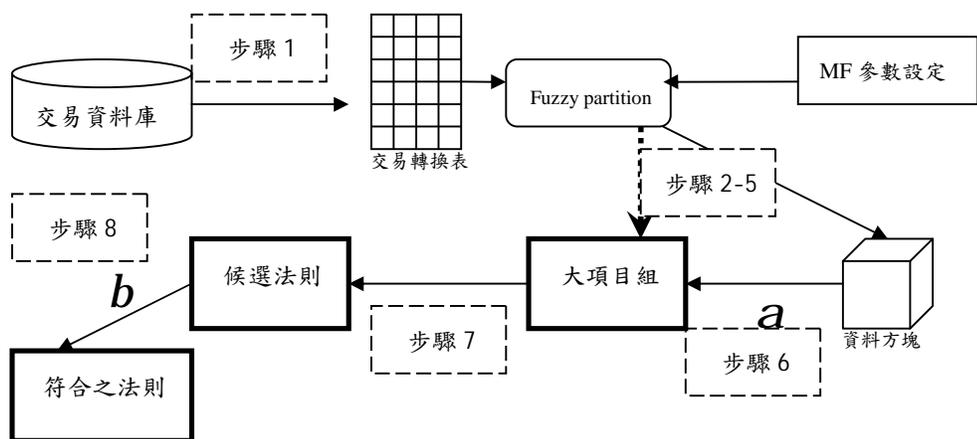


圖3-1：模糊切割之單層次關聯法則流程圖

三、範例說明

接下來我們將利用一個範例來說明模糊切割之單層次關聯法則的建構過程。首先交易資料表3-1以及我們的模糊歸屬函數（圖3-2）

表3-1: 交易資料表

交易ID	項目
1	(milk,3)(juice,4)(cookie,2)
2	(juice,3)(bread,7)(cookie,7)
3	(milk,2)(juice,10)(cookie,5)
4	(milk,7)(juice,2)(bread,8)(cookie,8)
5	(juice,2)(bread,8)(cookie,10)

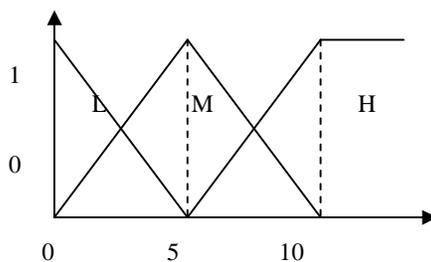


圖 3-2: 模糊歸屬函數

假設我們所設定的最小模糊支持度 $\alpha=1.0$ 與最小模糊信賴度 $\beta=0.51$ 。接下來開始進行我們的模糊切割之單層次關聯法則的建構。

步驟 1 從交易表格中的取出我們要的資訊（交易代號、交易商品項目以及每個商品項目的數值，如購買的量）如下表 3-2：

表3-2: 交易資料轉換表

交易ID	項 目
1	(A,3)(B,4)(D,1)
2	(B,3)(C,7)(D,7)
3	(A,2)(B,10)(D,5)
4	(A,7)(B,2)(C,8)(D,8)
5	(B,2)(C,8)(D,10)

步驟 2 將所有商品項目透過模糊切割而轉換成商品之模糊集合（A.L、A.M、A.H、B.L、B.M、B.H、C.L、C.M、C.H、D.L、D.M、D.H），並透過模糊歸屬函數計算每一個商品之模糊集合在每一筆的歸屬度值，最後加總為每一個商品之模糊集合（長度為1的項目組）的Fuzzy Count作為模糊支持度（FS）（如下表3-3）並儲存至資料方塊中。在表3-3中灰底為大項目組：

表3-3：長度為1之項目

	1	2	3	4	5	FS
A.L	0.4	0.0	0.6	0.0	0.0	1.0
A.M	0.6	0.0	0.4	0.6	0.0	1.6
A.H	0.0	0.0	0.0	0.4	0.0	0.4
B.L	0.0	0.4	0.0	0.6	0.6	1.6
B.M	0.0	0.6	0.0	0.4	0.4	1.4
B.H	0.0	0.0	1.0	0.0	0.0	1.0
C.L	0.2	0.0	0.0	0.0	0.0	0.2
C.M	0.8	0.6	0.0	0.4	0.4	2.2
C.H	0.0	0.4	0.0	0.6	0.6	1.6
D.L	0.8	0.0	0.0	0.0	0.0	0.8
D.M	0.2	0.6	1.0	0.4	0.0	2.2

D.H	0.0	0.4	0.0	0.6	1.0	2.0
-----	-----	-----	-----	-----	-----	-----

步驟 3 依據每一個長度為 1 之候選項目進行合併，成為長度為 2 之候選項目組，如下表 3-4：

步驟 4 利用公式9來計算長度為2之候選項目組在每一筆的歸屬度值，並加總歸屬度值為長度為2之候選項目組的支持度，例如A.L*B.L的第一筆的歸屬值計算為 $A.L(1) = 0.4$ 去乘上 $B.L(1) = 0.0$ 得到A.L*B.L的第一筆的歸屬值為0.0。依此類推算完A.L*B.L所有五筆的歸屬值並加總為0.0（如表3-4），其他項目也依此類推計算之。最後儲存至資料方塊中（如表3-5）。在表3-5中灰底為大項目組：

表3-4：A.L*B.L範例表

交易ID	A.L	B.L	A.L*B.L
1	0.4	0.0	0.0
2	0.0	0.4	0.0
3	0.6	0.0	0.0
4	0.0	0.6	0.0
5	0.0	0.6	0.0
Fuzzy count			0.0

表3-5：長度為2之項目

項目組	FS	項目組	FS
A.L*B.L	0.00	B.L*C.L	0.00
A.L*B.M	0.32	B.L*C.M	0.72
A.L*B.H	0.60	B.L*C.H	0.88
A.L*C.L	0.08	B.L*D.L	0.00
A.L*C.M	0.32	B.L*D.M	0.48
A.L*C.H	0.00	B.L*D.H	1.12
A.L*D.L	0.32	B.M*C.L	0.00
A.L*D.M	0.68	B.M*C.M	0.52
A.L*D.H	0.00	B.M*C.H	0.72
A.M*B.L	0.36	B.M*D.L	0.00
A.M*B.M	0.24	B.M*D.M	0.52
A.M*B.H	0.40	B.M*D.H	0.88

A.M*C.L	0.12	B.H*C.L	0.00
A.M*C.M	0.72	B.H*C.M	0.00
A.M*C.H	0.36	B.H*C.H	0.00
A.M*D.L	0.48	B.H*D.L	0.00
A.M*D.M	0.76	B.H*D.M	1.00
A.M*D.H	0.36	B.H*D.H	0.00
A.H*B.L	0.24	C.L*D.L	0.16
A.H*B.M	0.16	C.L*D.M	0.04
A.H*B.H	0.00	C.L*D.H	0.00
A.H*C.L	0.00	C.M*D.L	0.64
A.H*C.M	0.16	C.M*D.M	0.68
A.H*C.H	0.24	C.M*D.H	0.88
A.H*D.L	0.00	C.H*D.L	0.00
A.H*D.M	0.16	C.H*D.M	0.48
A.H*D.H	0.24	C.H*D.H	1.12

步驟 5 反覆步驟3與步驟4的方式產生其他長度之項目組以及其支持度，直到產生項目長度為4之項目組為止。

步驟 6 從每一個長度為1之候選項目的模糊支持度到每一個長度為4之候選項目組的模糊支持度與最小模糊支持度1.0進行比對，大於等於1.0的項目組為大項目組，結果如下表3-6：

表3-6： Large itemsets

大項目組	FS	大項目組	FS
A.L	1.0	C.H	1.6
A.M	1.6	D.M	2.2
B.L	1.6	D.H	2.0
B.M	1.4	B.L*D.H	1.12
B.H	1.0	B.H*D.M	1.00
C.M	2.2	C.H*D.H	1.12

步驟 7-8 由大項目組產生候選法則，我們以大項目組(B.L, D.H)為例，則可產生以下2條法則並計算法則之信賴度，就以B.L→D.H為例其信賴度0.8是由B.L*D.H的支持度1.12去除以B.L的支持度1.4得之，結果如下表3-7。其

他的大項目組也依此類推算出候選法則之信賴度並與最小模糊信賴度 $\beta = 0.51$ 比較，得到符合的法則如下表3-8：

表3-7：(B.L, D.H)為例之候選法則

法則	信賴度
B.L→D.H	$(B.L * D.H) / B.L = 1.12 / 1.4 = 0.80$
D.H→B.L	0.509

表3-8：符合最小模糊信賴度之法則

符合法則	
B.L→D.H (0.8)	C.H→D.H (0.7)
B.H→D.M (1.0)	D.H→C.H (0.56)

□

四、相關衍生問題的解決與應用

我們利用模糊理論的概念來處理傳統單層次關聯法則演算法無法處理的數值型資料之外，我們並結合模糊切割與資料方塊技術，來建構一個能夠解決現實資料庫是經常性變動與使用者可以隨時調整門檻值所造成的問題之單層次關聯法則演算法。

我們將利用這新的方法，應用於顧客關係管理中，在分析顧客消費行為相當著名的方法：RFM Model 來進行改良。此改良型 RFM Mode 我們稱之為 Fuzzy RFM Model。在介紹 Fuzzy RFM Model 之前，我們先介紹傳統 RFM Model 以及其缺失於第二節；最後，再介紹 Fuzzy RFM Model 於第三節並進行 Fuzzy RFM Model 的實驗與探討於第四節。

第二節 傳統 RFM Model

在顧客關係管理的許多研究中，發展出相當多的模式來提供企業有效的工具，來幫助經營者瞭解其顧客，進而制訂正確且能滿足顧客需求的行銷策略。而在這許多模式中，RFM Model 是其中相當著名的方法。RFM Model 主要是在分析且衡量顧客的消費行為，由於顧客的消費行為是相當抽象且難以衡量的，因此必須透過一些可衡量的指標性參數來進行分析與衡量。而其中最近購買時間（**R**ecency）、購買頻率（**F**requency）與購買金額（**M**onetary）三個參數，就是一個具參考性且指標性參數，簡稱為 RFM。

分析顧客過去的購買行為以判斷那些顧客值得進一步的接觸是發展行銷策略上重要的議題[12,14,44,50]。不同特性的顧客適合不同的行銷策略，因此，顧客分群是客製化行銷策略決策上首要的工作。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。R 值愈小，則隱含著該顧客再次選購此產品的購買程度愈高；反之，R 值愈大，隨著時間之拉長該顧客之持續購買慾隨之降低，則表示著此顧客的購買行為可能改變或是變節至他處消費。F 值主要是在測量顧客在此時間內與公司之互動程度，若 F 值愈高則代表此顧客與公司互動程度愈高；該顧客對此產品有愈高的熱衷程度，持續購買之動機亦較強。M 值主要是代表顧客對此產品之興趣指標，也是對企業之實質金錢貢獻。某顧客的 M 值愈多，代表該顧客大量購買此產品，對此商品具有大量之需求。

我們可以在許多文獻中看到許多 RFM Model 改良以及應用 [8,26,31,45,46]。[46]利用 SOM 將顧客的 RFM 先分群，然後，比較各群的 RFM 平均值和全部的 RFM 平均值，觀察其變化是否上升或下降，來判斷各群的顧客是否為忠誠度高的黃金顧客還是潛在顧客或是即將流失的顧客。其缺點是有些變化的組合很難去解釋、以及變化的幅度未考量。例如同樣是 R↑、F↑、M↑但上升的幅度不同，其實所代表的意義可能是不一樣的。

Stone 在 1989 年提出的 Stone RFM Model[45]來計算 R、F、M 分數，在 R 值部份，時間分為本季、距今六個月內、距今九個月內以及今年分別給予 24、12、6、3 分等分數；在 F 值部分，把購買次數乘以 4 當作 F 值的分數；在 M 值計算部份，就是以消費金額的百分之十當作分數，但若 M 值大於 9，則只取 9，主要是避免購買頻率低卻具有大量消費金額的情形。將顧客的 R、F、M 分數予以加總後，若其 RFM 總分大於使用者設定的門檻值，則為潛在型顧客或為黃金顧客。

顧客五等分法[31]是將顧客消費記錄之最近購買時間、購買次數及購買金額等三個維度值分別平均分成五等分，亦即 (R,F,M)=(1,1,1)...(5,5,5)，至多將分出 125 個級別之顧客。(5,5,5)即為 15 分，而(5,4,3)為 12 分，分數愈高者代表後續購買某產品之潛在持續購買力愈大。

行為五等分法[31]是依顧客購買行為將顧客排序。和顧客五等分法一樣，也將三個維度各分成五等分，但不同的是 R 值與 F 值部份。其 R 值分為前三個月、前四至六個月間、前七至十二個月間、前三至二十四個月以及二十五個月以上等五類，分別給予 5、4、3、2、

1 分；F 值是先將只購買一次的顧客分為一等分，然後計算其餘顧客的平均購買次數，高於此平均者分為一等分，再計算其餘顧客的平均購買次數，高於此平均者再分為一等分，重複此方法，將購買頻率由高至低分為 5 群，分別給予 5 分至 1 分；M 值是依據購買金額由多至寡分別給予分數 5 分至 1 分。將顧客的 R、F、M 分數予以加總後，若其 RFM 總分大於使用者設定的門檻值，則為潛在型顧客或為黃金顧客。

雖然 RFM Model 在顧客關係管理上算是最讓人簡單易懂，但上述介紹的 RFM Model 仍存在一些缺失與限制。例如，Stone RFM 與行為五等分法在 R 的部分採用固定式區間長度的切割方式來給分，但並非所有產品都適用這樣子的分法，因為每樣產品都有它們自己的生命週期，像汽車可以用到五年至十年左右但電腦只可以用到三年左右，這兩個產品的生命週期並不相同，很難用這樣的標準去給分。另外，這三個方法在給分上也過於主觀認定，同時，也不易區分與解釋同分的情形，例如，(1,4,5)，(4,1,5)，(5,1,4) 這三者彼此所代表的意義並不相同。再者，顧客與行為五等分法至多分成 125 群顧客，有時當顧客人數以及消費記錄未具規模時，實際的群數常小於 125。我們將這四個 RFM Model 的缺點整理成如下表 3-9：

表 3-9：傳統 RFM Model 的缺點比較表

RFM Model 缺點	[46]	Stone RFM Model	顧客五等分	行為五等分法
過於主觀	×	●：1. 門檻值設定 2. 給分偏向行銷人員主觀認定	●：門檻值設定	●：1. 門檻值設定 2. 給分偏向行銷人員主觀認定

產品生命週期問題	×	●：R 的部分採用固定式區間長度的切割方式來給分，難適合所有產品	●：R 的部分採用固定式區間長度的切割方式來給分，難適合所有產品	●：R 的部分採用固定式區間長度的切割方式來給分，難適合所有產品
同分問題	×	●：(12,24,4)，(6,32,2)，(24,8,8)同樣為 40 分但代表意義不同	●：(1,4,5)，(4,1,5)，(5,1,4)同樣為 10 分但代表意義不同	●：(1,4,5)，(4,1,5)，(5,1,4)同樣為 10 分但代表意義不同
分群的問題	×	×	●：分成 125 群，但有時顧客人數及消費紀錄未具如此規模時	●：分成 125 群，但有時顧客人數及消費紀錄未具如此規模時
其他問題	1. 有些組合難解釋 2. 變化的幅度也未被考量到	×	×	×

由這些研究可知在 R、F、M 值的量化上，分界點的決定是一件不容易的事。不同領域的問題需要不同的分割位置與群數。而傳統的區間分割方式會造成鮮明邊界 (sharp boundary) 的問題。另外，人們在描述知識及進行決策時常會使用易於瞭解的含糊的用詞 (vague terms) 以便處理不精確與不完全的資訊。模糊邏輯利用模糊集合的歸屬函數提供不同集合之間的模糊邊界 (smooth boundary) [3,20,21,27,28,49,50]，而每一集合可用人們常用的自然語言用詞描述。因此本研究利用模糊分割 (Fuzzy partition) [20,21,28,50] 的觀念來建構一模糊 RFM Model，以解決傳統的 RFM Model 在 R、F、M

這三個參數上其量化上的問題，以便更有彈性地取得顧客特性與分割等資訊。

再者，探討 R、F、M 的關聯性，在企業的顧客關係管理中也是相當重要的。舉個例子來說，若某企業發現”最近購買時間長”與”購買次數少”具有關聯性，則可進一步加以分析其原因，以決定因應的策略。例如，從產品在市場的特性來分析，若企業所賣的產品屬於高價或是耐久性的產品，例如，汽車、珠寶等，則在行銷策略上就必須強調售後服務；但若是從顧客的消費者行為來分析，則表示企業有許多的顧客都是屬於即將快流失的顧客；若從市場狀況來分析，可能是由於經濟景氣不佳的影響所導致的現象。因此，探討 R、F、M 的關聯性，除了能夠讓行銷決策者更清楚瞭解顧客的行為之外，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的達成行銷決策。文獻上幾乎沒有關於此議題的研究與應用。因此，本研究將探討挖掘模糊 RFM 關聯法則的方法。

我們提出的 Fuzzy RFM Model 將 RFM 值量化後表示成 3 維的資料方塊，除了可以彈性地讓企業管理者分析顧客特性與分群外，還可以找出 R、F、M 這三者之間的關聯性。這些挖掘出的知識有望能夠讓企業用最全面的角度且更深一層地瞭解企業所在的市場與顧客特性，以作為企業在行銷決策上重要的依據。接下來，第三節，我們將介紹 Fuzzy RFM Model 的流程以及其功能。

第三節 Fuzzy 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 3-3 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。

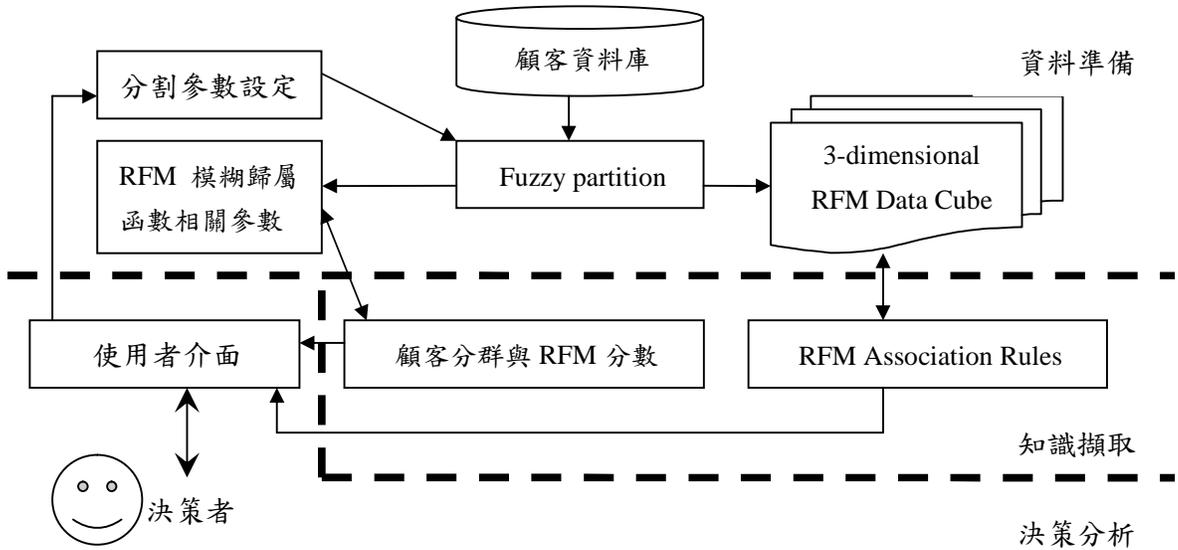


圖 3-3：模糊 RFM Model 流程架構圖

根據 Fuzzy RFM Model 的流程，我們可分為三個部分。分別為資料準備的部分、知識擷取的部分、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) 中詳細說明之了，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、顧客 RFM 分數以及 RFM Association Rules 這三部分，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[1,11]。但當資料庫中的資料量很龐大時，Apriori 演算法[34,37]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[1,11]，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[1,11]。RFM 值的 3 維資料方塊的圖示如下圖 3-4:

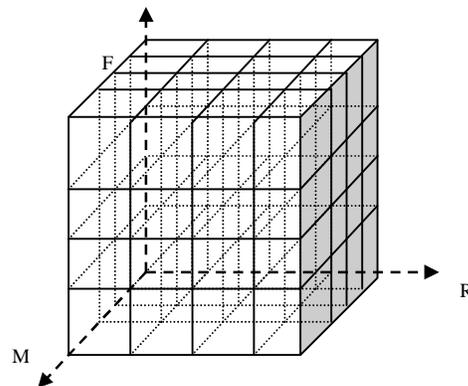


圖 3-4：RFM 3 維資料方塊

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

接下來，我們用一範例說明之：

- 步驟 1 假設某企業的顧客交易資料庫轉為顧客 RFM 資料表如下表 3-10
- 步驟 2 假設此企業行銷主管設定 R 分成三部分（最近購買時間短 (R.L) / 中 (R.M) / 長 (R.H)）；F 也分成三部分（最近購買次數少 (F.L) / 普通 (F.M) / 高 (F.H)）；M 也分成三部分（購買金額小 (M.L) / 普通 (M.M) / 大 (M.H)），並分別建立 R、F、M 的三角形歸屬度函數，如下圖 3-5 所示：

表 3-10：顧客 RFM 資料表

	R(天數)	F(次數)	M(金額)
1	8	5	4500
2	10	10	3000
3	19	3	1500
4	25	7	6000
5	5	6	3500
6	1	5	10000
7	51	4	15000
8	14	2	8500
9	34	9	21500
10	40	8	5000

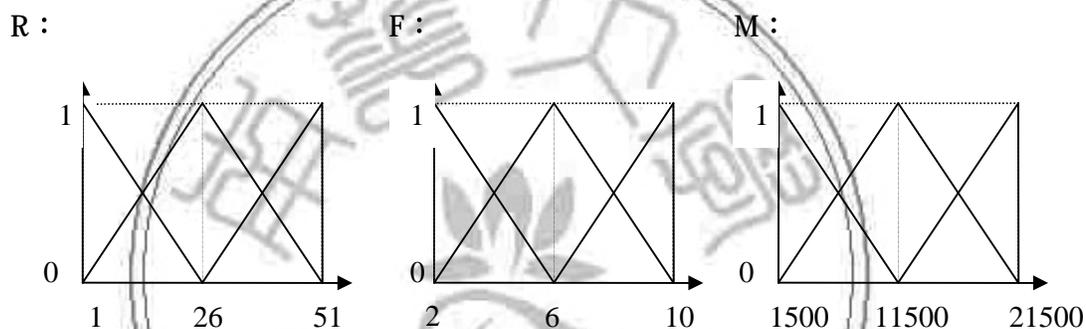


圖 3-5：RFM 三角形歸屬度函數

步驟 3 計算資料方塊 (Data Cube) 中，各 itemsets 的模糊支持度值，就以 1-itemset 的 R.L 為例，分別計算出其十筆的歸屬度分別為 0.72、0.64、0.28、0.04、0.84、1.00、0.00、0.52、0.00、0.00，並將這些歸屬度加總再除以總筆數得到 0.404，即為 R.L 的模糊支持度值，將該值存入座標為(1,0,0)之中。所有 1-itemsets 的模糊支持度值，如表 3-11 所示：

表 3-11：1-itemsets RFM 資料方塊

	1	2	3
R	0.404	0.412	0.188

	1	2	3
F	0.275	0.425	0.25

	1	2	3
M	0.5	0.365	0.135

而 2-itemsets 的計算就以 R.L×F.L 為例。就第一筆而言， $m_{x1}^{R \times F}(8,5) = m^R(8) \times m^F(5) = 0.72 \times 0.25 = 0.18$ ，其餘 9 筆的歸屬度分別為 0，0.21，0，0.84，0.75，0，0.52，0，0，如此可求得 R.L×F.L 的模糊支持度值為 0.112，存入 Data cube (1,1,0) 之中。所有的 2-itemsets 的模糊支持度值如下表 3.12 所示：

表 3-12：2-itemsets RFM 資料方塊

	F	1	2	3
R				
1		0.112	0.223	0.065
2		0.113	0.116	0.133
3		0.05	0.086	0.052

	M	1	2	3
R				
1		0.2316	0.1684	0.00
2		0.232	0.112	0.068
3		0.0364	0.0846	0.067

	M	1	2	3
F				
1		0.12625	0.13125	0.0175
2		0.2425	0.19	0.0425
3		0.13125	0.04375	0.075

依此類推計算出 3-itemsets 的模糊支持度值，如下

表 3-13 所示：

表 3-13： 3-itemsets RFM 資料方塊

M = 1				M = 2				M = 3			
F \ R	1	2	3	F \ R	1	2	3	F \ R	1	2	3
1	0.05175	0.1249	0.05495	1	0.06025	0.0981	0.01005	1	0.00	0.00	0.00
2	0.0745	0.0994	0.0589	2	0.0385	0.0496	0.0239	2	0.00	0.017	0.051
3	0.00	0.0182	0.0182	3	0.0325	0.0423	0.0098	3	0.0175	0.0225	0.024

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 3-14 的顧客分群表[4]。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 3-14：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

以第 9 位顧客為例，他的最近購買時間與計算日的時間距離是中 (R.M) 和長 (R.H) 的程度分別為 0.68 與 0.32；他的購買次數是普通 (F.M) 和高 (F.H) 的程度分別為 0.25 與 0.75；他的購買金額是大 (M.H) 的程度是 1.00。根據顧客分群表，我們可以把第 9 位顧客歸類為第 10 (R.M、F.H、M.H)、13 (R.M、F.M、M.H)、19 (R.H、F.H、M.H)、22 (R.H、F.M、M.H) 群的顧客，並且以這四群顧客的特性來對此顧客進行行銷策略。

雖然把第 9 位顧客歸類為第 10、13、19、22 群的顧客，但實際上第 9 位顧客在這四群的程度還是有差異的，所以我們

以（公式 8 或公式 9）算出所對應群的歸屬度，來為第 9 位顧客歸類的群作排序。

第 9 位顧客屬於第 10 群（R=M、F=H、M=H）歸屬度為 $0.68 \times 0.75 \times 1.00 = 0.51$ ；第 9 位顧客屬於第 13 群（R=M、F=M、M=H）歸屬度為 0.17；第 9 位顧客屬於第 19 群（R=H、F=H、M=H）歸屬度為 0.24；第 9 位顧客屬於第 22 群（R=H、F=M、M=H）歸屬度為 0.08。

然後經過排序得知第 9 位顧客屬於第 10 顧客群的程度較大，其次分別屬於第 19、第 13、第 22 顧客群。我們也可以設定一個 α -cut 的門檻值來選出每個顧客屬於程度較高的顧客群。例如當我們 α -cut 設為 0.3 時，則第 9 位顧客屬於第 10 顧客群。

（二）顧客 RFM 分數：

顧客分群是企業瞭解顧客行為與制定行銷策略時相當重要的工作，但除了顧客分群外，還有一個最重要的工作是找出對企業最重要且最具價值的顧客，這對於目前競爭激烈且資源有限的環境下，企業的行銷決策能夠有效地達到較佳的成效，這項工作是相當重要的，而計算每位顧客 RFM 分數就是一個評估企業最重要且最具價值的顧客的方法之一。在 Fuzzy RFM model 中，顧客 RFM 分數的計算步驟如下：

步驟 1 企業管理者分別針對 RFM 上定義的模糊集合對企業的重要性與價值給定權重值。原則上，對企業愈有利則權重值愈大，以便有效的反應出

顧客價值。例如，R 值愈小，則權重值愈高，表 3-15 是一個具體的例子。

步驟 2 計算 R 值在其各模糊集合的歸屬度及其對應之權重值的乘積，將這些乘積加總即得到 R 的分數，定義如下：

$$S_R(v_R) = \sum_{i=1}^{K_R} W_i^R \times m_i^R(v_R) \dots\dots\dots (公式 13)$$

同理，F 與 M 的分數定義如下：

$$S_F(v_F) = \sum_{i=1}^{K_F} W_i^F \times m_i^F(v_F) \dots\dots\dots (公式 14)$$

$$S_M(v_M) = \sum_{i=1}^{K_M} W_i^M \times m_i^M(v_M) \dots\dots\dots (公式 15)$$

步驟 3 將 R、F、M 的分數加總即得到顧客 RFM 分數，定義如下：

$$S_{RFM}(v_R, v_F, v_M) = S_R(v_R) + S_F(v_F) + S_M(v_M) \quad (公式 16)$$

顧客 RFM 分數愈高者則這些顧客對企業重要性與價值性就愈高，因此企業可以利用門檻值的方式篩選出企業最重要且最具價值的顧客。假定 R、F、M 這三個參數都分割成三部分 (K=3)，其 R、F、M 權重值分別如下表 3-15 所示：

表 3-15：RFM 權重表

RFM 狀態	權重值	RFM 狀態	權重值	RFM 狀態	權重值
R.L	$W_1^R = 3$	F.L	$W_1^F = 1$	M.L	$W_1^M = 1$
R.M	$W_2^R = 2$	F.M	$W_2^F = 2$	M.M	$W_2^M = 2$
R.H	$W_3^R = 1$	F.H	$W_3^F = 3$	M.H	$W_3^M = 3$

以第 9 位顧客為例，其 RFM 分數計算如下：

$$S_R(34) = m_1^R(34) * 3 + m_2^R(34) * 2 + m_3^R(34) * 1 = 0 * 3 + 0.68 * 2 + 0.32 * 1 = 1.68,$$

$$S_F(9) = m_1^F(9) * 1 + m_2^F(9) * 2 + m_3^F(9) * 3 = 0 * 1 + 0.25 * 2 + 0.75 * 3 = 2.75,$$

$$S_M(21500) = m_1^M(21500) * 1 + m_2^M(21500) * 2 + m_3^M(21500) * 3 = 0 * 1 + 0 * 2 + 1 * 3 = 3,$$

$$S_{RFM}(34, 9, 21500) = 1.68 + 2.75 + 3 = 7.43$$

依此類推計算出其他顧客的 RFM 分數，分數愈高者，表示此顧客對企業來說是愈重要的顧客。

(三) 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

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 $\min fs$ ，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 $\min fs$ ，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 $FC(R) \geq \min fc$ ，則 R 為所要之有效的關聯法則現在我們以表 3.10 的顧客 RFM 資料表為例來說明如何找出 RFM 的關聯法則。首先，假設最初的最小模糊支持度為 0.32，由 RFM 資料方塊我們得到 1-itemsets 的模糊支持度，如表 3-16 所示：

表 3-16：RFM1-itemsets 的 FS

	R.L	R.M	R.H	F.L	F.M	F.H	M.L	M.M	M.H
FS	0.404	0.408	0.188	0.425	0.275	0.300	0.495	0.340	0.165

經過最小模糊支持度門檻值 (0.32) 的比對後得到 L_1 ，如下表 3-17：

表 3-17：RFM 高頻 1- itemsets (L_1)

	R.L	R.M	F.L	M.L	M.M
FS	0.404	0.408	0.425	0.495	0.340

接下來，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 3-18 如下：

表 3-18： C_2 的 FS

	R.L×F.L	R.L×M.L	R.L×M.M	R.M×F.L
FS	0.25	0.2292	0.1748	0.125
	R.M×M.L	R.M×M.M	F.L×M.L	F.L×M.M
FS	0.2294	0.1106	0.2125	0.18

若按照之前設定的門檻值來比對，則將沒有任何法則生，因此，為了方便說明，我們在此階段調整門檻值，假設為 0.2，則產生了 L_2 ，如下表 3-19：

表 3-19： L_2 的 FS

	R.L×F.L	R.L×M.L	R.M×M.L	F.L×M.L
FS	0.25	0.2292	0.2294	0.2125

然後，繼續進行合併， C_3 ，如下表 3-20：

表 3-20： C_3 的 FS

	R.L×F.L×M.L	R.M×F.L×M.L
FS	0.12705	0.08575

同樣地，假設此階段的門檻值為 0.07，則 L_3 與 C_3 相同。因此我們獲得了這兩個最大長度為三的項目組。接下來，從這些高頻項目組可產生 22 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：

$$\frac{FS(A_1^R \times A_1^F)}{FS(A_1^R)} = \frac{0.25}{0.404} = 0.62$$

依此計算出其他法則的結果，如下表 3-21 所示：

表 3-21： RFM 關聯法則表

法則	FC
R is 短 => F is 少	0.62
R is 短 => M is 小	0.57
F is 少=> M is 小	0.50
F is 少=> R is 短	0.59
M is 小=> R is 短	0.46
M is 小=> F is 少	0.43
R is 普通=> F is 少	0.31
F is 少=> R is 普通	0.29
R is 普通=>M is 少	0.56
M is 少=> R is 普通	0.46
R is 短 且 F is 少=> M is 小	0.51
M is 小=> R is 短 且 F is 少	0.26
F is 少 且 M is 小=> R is 短	0.60
R is 短=> F is 少 且 M is 小	0.31
R is 短 且 M is 小=> F is 少	0.55
F is 少=> R is 短 且 M is 小	0.30
R is 普通 且 F is 少=>M is 小	0.69
M is 小=> R is 普通 且 F is 少	0.17
R is 普通 且 M is 小=> F is 少	0.37
F is 少=> R is 普通 且 M is 小	0.20
M is 小 且 F is 少=> R is 普通	0.40
R is 普通=> M is 小 且 F is 少	0.21

接下來，計算這 22 條法則的信賴度是否符合所制定的最小模糊信賴度（0.5），如下表 3-22：

表 3-22： RFM 關聯法則表（符合最小信賴度）

法則	FC
R is 短 => F is 少	0.62
R is 短 => M is 小	0.57
F is 少=> M is 小	0.50
F is 少 => R is 短	0.59
R is 普通=>M is 少	0.56
R is 短 且 F is 少 => M is 小	0.51
F is 少 且 M is 小=> R is 短	0.60
R is 短 且 M is 小 => F is 少	0.55
R is 普通 且 F is 少=>M is 小	0.69

總共產生了 9 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長而且購買次數少且購買金額還不大，這表示此企業的顧客大部分都是新來的顧客居多，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新顧客有很好的購物經驗以提升顧客的購買次數進而讓顧客花更多金額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，此企業有可能是賣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，或是整個所有國家甚至全球經濟的蕭條。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來度過經濟的寒冬。

另外，就 RFM 三者的關聯部分，假設有一條法則是”購買次數高=>購買金額高”，則表示此企業必須要想辦法讓顧客提高購買的次數，因為從法則來分析此企業的顧客每次購物都是花費相當大的金額來購買產品居多。

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。

第四節 實驗與探討

本節中主要對於 Fuzzy RFM Model 以及過去學者所提的 RFM Model 進行比較，並提出程式執行結果的實驗數據與事後的分析與探討，提供較客觀的數據比較。本實驗實作了一個親和力高的使用介面 RFM 系統，所以我們使用了 Visual Basic 6.0 視窗程式設計軟體撰寫，軟體本身提供了視窗程式設計所必須用到的圖形介面工具與程式開發工具來開發此系統。

一、實驗測試環境：

為了實作 RFM Model，我們利用以下的環境來進行實作與測試：

(一) 實驗平台：

1. CPU：Intel Pentium 4 2.4GHz
2. RAM：DDR 256Mb，HD：Maxtro 80Gb Ultra DMA133
3. OS：Windows 2000 Professional
4. DB：Access 2002

(二) 實作之程式語言：

為了比較能較客觀、公平，我們均利用 Visual Basic 6.0 來實作使用者介面、Fuzzy RFM 分群、Fuzzy RFM 分數、Fuzzy RFM Association Rules、Stone RFM 分數、顧客五等分、行為五等分，來進行分析與比較。

(三) 實驗資料庫：

本實驗的資料來源為 Microsoft SQL Server 2000 中的範例資料庫—Food Mart 在 1997 年與 1998 年的交易資料庫，我們依據此範例資料庫進行轉換成以”顧客 ID”為主鍵的 RFM 資料表，分別為 1997 年 RFM 資料表、1998 年 RFM 資料表，做為 RFM Model 的資料來源。每個 RFM 資料表各有 10282 筆顧客的 RFM 資料，畫面如下圖 3-6：

The screenshot displays three data tables from the RFM Model Systems software. The first table, 'RFM Model Systems', lists customer IDs (ID) and their corresponding RFM scores (R, F, M). The second table, 'RFM Data', shows the calculated RFM scores for each customer. The third table, 'RFM Results', displays the final customer segmentation into clusters (e.g., 1, 2, 3, 4, 5) based on the RFM scores.

圖 3-6：RFM Model Systems 與 RFM 資料表

二、實驗設計、結果與分析：

(一) 分群實驗

傳統 RFM Model，每一位顧客最多只屬於某一群，但現實環境下，會有一些顧客是落在群界邊緣附近，而這些顧客其實對於其他群還是有其貢獻度。若按傳統 RFM Model 來分析，並不會去考慮到這些顧客在其他群之貢獻度而導致分析顧客特性時，會有不完整的現象發生。模糊集合在元素的歸類上允許某個範圍的灰色區域，因此，我們將透過簡單的實驗來比較傳統 RFM Model 與 Fuzzy RFM Model 在顧客分群上的績效，我們設計了以下實驗：

1. 實驗 1：門檻值 α -cut 的設定對於顧客所擁有的群數之影響

(1) 實驗目的：

Fuzzy RFM Model 中，每一位顧客可能同時屬於不同的群，其群數會受 α -cut 值影響並找出恰當的門檻值。

(2)實驗方法：

我們先探討 α -cut 值對於 Fuzzy RFM Model 分群中之顧客所屬群數的影響。在此，Fuzzy RFM Model 的 K_R ， K_F ， K_M 都設為 5。我們利用公式 10 算出每位顧客在各群的歸屬度，並且累計歸屬度超過 α -cut 值之群的數目，即為該顧客所屬之群數並將所屬群數相同之顧客數目累計。

(3)實驗結果：

如下表 3-23 與表 3-24

表 3-23： α -cut 值對於不同顧客擁有群數之總人數影響（1997 年 RFM 資料表）

α -cut \ 群數	0	1	2	3	4	5	6	7	8
0.0	0	4665	1	0	24	0	0	0	5592
0.25	0	4665	33	233	917	682	1044	894	1814
0.5	0	5078	2510	1172	931	0	0	0	0
0.6	266	7363	2597	50	6	0	0	0	0
0.7	1581	8291	410	0	0	0	0	0	0
0.8	3578	6704	0	0	0	0	0	0	0
0.9	5170	5112	0	0	0	0	0	0	0

表 3-24： α -cut 值對於不同顧客擁有群數之總人數影響（1998 年 RFM 資料表）

α -cut \ 群數	0	1	2	3	4	5	6	7	8
0.0	0	0	2460	0	188	0	0	0	7634
0.25	0	2460	306	370	1465	878	1751	1101	1951
0.5	0	3425	4038	1900	919	0	0	0	0
0.6	290	6452	3458	650	17	0	0	0	0
0.7	1652	7851	779	0	0	0	0	0	0

0.8	4126	6156	0	0	0	0	0	0	0
0.9	6816	3466	0	0	0	0	0	0	0

(4) 結果探討與分析：

由上二表，我們可以發現 α -cut 值愈高時，會造成顧客所屬群數有愈少的趨勢，甚至某些顧客不屬於任何群的情形。當 α -cut 值低於（含）0.5 時，所有的顧客至少屬於一群；當 α -cut 值設到 0.6 以上會有某些顧客都不屬於某一群的現象發生；當大於（含）0.8 時，所有的顧客至多屬於一群。綜合言之，Food Mart 有相當多的顧客其消費行為之特性是可分別隸屬於不同的群。透過 Fuzzy RFM Model 的分群方法，可以比傳統的 RFM Model 更能找出顧客隸屬於不同群的特性。這個特性對於企業經營者來說，可以更完整且全面地瞭解顧客，進而能夠制定較佳的行銷決策方案。

接下來，實驗 2 與 3 的部分，我們將從傳統的顧客五等分與行為五等分的分群方法所分的顧客群，其在 Fuzzy RFM Model 分群上的特性，來驗證 Fuzzy RFM Model 比顧客五等分與行為五等分在顧客特性分析上有較佳的效果。首先，實驗 2 的部分，先來探討 Food Mart 在傳統的顧客五等分與行為五等分的分群，其顧客所屬之群的狀態。接下來，透過實驗 2 的結果，我們找出代表性的群（屬於此群的人數最多或次多）中的所有顧客，再透過 Fuzzy RFM Model 的分群方法，來分析這些顧客所屬之群數。

2. 實驗 2：針對顧客五等分與行為五等分，找出代表的群

(1) 實驗目的：

探討 Food Mart 的顧客經過傳統的顧客五等分與行為五等分的分群後，各群之顧客人數的分佈狀態，並找出具代表性的群。

(2) 實驗方法：

我們先探討 Food Mart 經過傳統的顧客五等分與行為五等分的分群後，計算各群之顧客人數與其分佈狀態，並找出人數最多的群來當作我們所要找之代表性的群。

(3) 實驗結果：

如下圖 3-7 至圖 3-10

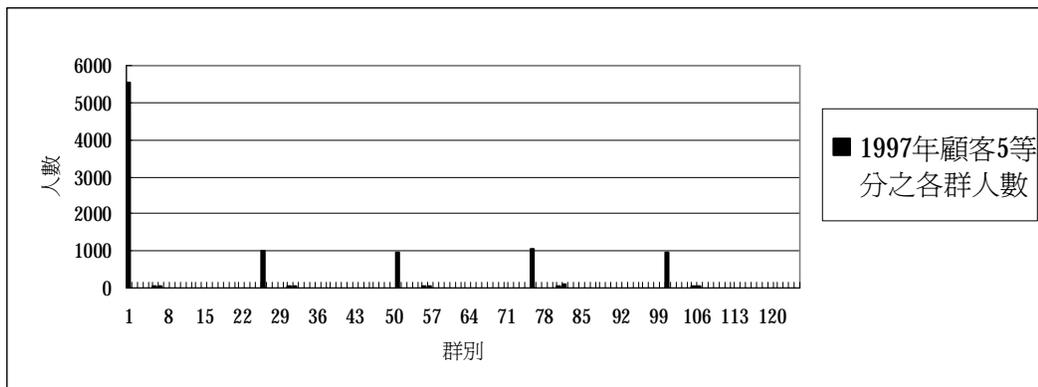


圖 3-7：1997 年顧客五等分之每一群之顧客人數

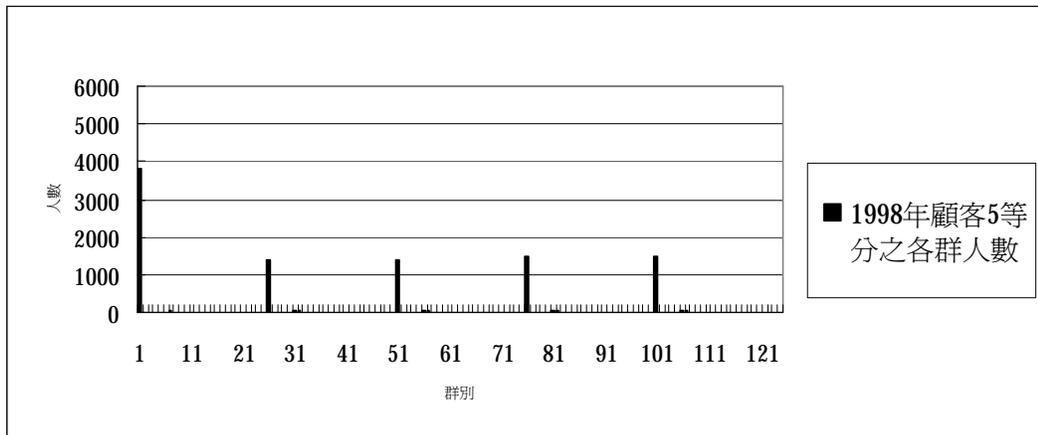


圖 3-8： 1998 年顧客五等分之每一群之顧客人數

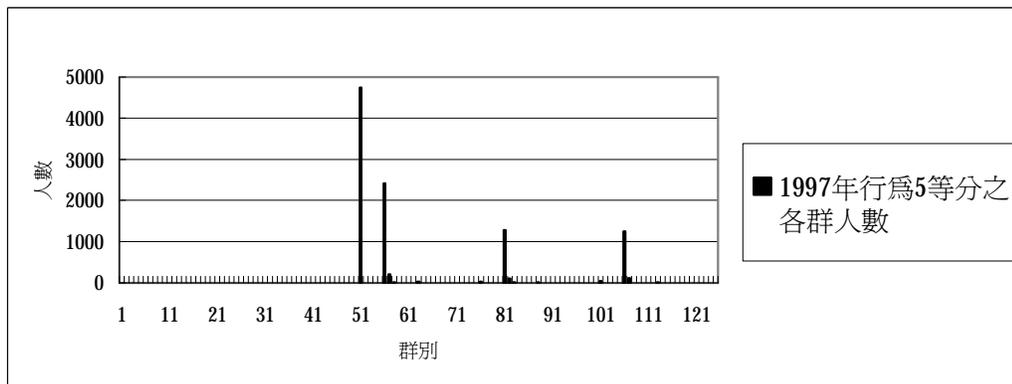


圖 3-9： 1997 年行為五等分之每一群之顧客人數

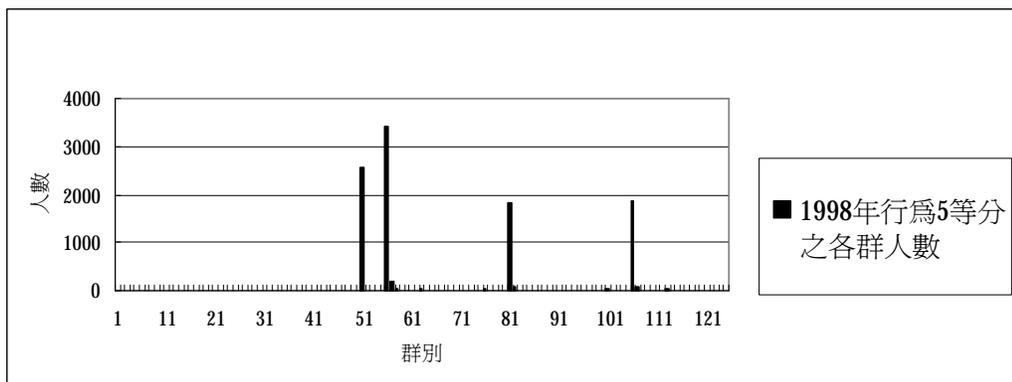


圖 3-10： 1998 年行為五等分之每一群之顧客人數

(4) 結果探討與分析：

從上四圖，我們可以發現 Food Mart 在顧客五等分中，大部分的顧客都集中於第 1 群顧客群（第 1 群顧客群在 1997 年有 5566 個顧客而在 1998 年有 3794 個顧客）；而在行為五等分，則分別在第 51 群與第 56 群（1997 年第 51 群有 4742 個顧客、第 56 群有 2412 個顧客而 1998 年第 51 群有 2544 個顧客、第 56 群有 3420 個顧客）。

3. 實驗 3：透過實驗 2 所找出來的代表群中之所有顧客，在 Fuzzy RFM Model 分群上的特性

(1) 實驗目的：

驗證顧客五等分與行為五等分的分群方法較無法完整且全面地瞭解顧客。

(2) 實驗方法：

利用實驗 2 所找出來的代表性顧客（顧客五等分：第 1 群；行為五等分：第 51 與 56 群）後，由這些顧客去對應其在 Fuzzy RFM Model 是否有歸屬於不同群的現象。以及計算其歸屬不同群的人數多寡。

(3) 實驗結果：

如下表 3-25 至表 3-30

表 3-25： 1997 年顧客五等分（第 1 群）在 Fuzzy RFM Model 分群上之狀態

α -cut \ 群數	0	1	2	3	4	5	6	7	8
0.0	0	4665	0	0	0	0	0	0	901
0.25	0	4665	6	34	94	104	187	134	343
0.5	0	4708	454	251	153	0	0	0	0

表 3-26： 1998 年顧客五等分（第 1 群）在 Fuzzy RFM Model 分群上之狀態

α -cut \ 群數	0	1	2	3	4	5	6	7	8
0.0	0	0	2459	0	37	0	0	0	1298
0.25	0	2458	56	46	165	112	364	191	402
0.5	0	2538	777	318	161	0	0	0	0

表 3-27： 1997 年行為五等分（第 51 群）在 Fuzzy RFM Model 分群上之狀態

α -cut \ 群數	0	1	2	3	4	5	6	7	8
0.0	0	4665	0	0	0	0	0	0	77
0.25	0	4665	9	54	13	1	0	0	0
0.5	0	4685	57	0	0	0	0	0	0

表 3-28： 1998 年行為五等分（第 56 群）在 Fuzzy RFM Model 分群上之狀態

α -cut \ 群數	0	1	2	3	4	5	6	7	8
0.0	0	0	0	0	0	0	0	0	3420
0.25	0	0	79	141	623	373	839	491	876
0.5	0	422	1786	829	383	0	0	0	0

表 3-29： 1997 年行為五等分（第 56 群）在 Fuzzy RFM Model 分群上之狀態

α -cut \ 群數	0	1	2	3	4	5	6	7	8
0.0	0	0	0	0	0	0	0	0	2412
0.25	0	0	2	43	356	307	463	390	851
0.5	0	161	1149	779	423	0	0	0	0

表 3-30：1998 年行為五等分（第 51 群）在 Fuzzy RFM Model 分群上之狀態

α -cut \ 群數	0	1	2	3	4	5	6	7	8
0.0	0	0	0	0	0	0	150	0	2394
0.25	0	0	85	119	426	219	387	342	966
0.5	0	445	1023	580	273	223	0	0	0

(4) 結果探討與分析：

由上六表，我們可以發現在 1997 年的顧客五等分與行為五等分之最大顧客群之顧客，在 Fuzzy RFM Model 分群上主要是以顧客只歸屬於一群的較多。這表示 1997 年的顧客消費特性之分群上，用 Fuzzy RFM Model、顧客五等分法或是行為五等分法，其效果是差不多的。但就 1997 年行為五等分之次多的顧客群（第 56 群）以及 1998 年的顧客五等分與行為五等分來作分析，我們可以很明顯地發現用 Fuzzy RFM Model 比顧客五等分法或是行為五等分法來的效果較佳。理由是，這些顧客群之顧客，有很多顧客在 Fuzzy RFM Model 還可以分別歸屬於一個以上之不同群。大體上而言，尤其在 1998 年的顧客在 Fuzzy RFM Model 的分群上比顧客五等分法或是行為五等分法更能反映出落在群界邊緣附近的顧客之特性。

(二) RFM 分數實驗

企業在進行行銷決策時，必須對顧客的特性要有更精確的瞭解，而在 RFM 分數上若能夠對顧客之持續購買能力作較精細的

區分，在客戶的選擇上將更能夠接近所預期效果。例如，共有 1000 位顧客，若兩種不同的 RFM 分數各分別出 10 等級（平均每等級 100 名顧客）以及 100 等級（平均每等級 10 名顧客），在行銷決策上，每 10 名顧客之級距比每 100 顧客之級距較易制定且較具實質效果。因此，RFM 分數的鑑別能力是評量 RFM 分數是否能夠對顧客之持續購買能力作較精細地區分之重要指標。除了 RFM 分數鑑別能力的分析之外，在 RFM 分數上還有一個重要的考量，那就是 RFM 分數在分數相同但其所代表意義是不同的情形發生機會。這種狀況會讓決策者在進行決策時，導致誤判的結果發生。因此，在 RFM 分數的實驗部分，我們將設計六項實驗來比較各個 RFM 分數的鑑別能力以及同分但其意義不相同的發生次數來證明 Fuzzy RFM 分數比其他 RFM 分數有更好的效果。在此，Fuzzy RFM Model 的 K_R, K_F, K_M 都設為 3； $w_1^R=6$ (R.L)； $w_2^R=3$ (R.M)； $w_3^R=1$ (R.H)； $w_1^F=1$ (F.L)； $w_2^F=3$ (F.M)； $w_3^F=6$ (F.H)； $w_1^M=1$ (M.L)； $w_2^M=3$ (M.M)； $w_3^M=6$ (M.H)。實驗分別如下：

1.實驗 1：計算傳統 RFM 分數與 Fuzzy RFM 分數在人數的分佈狀況以及 RFM Model 擁有多少不同的分數，來分析傳統 RFM 分數與 Fuzzy RFM 分數在整體鑑別度。其鑑別度判斷是依據 RFM 分數在人數的分佈狀況是否趨近於均勻分佈並且擁有多少不同的分數，分佈愈均勻且擁有的分數愈多則整體鑑別度愈高。

(1)實驗目的：

透過 RFM 分數在人數的分佈狀況以及計算每一個 RFMModel 用多少不同的分數來描述 10282 位顧客，來分析傳統 RFM 分數與 Fuzzy RFM 分數在整體鑑別度。

(2)實驗方法：

計算顧客在傳統 RFM 與 Fuzzy RFM 每一個分數上之人數、計算每一個 RFMModel 擁有多少不同的分數並且計算每一個 RFMModel 人數最少與最多人的分數之人數、每個分數平均人數以及 Entropy 值計算比較。在此，Entropy 值的公式如下：

$$Entropy = \sum_i^n -p_i \log p_i \dots\dots\dots (公式 17)$$

(3)實驗結果：

如下表 3-31 至表 3-35 與圖 3-11 至圖 3-18

表 3-31：1997 年顧客五等分各個分數之顧客人數表

分數	3	4	5	6	7	8	9	10	11	12	13	14	15
人數	5567	1055	1043	1202	1074	185	94	29	15	11	4	1	2

圖 3-11：1997 年顧客五等分各個分數之顧客人數狀況圖

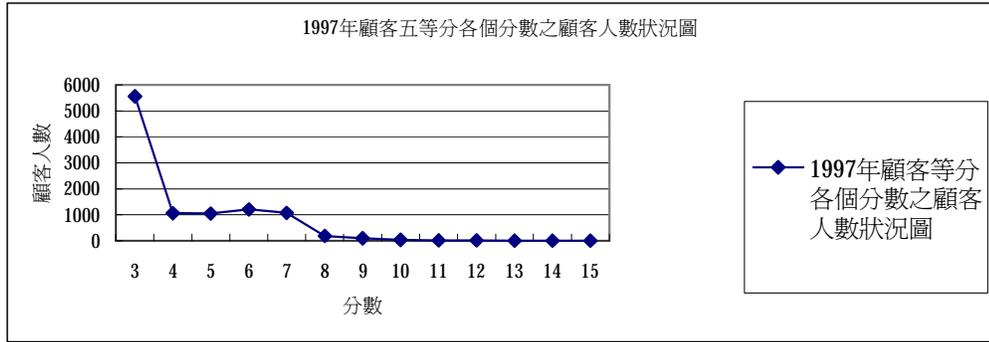


表 3-32：1998 年顧客五等分各個分數之顧客人數表

分數	3	4	5	6	7	8	9	10	11	12	13	14	15
人數	3794	1427	1473	1609	1616	144	95	55	36	21	10	2	0

圖 3-12：1998 年顧客五等分各個分數之顧客人數狀況圖

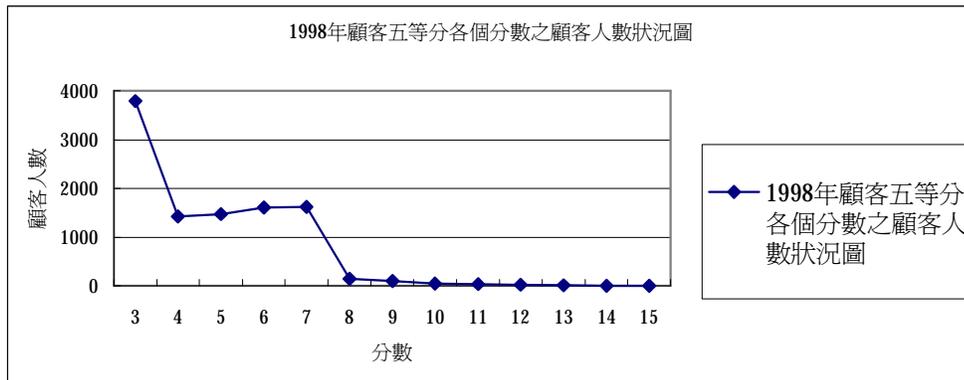


表 3-33：1997 年行為五等分各個分數之顧客人數表

分數	3	4	5	6	7	8	9	10	11	12	13	14	15
人數	0	0	4742	2442	1520	1371	154	23	14	12	2	1	1

圖 3-13：1997 年行為五等分各個分數之顧客人數狀況圖

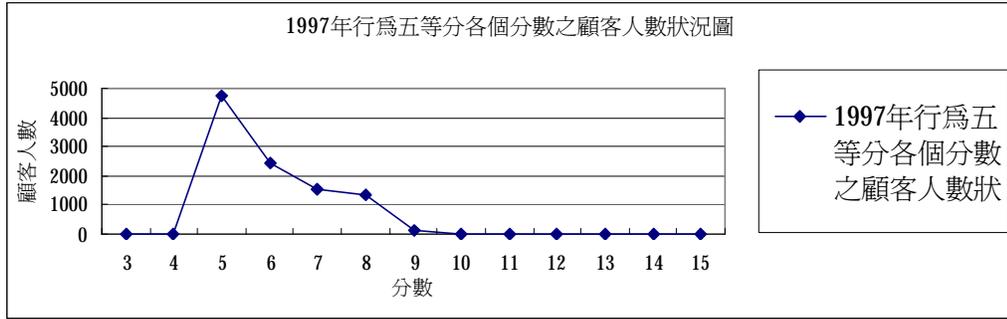


表 3-34：1998 年行為五等分各個分數之顧客人數表

分數	3	4	5	6	7	8	9	10	11	12	13	14	15
人數	0	0	2544	34589	2044	1972	150	45	49	14	4	2	0

圖 3-14：1998 年行為五等分各個分數之顧客人數狀況圖

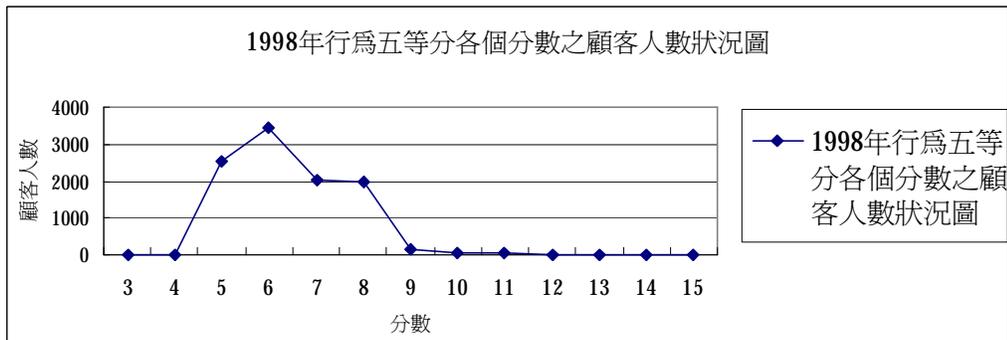


圖 3-15：1997 Stone RFM 各個分數之顧客人數狀況圖

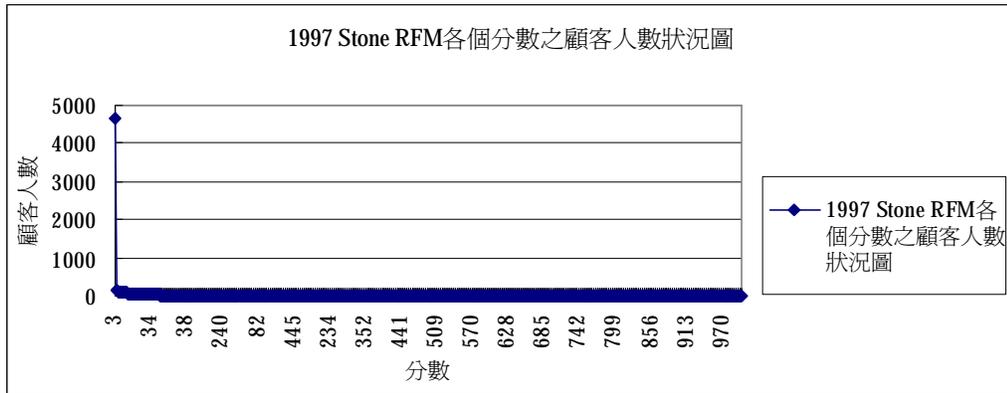


圖 3-16：1998 Stone RFM 各個分數之顧客人數狀況圖

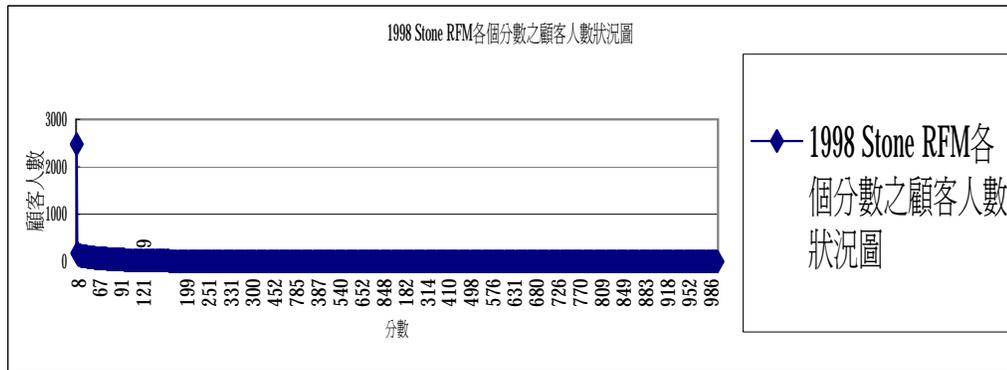


圖 3-17：1997 Fuzzy RFM 各個分數之顧客人數狀況圖

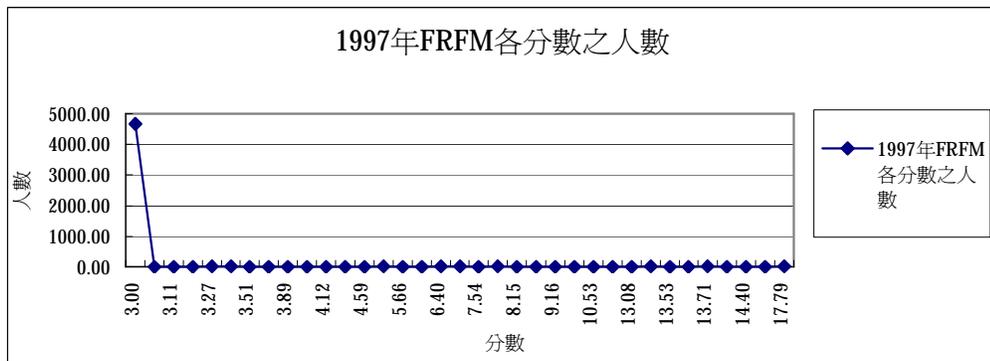


圖 3-18：1998 Fuzzy RFM 各個分數之顧客人數狀況圖

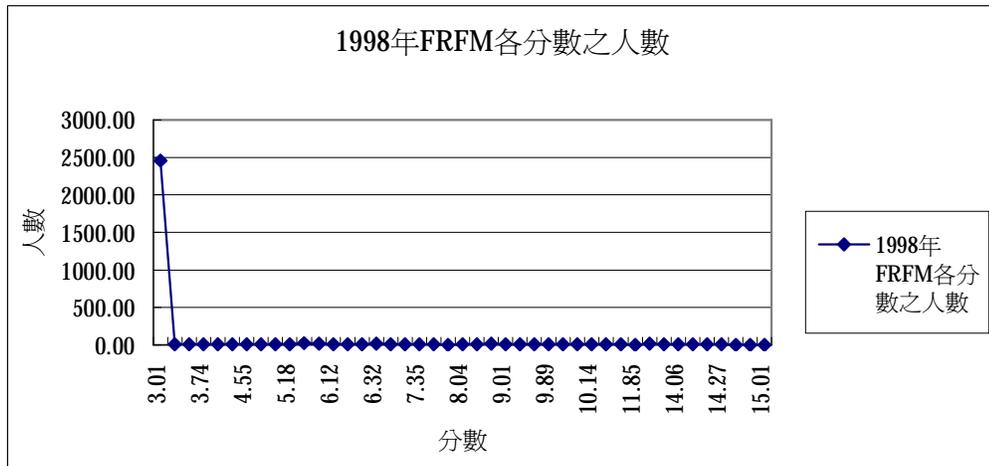


表 3-35：RFMModel 分數綜合比較表

RFMModel	顧客五等分		行為五等分		Stone RFM		Fuzzy RFM	
	1997	1998	1997	1998	1997	1998	1997	1998
年份	1997	1998	1997	1998	1997	1998	1997	1998
不同分數的個數 (range)	13	12	11	10	380	792	983	1067
人數最少之分數的人數	1	2	1	2	1	1	1	1
人數最多之分數的人數	5567	3794	4742	3458	4665	2481	4665	2458
每個分數之平均人數	791	857	934	1028	27	13	10	10
Entropy 值	0.58479	0.64	0.625	0.727	1.460	1.938	1.787	2.164

(4) 結果探討與分析：

我們可以由表 3-31 至表 3-34 與圖 3-11 至圖 3-18 發現 Stone RFM 與 Fuzzy RFM 在各個分數之顧客人數分佈狀況較顧客五等分與行為五等分均勻。因此，Stone RFM 與 Fuzzy RFM 的整體鑑別度相較於顧客五等分與行為五等分具較佳的效果。另一方面，我們也可以發現 RFM 分數較低的顧客人數比其他分數所擁有顧客人數來得多。這表示有很多顧客對 Food Mart 的持續性購買價值而言，其貢獻度是相當低。因此，這結果是作為 Food Mart 的經營決策者所該注意的。

由上表 3-35，不論是從 Entropy 值（愈大愈好）或是每個分數之平均人數（愈小愈好）我們可以發現 Fuzzy RFM 分數有較佳的整體鑑別能力，Stone RFM 次之，顧客五等分與行為五等分在整體鑑別度表現上較不理想。總體而言，Fuzzy RFM 分數比其他的 RFM 分數在整體鑑別度上有較佳的成效。

2.實驗 2：透過實驗 1，我們可以發現 Fuzzy RFM 分數比其他的 RFM 分數在整體鑑別度上有較佳的成效。接下來，實驗 2 的部分將探討 Fuzzy RFM 與 Stone RFM 在區域鑑別度上的比較。

(1)實驗目的：

探討 Fuzzy RFM 在區域鑑別度上是否有優於 Stone RFM。

(2)實驗方法：

由於 Stone RFM 最多人數的分數（1997 年為 3 分；1998 年為 8 分）都只有一種 RFM 組合，所以我們依據 Stone RFM 次多人數之分數（1997 年 45 分有 150 人；1998 年 40 分有 179 人）的顧客群，其 R、F、M 三個值的分佈狀況來比較。

(3)實驗結果：

如下圖 3-19 至圖 3-31

圖 3-19：1997 年 StoneRFM 次多人數之分數之顧客其各 R 值人數直條圖

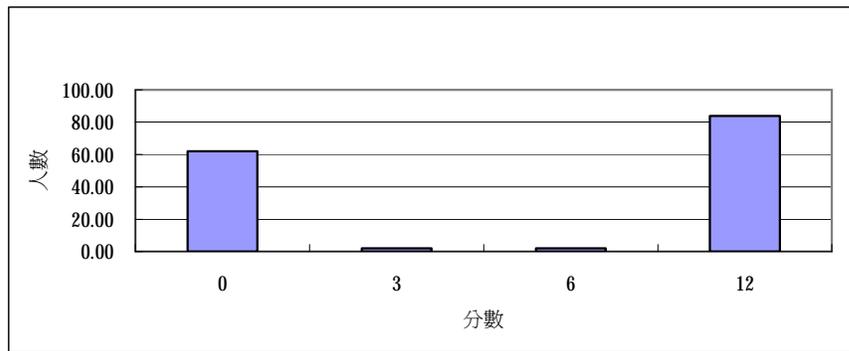


圖 3-20：1997 年 StoneRFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 R 值人數直條圖

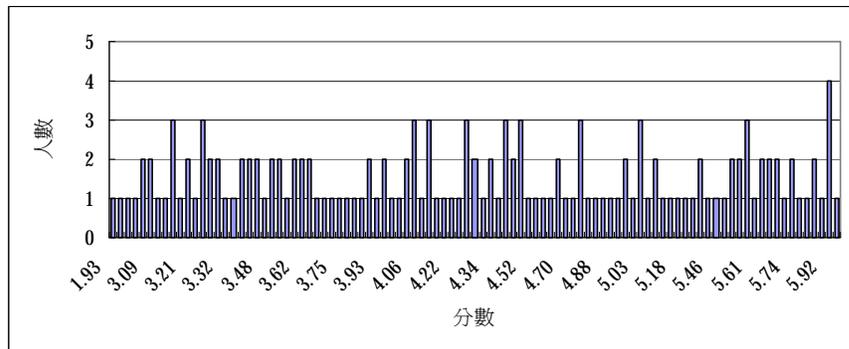


圖 3-21：1997 年 StoneRFM 次多人數之分數之顧客其各 F 值人數直條圖

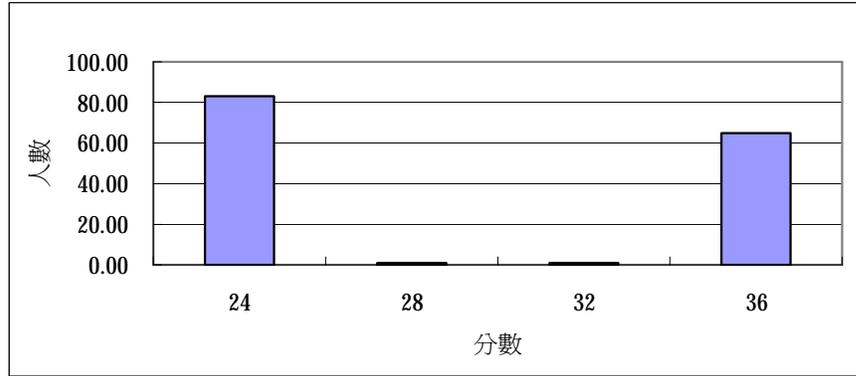


圖 3-22：1997 年 StoneRFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 F 值人數直條圖

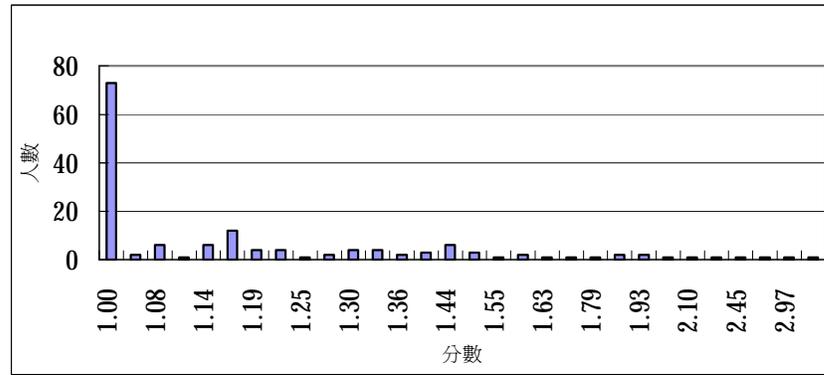


圖 3-23：1997 年 StoneRFM 次多人數之分數之顧客其各 M 值人數直條圖

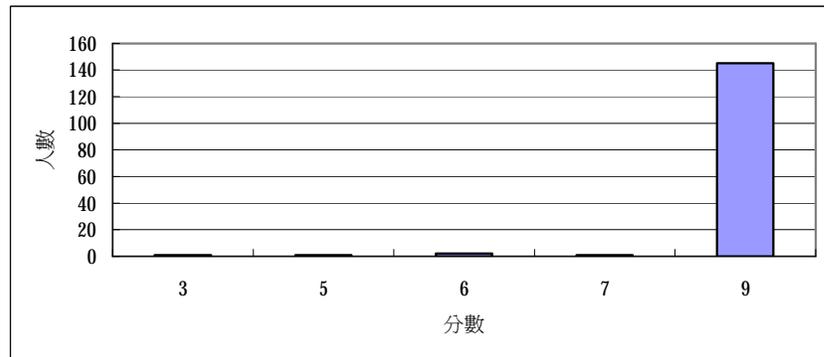


圖 3-24：1997 年 StoneRFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 M 值人數直條圖

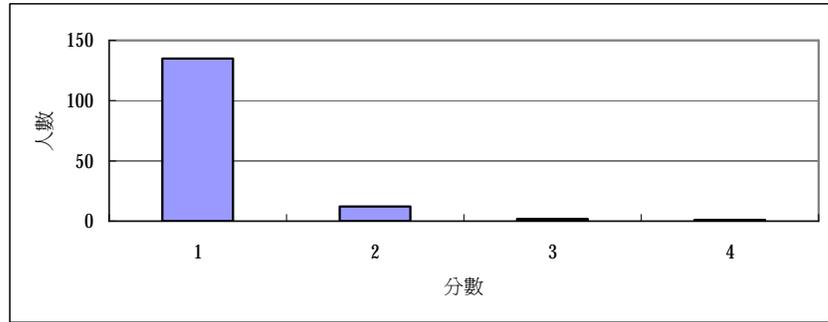


圖 3-25：1998 年 StoneRFM 次多人數之分數之顧客其各 R 值人數直條圖

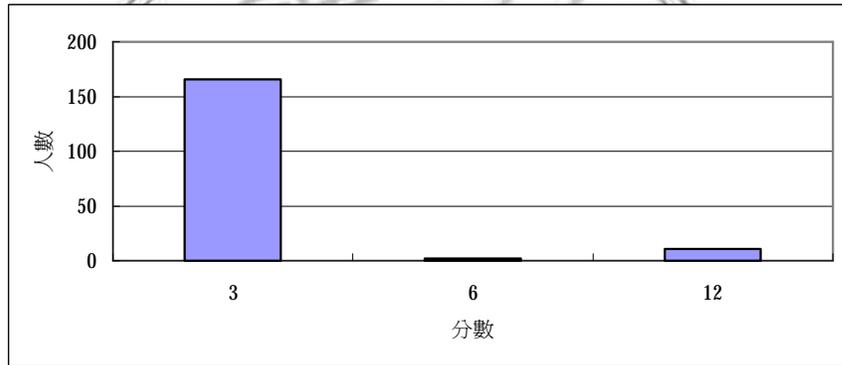


圖 3-26：1998 年 StoneRFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 R 值人數直條圖

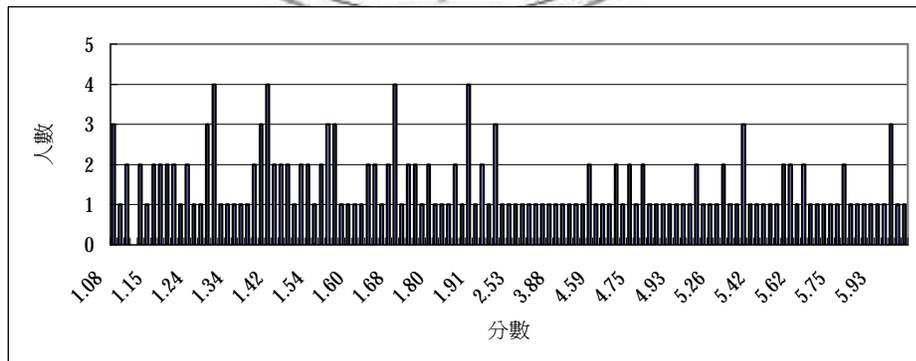


圖 3-27：1998 年 StoneRFM 次多人數之分數之顧客其各 F 值人數直條圖

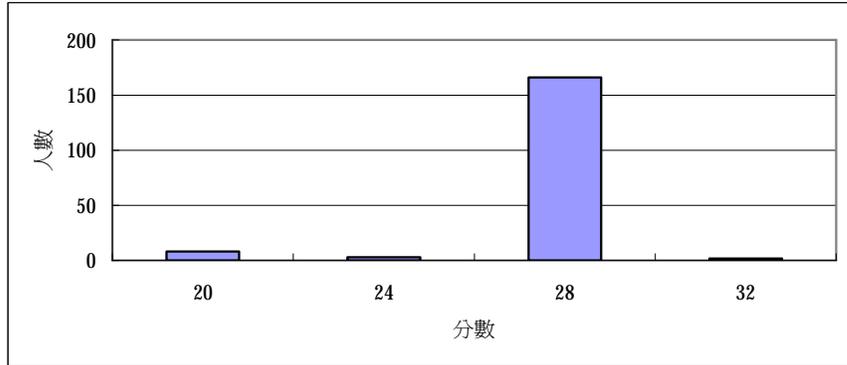


圖 3-28：1998 年 StoneRFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 F 值人數直條圖

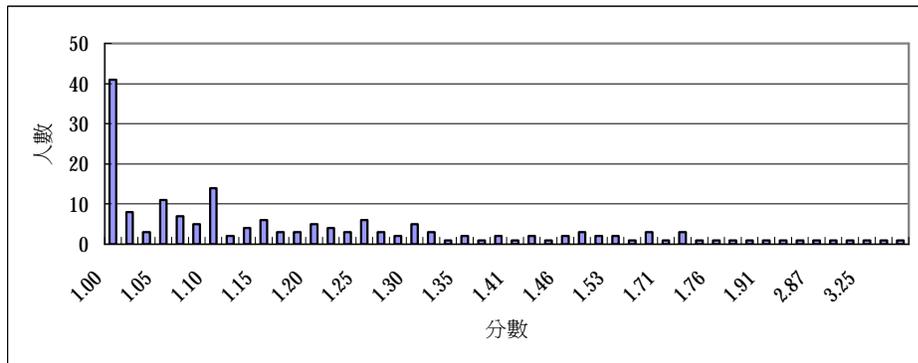


圖 3-29：1998 年 StoneRFM 次多人數之分數之顧客其各 M 值人數直條圖

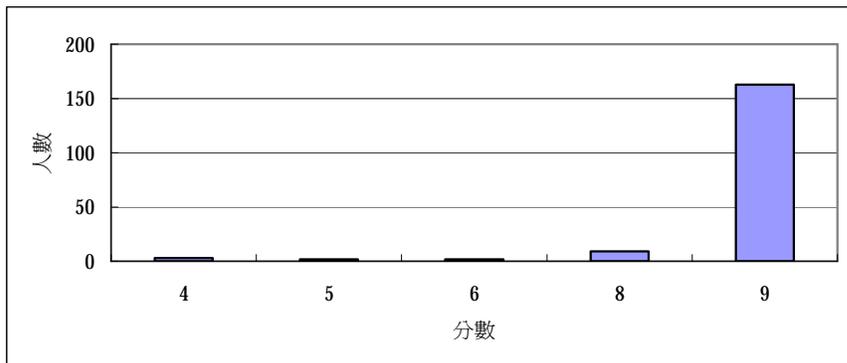
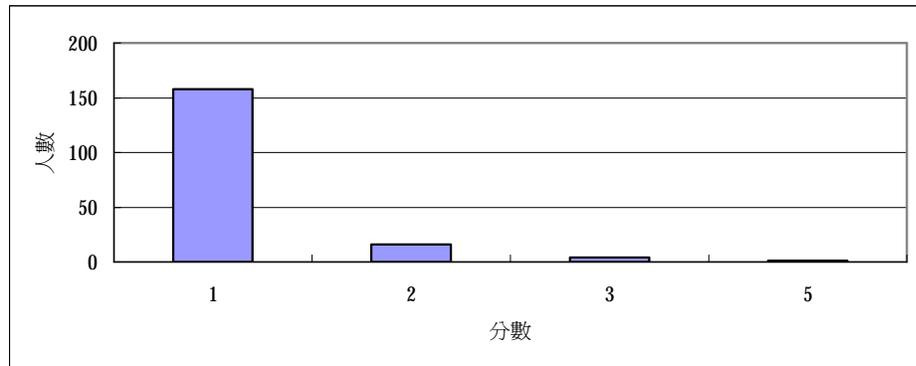


圖 3-30：1998 年 StoneRFM 次多人數之分數之顧客其在 Fuzzy RFM 之各 M 值人數直條圖



(4) 結果探討與分析：

由上圖我們可以發現在 Stone RFM 次多人數之分數的顧客群，其 R、F、M 三個值分別在 Stone RFM 與 Fuzzy RFM 的分佈狀況，除了 M 值分佈狀況類似之外，R 值與 F 值有明顯的差異。Stone RFM 的 R 值與 F 值較為集中落在少數個不同的值，而 Fuzzy RFM 的 R 值與 F 值落在多個不同的值。因此，Fuzzy RFM 在區域鑑別度較 Stone RFM 具較佳效果。

3.實驗 3：Fuzzy RFM 分數與傳統 RFM 分數在同分但其意義不同之發生次數

(1)實驗目的：

比較 Fuzzy RFM 分數以及傳統 RFM 分數發生同分但其意義不同的狀況的次數。

(2)實驗方法：

計算每一個 RFM 分數發生同分但其所代表之意義不同的情形之個數。例如，以 1997 年顧客五等分為例（如表 3-37），4 分有 (1,1,2)、(1,2,1)、(2,1,1) 這三個組合，而在 4 分的顧客都有這三種組合，表示 4 分發生同分且意義不同的情形之個數為 3。其他分數也依此類推算出同分且意義不同的情形之個數並加總。

(3) 實驗結果：

如下表 3-36

表 3-36：同分且意義不同之發生次數比較表

RFM 分數 \ 年份	1997	1998
Fuzzy RFM	0	0
Stone RFM	38	41
顧客五等分法	48	56
行為五等分法	25	28

(4) 結果探討與分析：

由上表，我們可以發現 Fuzzy RFM 分數並無發現同分且其意義不同的情形，相較傳統 RFM 分數其發生機會非常低，而顧客五等分發生的次數最多。

4.實驗 4：顧客五等分與行為五等分法在顧客人數超過 1000 人之分數以及 Stone RFM 次多顧客人數之分數其發生同分但其意義不同之發生次數。

(1)實驗目的：

探討顧客五等分、行為五等分法與 Stone RFM 在區域分數上，其發生同分但其意義不同之發生次數。

(2)實驗方法：

依據表 3-31 至表 3-34，找出人數超過 1000 人的分數以及 Stone RFM 次多顧客人數之分數並列出每個分數之 RFM 組合與組合所擁有之人數。

(3)實驗結果：

如下表 3-37 至表 3-41

表 3-37：1997 年顧客五等分分數之不同組合人數

分數	3 共 0 個同分意義不同組合														
RFM 組合	111														
人數	5567														
分數	4 共 3 個同分意義不同組合														
RFM 組合	112	121	211												
人數	4	46	1005												
分數	5 共 4 個同分意義不同組合														
RFM 組合	113	122	131	212	221	331									
人數	0	63	0	6	39	935									
分數	6 共 6 個同分意義不同組合														
RFM 組合	114	123	132	141	213	222	231	312	321	411					
人數	0	4	10	0	0	72	0	6	59	1051					
分數	7 共 6 個同分意義不同組合														
RFM 組合	115	124	133	142	151	214	223	232	241	313	322	331	412	421	511
人數	0	0	9	0	0	0	0	10	0	0	65	0	3	50	938

表 3-38：1998 年顧客五等分分數之不同組合人數

分數	3 共 0 個同分意義不同組合														
RFM 組合	111														
人數	3794														
分數	4 共 3 個同分意義不同組合														
RFM 組合	112	121	211												
人數	6	23	1398												
分數	5 共 4 個同分意義不同組合														
RFM 組合	113	122	131	212	221	331									
人數	0	36	0	1	27	1409									
分數	6 共 6 個同分意義不同組合														
RFM 組合	114	123	132	141	213	222	231	312	321	411					
人數	0	1	21	0	0	68	0	5	30	1484					
分數	7 共 6 個同分意義不同組合														
RFM 組合	115	124	133	142	151	214	223	232	241	313	322	331	412	421	511
人數	0	0	12	0	0	0	0	20	0	0	63	0	10	44	1467

表 3-39：1997 年行為五等分分數之不同組合人數

分數	5 共 0 個同分意義不同組合																	
RFM 組合	113	122	131	212	221	311												
人數	0	0	0	0	0	4742												
分數	6 共 2 個同分意義不同組合																	
RFM 組合	114	123	132	141	213	222	231	312	321	411								
人數	0	0	0	0	0	0	0	0	2412	30								
分數	7 共 3 個同分意義不同組合																	
RFM 組合	115	124	133	142	151	214	223	232	241	313	322	331	412	421	511			
人數	0	63	0	6	39	0	0	0	0	0	210	0	1273	0	37			
分數	8 共 6 個同分意義不同組合																	
RFM 組合	125	134	143	152	215	224	233	242	251	314	323	332	341	413	422	431	512	521
人數	0	0	0	0	0	0	0	0	0	0	16	3	0	0	99	0	0	1253

表 3-40：1998 年行為五等分分數之不同組合人數

分數	5 共 0 個同分意義不同組合																	
RFM 組合	113	122	131	212	221	311												
人數	0	0	0	0	0	2544												
分數	6 共 2 個同分意義不同組合																	
RFM 組合	114	123	132	141	213	222	231	312	321	411								
人數	0	0	0	0	0	0	0	0	3420	38								
分數	7 共 3 個同分意義不同組合																	
RFM 組合	115	124	133	142	151	214	223	232	241	313	322	331	412	421	511			
人數	0	0	0	0	0	0	0	0	0	0	188	0	0	1819	37			
分數	8 共 6 個同分意義不同組合																	
RFM 組合	125	134	143	152	215	224	233	242	251	314	323	332	341	413	422	431	512	521
人數	0	0	0	0	0	0	0	0	0	0	20	11	0	0	92	0	0	1849

表 3-41：Stone RFM 分數之不同組合人數

	1997 年		1998 年	
Stone RFM	45 分發生同分且意義不同的 RFM 組合數	150 人	40 分發生同分且意義不同的 RFM 組合數	179 人
RFM 組合	0, 36, 9	61 人	3, 28, 9	166 人
	3, 36, 6	2 人	6, 28, 6	2 人
	6, 36, 3	1 人	12, 20, 8	8 人
	6, 32, 7	1 人	12, 24, 4	3 人
	12, 24, 9	85 人		

(4) 結果探討與分析：

由上表，我們可以發現在人數超過 1000 人之分數，顧客五等分發生同分且意義不同的 RFM 組合數量（1997 年與 1998 年組合數量為 19）較行為五等分（1997 年與 1998 年組合數量為 11）多。由此，我們可以發現顧客五等分發生同分且意義不同的機會非常高。而 Stone RFM 在次多顧客人數之分數上，我們也可以發現有發生同分且意義不同的狀況。

三、實驗結論：

在 RFM 分群上，在現實環境下，會有一些顧客是落在群界邊緣附近，而這些顧客對於其他群還是有其貢獻度。透過實驗，我們證明了 Fuzzy RFM Model 在分群上比其他傳統 RFM Model 更能夠表現出落在群界邊緣附近顧客之特性。而在 RFM 分數實驗中，Fuzzy RFM Model 不論是 RFM 分數的鑑別能力以及同分而產生錯誤之狀況都比其他 RFM 分數的效果來得更具成效。

第五節 結論

我們利用模糊理論的概念來解決傳統單層次關聯法則演算法所無法處理的數值型資料之外，我們並結合模糊切割與資料方塊技術，來建構一個功能完整的單層次關聯法則演算法。此演算法能夠解決現實資料庫經常性變動以及使用者可以隨時調整門檻值所造成的問題。

而在其應用的部分，本研究利用模糊切割之單層次關聯法則演算法來建構一個新的 RFM Model—Fuzzy RFM Model，來改進過去

傳統 RFM Model 的缺失，例如，無法量化去做客觀的判斷、有些變化的組合很難去解釋、以及變化的幅度未考量等。我們利用 Fuzzy partition 的觀念來建立 Fuzzy RFM Model 的第一階段，當第一階段完成後，就可以在第二階段做很多應用，像顧客分群以及計算每一個顧客的 RFM 分數。其中特別的應用是，這個 Fuzzy RFM Model 也可以結合關聯法則來進行 R、F、M 這三個參數彼此之間的關聯性探討，這是目前在許多 RFM 研究上非常少見的，但實際上探討 R、F、M 這三個參數彼此之間的關聯性是有其必要性。從這三個變數之關聯性探討，企業經營主管可以從中獲得更多隱藏的資訊來更深一層地瞭解顧客的需求與行為以及市場上的特性與變化，以提供企業經營主管更容易且更精確地制定最佳的決策。另外，3 維 Data cube 的表示法可以快速地找出 RFM 的模糊關聯法則，透過這樣的儲存方式並可支援 on-line mining 與 incremental mining 的功能[1,11]。

本研究所提出 Fuzzy RFM Model 的方法，還有許多地方是值得未來去深入探討的。例如，(a)在模糊理論中有相當多種模糊歸屬度函數，不同的模糊隸屬函數可能會對 Fuzzy RFM Model 產生不同影響；以及(b)R、F、M 中所切割的數量 K 要各設多少會有比較好的鑑別度等，未來我們將實際深入探討這些層面的影響。另外，本研究所提的 Fuzzy RFM Model 可實際應用在很多電子商務系統的經營與行銷上。未來，我們將結合其他技術[12,14]以便廣泛地應用到電子商務系統的个人化行銷、線上推薦與增進顧客價值等功能上。

第四章 模糊切割在多階層關聯法則之研究

本章我們將介紹如何利用模糊切割的概念來建構一多階層關聯法則。首先，在第一節部分，我們將介紹如何建構這個新的方法以及其流程，並以一範例說明之；在第二節部分，我們將介紹這新的方法在漸進式多階層關聯法則與線上挖掘等特殊問題上的應用；第三節介紹模糊切割之關聯法則在漸進式挖掘與線上挖掘的績效實驗比較；最後第四節部分，為此章之結論與討論。

第一節 模糊切割於多階層關聯法則之建構

一、問題描述

在過去多階層關聯法則的研究中，如同第三章單層次關聯法則在數值型資料、動態資料庫的漸進式挖掘以及線上挖掘等問題上都有其限制。雖然近幾年，有學者針對這三個問題的一兩點問題來進行解決，但能夠完全處理這三個問題的多階層關聯法則演算法並不多見。因此，我們還是利用模糊切割來處理數值型資料，以及資料方塊的結構優勢來處理漸進式挖掘與線上挖掘等問題。當我們在商品項目進行模糊切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（Item），如此可以利用 Apriori-like 演算法擷取商品項目之間的模糊關聯法則[11]。但當資料庫中的資料量很龐大時，Apriori-like 演算法[34,37]是相當費時的。我們可以事先計算這些項目組的模糊支持度（Fuzzy support）值並儲存在資料方塊（Data cube）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

(On-line mining) 的功能[11]，而當新增紀錄時，亦可即時加以處理並更新資料方塊 (Data cube) 中對應項目組的支持度，以達到漸進式挖掘(Incremental mining)的目的[11]。

二、模糊切割之多階層關聯法則演算法

假定我們在每個商品項目上分割之模糊集合個數為 K ；使用者所設定的門檻值分別為最小模糊支持度 α 與最小模糊信賴度 β ，則模糊切割之多階層關聯法則的步驟說明與步驟流程圖4-1如下：

- 步驟 1 從交易資料庫中的原始交易表格，透過分類樹來增加祖先項目，並且計算祖先項目的數量（與此祖先相關的分支項目的數量加總）來建立一個含有祖先項目的交易表
- 步驟 2 透過模糊歸屬函數來計算每一個產品項目以及產品之祖先項目的歸屬度值
- 步驟 3 計算每一個模糊集合的歸屬度總值，而得到每一個模糊集合（長度為1之候選項目 C_1 ）的支持度（ $FS(C_1)$ ）
- 步驟 4 並把步驟3的支持度儲存在資料方塊中
- 步驟 5 依據每一個 C_1 進行合併，成為長度為2之候選項目組 C_2 ，並計算長度為2之候選項目組在每一筆之歸屬度值（利用公式9）並加總（ $FS(C_2)$ ）。最後儲存在候選項目為二之資料方塊中
- 步驟 6 重複步驟5之方法直到產生項目長度為 k 之候選項目組 C_k ，而無法合併產生項目長度為 $(k+1)$ 之候選項目組為止

步驟 7 從每一個 C_i 的模糊支持度到 C_k 的模糊支持度與最小模糊支持度 α 進行比對，符合 $(FS(C_i) \geq \alpha | t=1 \text{ to } k)$ 的項目組為大項目組

步驟 8 由大項目組來產生候選法則

步驟 9 計算候選法則的模糊信賴度 $FC(R)$ 並與最小模糊信賴度 β 進行比對，符合 $(FC(R) \geq \beta)$ 的法則為最終符合我們需求的法則

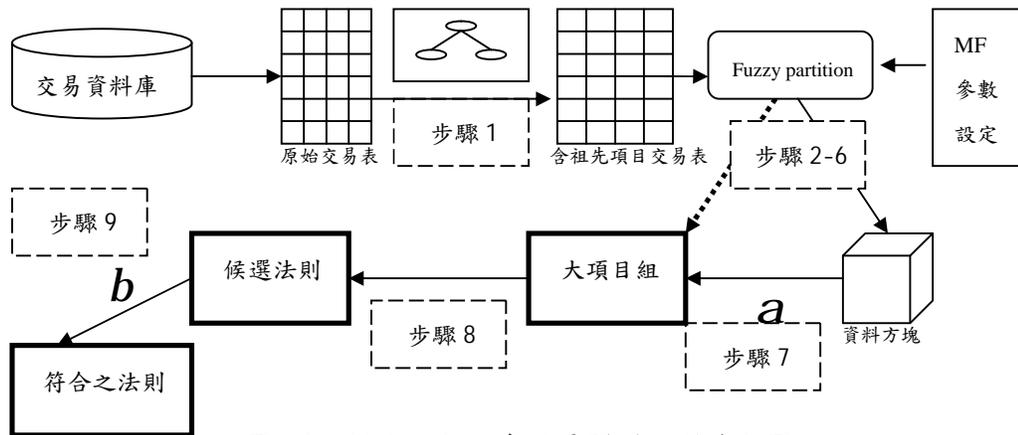


圖4-1：模糊切割之多階層關聯法則流程圖

三、範例說明

接下來我們將利用一個範例來說明模糊切割之多階層關聯法則的建構過程。首先交易資料表4-1如下

表4-1：交易資料表

交易ID	項目
1	(milk,3)(bread,4)(cookie,1)
2	(juice,3)(bread,7)(cookie,7)
3	(milk,2)(juice,10)(cookie,5)
4	(milk,7)(juice,2)(bread,8)(cookie,8)
5	(juice,2)(bread,8)(cookie,10)

其商品項目分類規則與模糊歸屬函數分別如下圖 4-2 與圖 4-3

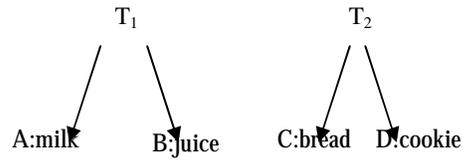


圖 4-2: 商品項目分類圖

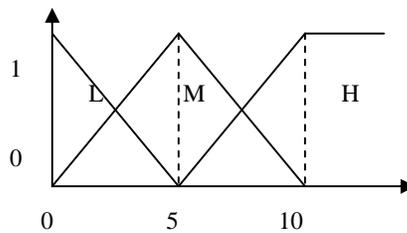


圖 4-3: 模糊歸屬函數

假設我們所設定的最小模糊支持度 $\alpha=1.6$ 與最小模糊信賴度 $\beta=0.7$ 。接下來開始進行我們的模糊切割之多階層關聯法則的建構。

步驟1 從表4.1中透過分類規則來增加祖先項目 T_1 與 T_2 並計算 T_1 與 T_2 的數量（例如第一筆的 T_2 的數量是由第一筆的C的數量4與D的數量1的加總為5）來建立一個含有祖先項目的交易表4-2

表4-2：含有祖先項目的交易表

交易ID	項 目
1	(A,3)(C,4)(D,1)(T ₁ ,3)(T ₂ ,5)
2	(B,3)(C,7)(D,7)(T ₁ ,3)(T ₂ ,14)
3	(A,2)(B,10)(D,5)(T ₁ ,12)(T ₂ ,5)
4	(A,7)(B,2)(C,8)(D,8)(T ₁ ,9)(T ₂ ,16)
5	(B,2)(C,8)(D,10)(T ₁ ,2)(T ₂ ,18)

步驟2-4 透過模糊歸屬函數(圖4-3)來計算每筆項目A,B,C,D,T₁,T₂的歸屬度值，結果如下表4-3。然後再計算每一個模糊集合的歸屬度總值(如表4-3中的FS)，其為每一個項目屬於某一語意值的支持度。並儲存支持度於在候選項目為一的資料方塊中。在表4-3灰底為大項目組：

表4-3：長度為1之候選項目

項目組	1	2	3	4	5	FS
A.L	0.4	0.0	0.6	0.0	0.0	1.0
A.M	0.6	0.0	0.4	0.6	0.0	1.6
A.H	0.0	0.0	0.0	0.4	0.0	0.4
B.L	0.0	0.4	0.0	0.6	0.6	1.6
B.M	0.0	0.6	0.0	0.4	0.4	1.4
B.H	0.0	0.0	1.0	0.0	0.0	1.0
C.L	0.2	0.0	0.0	0.0	0.0	0.2
C.M	0.8	0.6	0.0	0.4	0.4	2.2
C.H	0.0	0.4	0.0	0.6	0.6	1.6
D.L	0.8	0.0	0.0	0.0	0.0	0.8
D.M	0.2	0.6	1.0	0.4	0.0	2.2
D.H	0.0	0.4	0.0	0.6	1.0	2.0
T ₁ .L	0.4	0.4	0.0	0.0	0.6	1.4
T ₁ .M	0.6	0.6	0.0	0.2	0.4	1.8
T ₁ .H	0.0	0.0	1.0	0.8	0.0	1.8
T ₂ .L	0.0	0.0	0.0	0.0	0.0	0.0
T ₂ .M	1.0	0.0	1.0	0.0	0.0	2.0
T ₂ .H	0.0	1.0	0.0	1.0	1.0	3.0

步驟5 產生項目長度為2之候選項目組並計算長度為2之候選項目組在每一筆之歸屬值並加總得其長度為2之候選項目的支持度。如下表4-5最後把支持度儲存在候選項目為二的資料方塊中。例如A.L*B.L的第一筆的歸屬值計算為 $A.L(1) = 0.4$ 去乘上 $B.L(1) = 0.0$ ，而得到 $A.L(1)*B.L(1)$ 歸屬值為0.0。依此類推算完A.L*B.L五筆的歸屬值並加總為0.0，如下表4-4。其他項目也依此類推計算之。在表4-5灰底為大項目組：

表4-4：A.L*B.L範例表

交易ID	A.L	B.L	A.L*B.L
1	0.4	0.0	0.0
2	0.0	0.4	0.0
3	0.6	0.0	0.0
4	0.0	0.6	0.0
5	0.0	0.6	0.0
Fuzzy count			0.0

表4-5：長度為2之候選項目

項目組	FS	項目組	FS	項目組	FS	項目組	FS
A.L*B.L	0.00	A.H*C.M	0.16	B.M* T ₂ .H	1.40	C.H* T ₁ .L	0.40
A.L*B.M	0.32	A.H*C.H	0.24	B.H*C.L	0.00	C.H* T ₁ .M	0.72
A.L*B.H	0.60	A.H*D.L	0.00	B.H*C.M	0.00	C.H* T ₁ .H	0.48
A.L*C.L	0.08	A.H*D.M	0.16	B.H*C.H	0.00	C.H* T ₂ .L	0.00
A.L*C.M	0.32	A.H*D.H	0.24	B.H*D.L	0.00	C.H* T ₂ .M	0.00
A.L*C.H	0.00	A.H* T ₁ .L	0.00	B.H*D.M	1.00	C.H* T ₂ .H	1.60
A.L*D.L	0.32	A.H* T ₁ .M	0.08	B.H*D.H	0.00	D.L* T ₁ .L	0.32
A.L*D.M	0.68	A.H* T ₁ .H	0.32	B.H* T ₁ .L	0.00	D.L* T ₁ .M	0.48
A.L*D.H	0.00	A.H* T ₂ .L	0.00	B.H* T ₁ .M	0.00	D.L* T ₁ .H	0.00
A.L* T ₁ .L	0.16	A.H* T ₂ .M	0.00	B.H* T ₁ .H	1.00	D.L* T ₂ .L	0.00
A.L* T ₁ .M	0.24	A.H* T ₂ .H	0.40	B.H* T ₂ .L	0.00	D.L* T ₂ .M	0.80
A.L* T ₁ .H	0.60	B.L*C.L	0.00	B.H* T ₂ .M	1.00	D.L* T ₂ .H	0.00
A.L* T ₂ .L	0.00	B.L*C.M	0.72	B.H* T ₂ .H	0.00	D.M* T ₁ .L	0.32
A.L* T ₂ .M	1.00	B.L*C.H	0.88	C.L*D.L	0.16	D.M* T ₁ .M	0.56
A.L* T ₂ .H	0.00	B.L*D.L	0.00	C.L*D.M	0.04	D.M* T ₁ .H	1.32
A.M*B.L	0.36	B.L*D.M	0.48	C.L*D.H	0.00	D.M* T ₂ .L	0.00
A.M*B.M	0.24	B.L*D.H	1.12	C.L* T ₁ .L	0.08	D.M* T ₂ .M	1.20
A.M*B.H	0.40	B.L* T ₁ .L	0.40	C.L* T ₁ .M	0.12	D.M* T ₂ .H	1.00
A.M*C.L	0.12	B.L* T ₁ .M	0.72	C.L* T ₁ .H	0.00	D.H* T ₁ .L	0.56
A.M*C.M	0.72	B.L* T ₁ .H	0.48	C.L* T ₂ .L	0.00	D.H* T ₁ .M	0.84
A.M*C.H	0.36	B.L* T ₂ .L	0.00	C.L* T ₂ .M	0.20	D.H* T ₁ .H	0.48
A.M*D.L	0.48	B.L* T ₂ .M	0.00	C.L* T ₂ .H	0.00	D.H* T ₂ .L	0.00
A.M*D.M	0.76	B.L* T ₂ .H	1.60	C.M*D.L	0.64	D.H* T ₂ .M	0.00
A.M*D.H	0.36	B.M*C.L	0.00	C.M*D.M	0.68	D.H* T ₂ .H	2.00
A.M* T ₁ .L	0.24	B.M*C.M	0.52	C.M*D.H	0.88	T ₁ .L* T ₂ .L	0.00
A.M* T ₁ .M	0.48	B.M*C.H	0.72	C.M* T ₁ .L	0.72	T ₁ .L* T ₂ .M	0.40
A.M* T ₁ .H	0.88	B.M*D.L	0.00	C.M* T ₁ .M	1.16	T ₁ .L* T ₂ .H	0.80
A.M* T ₂ .L	0.00	B.M*D.M	0.52	C.M* T ₁ .H	0.32	T ₁ .M* T ₂ .L	0.00
A.M* T ₂ .M	1.00	B.M*D.H	0.88	C.M* T ₂ .L	0.00	T ₁ .M* T ₂ .M	0.40
A.M* T ₂ .H	0.60	B.M* T ₁ .L	0.40	C.M* T ₂ .M	0.80	T ₁ .M* T ₂ .H	0.40
A.H*B.L	0.24	B.M* T ₁ .M	0.68	C.M* T ₂ .H	1.40	T ₁ .H* T ₂ .L	0.00
A.H*B.M	0.16	B.M* T ₁ .H	0.32	C.H*D.L	0.00	T ₁ .H* T ₂ .M	1.00
A.H*B.H	0.00	B.M* T ₂ .L	0.00	C.H*D.M	0.48	T ₁ .H* T ₂ .H	0.80
A.H*C.L	0.00	B.M* T ₂ .M	0.00	C.H*D.H	1.12		

步驟6 重複步驟5之方法直到產生項目長度為6之候選項目組為止

步驟7 從每一個項目長度為1之候選項目的歸屬總值（模糊支持度）到每一個項目長度為6之候選項目的模糊支持度與最小模糊支持度進行比對，大於等於1.6的項目組為大項目組。以長度為1的大項目組為例，結果如下表4-6：

表4-6： Large 1-itemsets

項目組	FS
A.M	1.6
B.L	1.6
C.M	2.2
C.H	1.6
D.M	2.2
D.H	2.0
T ₁ .M	1.8
T ₁ .H	1.8
T ₂ .M	2.0
T ₂ .H	3.0

長度為2的大項目組，如下表4-7：

表4-7： Large 2-itemsets

項目組	FS
B.L* T ₂ .H	1.60
C.H* T ₂ .H	1.60
D.H* T ₂ .H	2.00

步驟8-9 由大項目組來產生候選法則，如下表4-8。並與最小模糊信賴度0.7比較得到符合的法則如下表4-9：

表4-8： 候選法則

法則	信賴度
$B.L \rightarrow T_2.H$	$(B.L * T_2.H) / B.L = 1.6 / 1.6 = 1$
$C.H \rightarrow T_2.H$	1
$D.H \rightarrow T_2.H$	1
$T_2.H \rightarrow B.L$	0.53
$T_2.H \rightarrow C.H$	0.53
$T_2.H \rightarrow D.H$	1

表4-9： 符合最小模糊信賴度之法則

符合法則	
$B.L \rightarrow T_2.H$	$D.H \rightarrow T_2.H$
$C.H \rightarrow T_2.H$	$T_2.H \rightarrow D.H$

□

第二節 模糊切割之漸進式與線上多階層關聯法則

在此節我們將探討模糊切割之多階層關聯法則在漸進式探勘以及線上探勘等特殊問題上的應用。在文獻探討中，我們可以發現過去傳統的多階層關聯法則很難進行漸進式探勘以及線上探勘等問題。但透過我們的模糊切割之多階層關聯法則便能夠有效地進行解決。接下來，我們將分別介紹模糊切割之漸進式多階層關聯法則以及模糊切割之線上探勘多階層關聯法則。

一、模糊切割之漸進式多階層關聯法則

在第一節介紹模糊切割之多階層關聯法則時，我們是先計算出長度為 1 的候選項目組的支持度到長度為 k 的候選項目組的支持度，並將這些支持度儲存於資料方塊中。當資料庫有所異動時，便能很快速地找到其異動所影響到的項目組來進行此項目組的支持度

變更。接下來，我們將以資料庫異動的不同狀況來說明我們所因應之策略，並用範例說明之，如下：

1. 當新增幾筆資料時：

先將新增的資料先進行第一節中的步驟 1 至步驟 6 的程序，而會得到新增部分 $FS(C_t)_{new} | t=1 \text{ to } k$ 。然後將原有資料之長度為 1 的候選項目組支持度加上新增部分之長度為 1 的候選項目組支持度，就是新增後之長度為 1 的候選項目組支持度。依此類推計算出其他長度的候選項目組支持度

$$\left(FS(C_t)_{new} + FS(C_t)_{old} | t=1 \text{ to } k \right)。$$

範例 1:

假設新增的資料如下表 4-10：

表4-10：新增資料表

交易ID	項 目
6	(A,3)(C,4)(D,1)(T ₁ ,3)(T ₂ ,5)
7	(B,3)(C,7)(D,7)(T ₁ ,3)(T ₂ ,14)
8	(A,2)(B,10)(D,5)(T ₁ ,12)(T ₂ ,5)

就以長度為1的項目組為例：

我們首先先計算這三筆每個項目組的歸屬度值並加總得到每一個項目組的支持度，如下表4-11。接下來，將新增之部分的支持度加到資料方塊中，新增後的長度為1的模糊支持度如下表4-11，灰底為大項目組。

表4-11：新增部分的支持度

項目組	新增部分之FS	新增前之FS	新增後之FS
A.L	1.0	1.0	2.0
A.M	1.0	1.6	2.6
A.H	0.0	0.4	0.4
B.L	0.4	1.6	2.0
B.M	0.6	1.4	2.0
B.H	1.0	1.0	2.0

C.L	0.2	0.2	0.4
C.M	1.4	2.2	3.6
C.H	0.4	1.6	2.0
D.L	0.8	0.8	1.6
D.M	0.8	2.2	3.0
D.H	0.4	2.0	2.4
T ₁ .L	0.8	1.4	2.2
T ₁ .M	1.2	1.8	2.0
T ₁ .H	1.0	1.8	2.8
T ₂ .L	0.0	0.0	0.0
T ₂ .M	2.0	2.0	4.0
T ₂ .H	1.0	3.0	3.0

其他維度之項目也依此類推來進行更新，找出各維度的大項目組後，並產生法則。

□

2. 當刪除原有資料幾筆時：

直接找出與要刪除那筆資料相關的項目組歸屬值，並將原有的模糊支持度扣除之。

範例 2: 假設要刪除第 5 筆資料

直接找出與第 5 筆資料相關的項目之歸屬值，如下表 4-12：

表4-12：與第5筆資料相關之項目的歸屬值

項目組	5
A.L	0.0
A.M	0.0
A.H	0.0
B.L	0.6
B.M	0.4
B.H	0.0
C.L	0.0
C.M	0.4
C.H	0.6
D.L	0.0
D.M	0.0
D.H	1.0
T ₁ .L	0.6
T ₁ .M	0.4

T ₁ .H	0.0
T ₂ .L	0.0
T ₂ .M	0.0
T ₂ .H	1.0

以長度為 2 的 B.M* D.H 為例：

首先計算出 B.M* D.H 之歸屬值 $0.4 \times 1.0 = 0.4$ 並將原有的模糊支持度 0.88 扣除之得結果為 0.4。

□

二、模糊切割之線上挖掘多階層關聯法則

當最小模糊支持度變動時，有兩種狀況，一種是新的最小模糊支持度比舊的最小模糊支持度還小時以及新的最小模糊支持度比舊的最小模糊支持度還大時。接下來將分別介紹這兩種狀況時，我們的因應策略。

1. 新的最小模糊支持度比舊的最小模糊支持度還小時：

從長度為 1 至 k 的候選項目組中找出其支持度介於舊的最小模糊支持度與新的最小模糊支持度之間的項目組，這些項目組原本是非大項目組改成為大項目組。

2. 新的最小模糊支持度比舊的最小模糊支持度還大時：

從長度為 1 至 k 的候選項目組中找出其支持度介於舊的最小模糊支持度與新的最小模糊支持度之間的項目組，這些項目組原本是大項目組改成為非大項目組。

第三節 實驗與探討

本節中主要對於在漸進式挖掘以及線上挖掘的環境下，沒有使用資料方塊預存項目組的資訊以及有利用資料方塊預存項目組的資訊在效率上的比較。

一、實驗測試環境：

為了實作本實驗，我們利用以下的環境來進行實作與測試：

(一) 實驗平台：

1. CPU：Intel Pentium 4 2.4GHz
2. RAM：DDR 256Mb，HD：Maxtro 80Gb Ultra DMA133
3. OS：Windows 2000 Professional
4. DB：Access 2002

(二) 實作之程式語言：

為了比較能較客觀、公平，我們均利用 Visual Basic 6.0 來實作一模糊多階層關聯法則在漸進式挖掘與線上挖掘環境下，無資料方塊與有資料方塊的兩種情況之比較與分析。

(三) 實驗資料庫：

本實驗的資料來源為亂數產生 10000 筆資料，共有四種商品項目(A, B, C, D)與此商品項目之祖先項目(T1, T2)，如下圖 4-4，並且模糊切割 K 設為 3。而在資料方塊的部分，本研究是利用主記憶體模擬資料方塊，而並非儲存至硬碟的資料方塊。

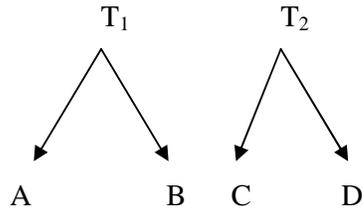


圖 4-4：商品項目與祖先項目關係圖

二、實驗設計、結果與分析：

(一) 漸進式挖掘之實驗

在動態資料庫的環境下，資料庫中的資料是不斷新增。若無資料方塊預先儲存每一個項目組的支持度時（與傳統關聯法則同），會導致部份項目組的支持度之資訊遺失而必須重新執行關聯法則程序。但若有利用資料方塊預先儲存項目組的支持度時，只要計算新增部分的每個項目組之支持度並累加至所對應的資料方塊位置。最後在由門檻值—最小支持度進行比對，就能找到大項目組。因此，在本實驗將驗證有利用資料方塊與沒有利用資料方塊的差異。

1. 實驗目的：

探討在漸進式挖掘的環境下，模糊多階層關聯法則在有利利用資料方塊來預先儲存項目組支持度以及無資料方塊的績效比較，以驗證有利用資料方塊來預先儲存項目組支持度在漸進式挖掘的優勢。

2. 實驗方法：

此實驗的門檻值分別設為 0.05、0.1、0.2、0.3。首先，先執行 10000 筆資料的關聯法則程序，然後再計算新增 4000 筆、5000 筆、7000 筆時，無資料方塊與有資料方塊在各個支持度與新增筆數所花費的時間。

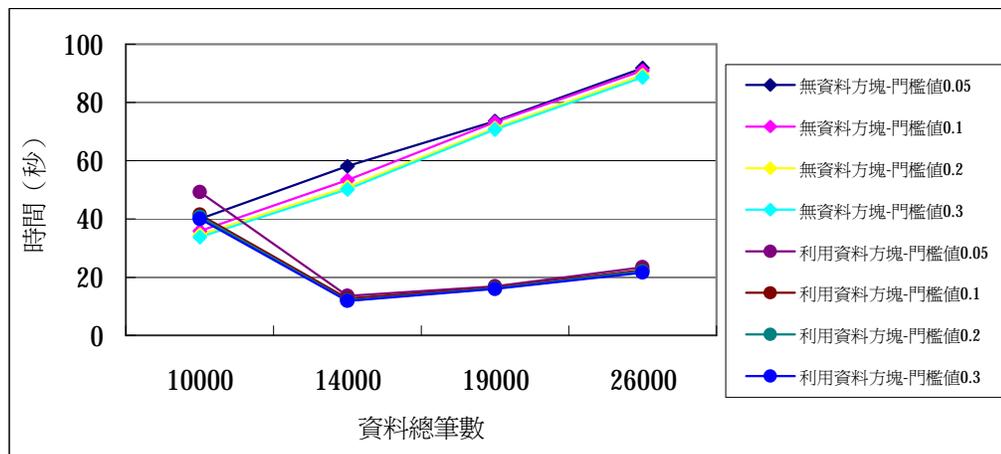
3.實驗結果：

如下表 4-13 與圖 4-5

表 4-13：資料筆數漸增示意表

門檻值	筆數	10000	14000	19000	26000
無資料方塊	0.05	40.01	58.18	73.64	91.86
	0.1	35.93	53.29	73.16	90.93
	0.2	34.64	50.93	71.27	89.34
	0.3	33.75	50.16	70.77	88.59
有資料方塊	0.05	49.21	13.59	16.89	23.43
	0.1	41.47	12.67	16.45	22.46
	0.2	40.73	12.11	16.17	21.99
	0.3	39.97	11.88	15.94	21.56

圖 4-5：資料筆數漸增示意圖



4.結果探討與分析：

由上圖表，我們可以發現有利用資料方塊預先儲存項目組的支持度之資訊，在 10000 筆時比沒利用資料方塊所花費

的時間多，主要是因為是沒利用資料方塊所產生的項目組較有利用資料方塊來的少。可是當資料由 10000 筆新增為 14000 筆、19000 筆、26000 筆時有利用資料方塊所花費的時間比沒利用資料方塊來得少，而且時間差距有拉大的趨勢。因此，在漸進式挖掘的環境下，利用資料方塊來預先儲存所有項目組的支持度是績效最佳的方式。

(二) 線上挖掘之實驗

在線上挖掘的環境下，可以讓使用者任意調整門檻值。若無資料方塊預先儲存每一個項目組的支持度時（與傳統關聯法則同），會導致部分項目組的支持度之資訊遺失而必須重新執行關聯法則程序。但若有利用資料方塊預先儲存項目組的支持度時，只要比對儲存在資料方塊中的每一個項目組之支持度與新的門檻值—最小支持度，就能找到大項目組。因此，在本實驗將驗證有利用資料方塊與沒有利用資料方塊的差異。

1. 實驗目的：

探討在線上挖掘的環境下，模糊多階層關聯法則在有利利用資料方塊來預先儲存項目組支持度以及無資料方塊的績效比較，以驗證有利利用資料方塊來預先儲存項目組支持度在線上挖掘的優勢。

2. 實驗方法：

先執行門檻值為 0.5 的關聯法則程序，然後在計算動態調整門檻值為 0.05、0.1、0.2、0.3、0.4、0.6、0.7 時，無資料方塊與有資料方塊在各個支持度所花費的時間。

3.實驗結果：

如下表 4-14 與圖 4-6

表 4-14：動態調整門檻值示意表

調整門檻值	0.05	0.1	0.2	0.3	0.4	0.6	0.7
無資料方塊	35.07	34.94	34.23	33.75	33.42	32.97	31.68
有資料方塊	0.00197	0.00197	0.00169	0.00145	0.00145	0.00145	0.00043

時間單位：秒

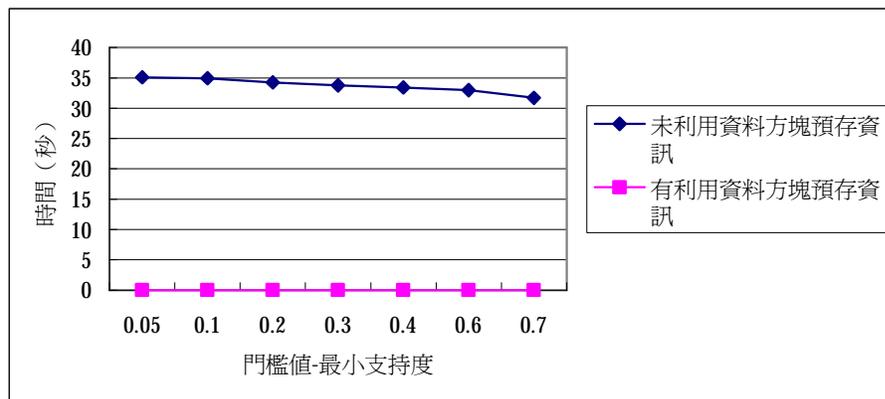


圖 4-6：動態調整門檻值示意圖

4.結果探討與分析：

由上圖表，我們可以發現有利用資料方塊預先儲存項目組的支持度之資訊，在調整門檻值時所花費的時間趨近於 0 秒。而沒有利用資料方塊所花費的時間超過 30 秒以上。因此，在線上挖掘的環境下，利用資料方塊預先儲存項目組之支持度是績效最佳的方式。

三、實驗結論：

透過上述兩個實驗，我們可以發現本研究的模糊切割為基礎關聯法則演算法有利用資料方塊預先儲存項目組的資訊，不論是在動態資料庫新增的漸進式挖掘或是動態調整門檻值的線上挖掘都有不錯的績效表現。主要理由是，沒有利用資料方塊時（與傳統關聯法則同），非大項目組的資訊將會遺失掉，而必須重新執行費時且費力的關聯法則程序。但若能預先儲存這些資訊，我們就只要考量新增的項目組之部份或是掃描資料方塊的資訊而不必重新執行費時且費力的關聯法則程序。因此，本研究的方法在漸進式挖掘和線上挖掘較傳統關聯法則演算法有不錯的績效表現。

第四節 討論與結論

在現今產品項目的多樣化，往往很難求出能夠滿足最小支持度的關聯法則，有時候會將一些具有高價值性的資訊遺漏而不自知。為了避免這種遺憾發生，便有研究利用歸納的方法，將商品項目進行階層化分類動作。透過不同層次的分類，我們便可以對動同層次作關聯法則的求取。但過去傳統的多階層的關聯法則大部分適用在二值屬性的資料，對於其他類型的資料就很難去進行處理。加上，現今的資料庫的資料隨時在變動，導致過去的演算法無法有效地運用，使得當資料變更時必須重新做一次關聯法則挖掘。

因此，我們利用模糊理論的概念來解決傳統多階層關聯法則演算法所無法處理的數值型資料之外，我們並結合模糊切割與資料方塊技術，來建構一個功能完整的多階層關聯法則演算法。此演算法能夠解決現實資料庫經常性變動以及使用者可以隨時調整門檻值所造成的問題。以便讓企業界的決策者在面對如此快速變化且競爭的

環境下，能夠有效地掌握住產品供需的大方向，以確認發展的目標並下達正確的決策。

第五章 結論與未來展望

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。由於在數值型資料、資料庫動態更新以及使用者能夠線上來進行調整門檻值來進行探勘等問題上，傳統過去關聯法則演算法無法有效進行處理。因此，本研究提出了以模糊切割技術為基礎來建構關聯法則演算法－FPBARA，來彌補過去傳統關聯法則演算法之缺失。

我們總共提出了兩種以模糊切割技術為基礎之關聯法則演算法，分別為模糊切割之單層次關聯法則演算法以及模糊切割之多階層關聯法則演算法。除了這兩種新方法能夠有效處理數值型資料、資料庫動態更新以及使用者能夠線上來進行調整門檻值的問題之外，我們還再針對這兩個方法，提出在顧客關係管理上的應用。其分別為，(一)利用 FPBARA 之單層次關聯法則演算法於 Fuzzy RFM Model 的建構，來改善過去 RFM Model 的缺失，並讓企業經營者能夠以全面的角度來分析顧客行為以及制訂更完整的行銷決策；(二)利用 FPBARA 之多階層關聯法則演算法於一般商品上的關聯探討，

並提供不同階層的管理者其所需的資訊，來幫助他（她）們來進行決策制訂等。

最後，本研究未來可繼續探究的方向有以下三點，來提供未來研究者研究的方向：

一、改善儲存空間上的限制：

為了能夠處理數值型資料以及資料庫變動和使用者能夠線上調整門檻值，我們的方法就會產生相當大量的項目組，而且所有項目組支持度必須預存至儲存空間，這對於記憶體的儲存與管理上將是一個極大的考驗。因此，我們將提出以下三點可行解決方案來作為未來研究者研究方向。

（一）在高頻項目組與非高頻項目組間設定一個準高頻項目組 (pre-large itemsets) [2]，作為一個有效的緩衝

（二）利用分散式處理的技術，來以減少資料處理的負荷量

（三）利用快取的概念來找出比較經常性變動的項目來作為儲存的對象，而不必儲存過多的而較為無用的資料

二、模糊歸屬函數以及其他參數之設定：

在本研究所運用的模糊歸屬函數是採用一般具有不錯效果的三角形模糊歸屬函數以及模糊切割之數目等。未來，可以利用基因演算法或是類神經網路等學習演算法找出最佳的模糊歸屬函數以及最佳的參數值，以能適切不同的應用中。

三、在序列型樣之關聯法則的發展：

本研究的 FPBARA 主要是用於單層次關聯法則演算法與多階層關聯法則演算法，未來，還可利用 FPBARA 於序列型樣的關聯法則演算法之研究中與其應用。

四、在應用上的展望：

本研究利用模糊切割為基礎之關聯法則應用於RFM Model的改良、一般商品之跨階層關聯探討以及網路探勘等電子商務系統的經營與行銷，以作為制訂行銷策略的參考。未來，可以利用這三個方法（單階層、多階層、序列型樣）的混合或是結合其他技術來應用於其他領域。例如：

- （一）我們可以結合模糊切割之多階層關聯法則與模糊切割之序列型樣所建構之模糊切割之多階層序列型樣，來進行網路探勘中的Web Structure Mining與Web Usage Mining問題上[6,38]。
- （二）藉由所建立之專屬會員制度，透過本研究的FPBARA，從龐大的會員交易資料中挖掘消費特性，以實現個人化服務，有效區隔市場與訂定行銷決策[13]。

參考文獻

一、中文部分：

- [1] 毛立仁，「動態新增資料集中敏感資料挖掘技術之研究」，南華大學資訊管理研究所碩士論文，民國 91 年。
- [2] 王慶堯，「利用準大項目集之漸進式挖掘」，義守大學資訊工程研究所碩士論文，民國 89 年。
- [3] 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。
- [4] 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。
- [5] 林高弘，「有趣性關聯法則之線上調適性挖掘法的研究」，南華大學資訊管理研究所碩士論文，民國 92 年。
- [6] 林傑彬、劉明德、陳湘，資料採掘與 OLAP 理論與實務，台北：文魁資訊股份有限公司，民國 91 年。
- [7] 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。
- [8] 許中川、林勇助、林旻宏，「探勘持續性購買行為—以銷售資料為例」，第三屆產業資訊管理學術暨新興科技實務研討會，pp.558~569，台北，民國 91 年。
- [9] 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。
- [10] 楊昇樺，「可互動調適性挖掘關聯法則之整合式架構的研究」，南華大學資訊管理研究所碩士論文，民國 92 年。
- [11] 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。
- [12] 張晉赫、謝文天、蔡佳偉、邱宏彬，「運用資料探勘技術進行一對一行銷之電子商務推薦系統」，2003 企業管理學術研討會暨 2003 電子商務經營管理研討會，pp.387~392，台中，民國 92 年。

- [13] 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，第十卷第一期，pp.181-204，民國 92 年。
- [14] 蘇建源、邱美倫、邱宏彬、吳光閔，「應用資料探勘技術支援顧客導向影片檢索及推薦之智慧型人機介面」，2003 企業管理學術研討會暨 2003 電子商務經營管理研討會，pp.427~432，台中，民國 92 年。

二、英文部分：

- [15] A. Savasere, E. Omiecinski, S. Navathe, “An efficient Algorithm for Mining Association Rules in Large Databases,” *Proc. of the 21st Conf. on Very Large Data Bases*, pp.432-444, September, 1995.
- [16] C. C. Aggarwal and P. S. Yu, “Online generation of association rules,” *Proc. of the IEEE ICDE '98*, pp.402-411, 1998.
- [17] C. Kleissner, “Data mining for the enterprise,” *Proc. of the Thirty-First Hawaii International Conference*, Vol. 7, pp. 295-304, 1998.
- [18] C.M. Kuok, A. Fu, and M.H. Wong, ”Mining fuzzy association rules in databases,” *ACM SIGMOD record* 27, pp.41-46, 1998.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] H. Toivonen, “Sampling large databases for association rules,” *Proc. of the 22nd Conf. on VLDB*, pp.134-145, 1996.
- [23] J. Hipp, U. Guntzer and Gholamreza, “Algorithms for association rule mining – A general survey and comparison,” *Proc. of ACM SIGKDD*, pp.58-64, 2000.
- [24] Jiawei Han and Youngjian Fu, “Discovery of multiple-level association rules from large database,” *The Internat. Conf. on Very Large Databases*, 1995.
- [25] J. S. Park, M. S. Chen, and P. S. Yu, “An effective hash based algorithm for mining association rules,” *Proc. of ACM SIGMOD*, pp.175-186, 1995.

- [26] Kahan Ron, "Using Database Marketing Techniques to Enhance Your One-to-One Marketing Initiatives", *Journal of Consumer Marketing*, Vol.15(5), pp.491-493,1998.
- [27] L. A. Zadeh, "Fuzzy sets" ,*Information Control*,Vol.8(3), pp.338-353,1965.
- [28] M. Ait Kbir, H.Benkirane, K. Maalmi, and R. Benslimane, "Hierarchical fuzzy partition for pattern classification with fuzzy if-then rules" *Pattern Recognition Letters*,21, pp.503-509,2000.
- [29] M. Berry and G. Linoff, "Data Mining Techniques for marketing, sales, and Customer Support," *New York. Wiley Computer Publishing*,1997.
- [30] M. Goebel and L. Gruenwald, "A survey of data mining and knowledge discovery software tools," *ACM SIGKDD*, Val. 1, pp. 20- 33,1999.
- [31] Miglautsch John, "Thoughts on RFM Scoring",*Journal of Database Marketing*,Vol.8(1),2000.
- [32] 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.
- [33] 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.
- [34] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," *Proc. of the 20th Conf. on Very Large Databases*, pp.487-499,1994.
- [35] R. Agrawal and R. Srikant, "Mining Sequential Patterns," *Proc. of the 11th Conf. on Data Engineering*, pp.3-14,1995.
- [36] 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.
- [37] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," *Proc. of the ACM SIGMOD Conf. on Management of Data*, pp.207-216,1993.
- [38] R. Kosala and H. Blockeel," Web Mining Research : A Survey," In *ACM SIGMOD*,Vol.2-1,pp.1-15,2000.

- [39] R. Ng, L. V. S. Lakshmanan, J. Han, and A. Pang, "Exploratory mining and pruning optimizations of constrained association rules," *Proc. of ACM SIGMOD Conf. on Management of Data*, pp.13-24,1998.
- [40] R. Ng, L. V. S. Lakshmanan, J. Han, and T. Mah, "Exploratory mining via constrained frequent set queries," *Proc. of ACM SIGMOD Conf. on Management of Data*, pp.556-558,1999.
- [41] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," *Proc. of ACM SIGMOD Conf. on Management of Data*, pp. 1-12,1996.
- [42] R. Srikant and R. Agrawal, "Mining generalized association rules,"*Proc. of the 21st VLDB*, pp. 407-419,1997.
- [43] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic itemset counting and implication rules for marketing basket data," *ACM SIGMOD Conf. on Management of Data*, pp.255-264,1997.
- [44] Shaw Michael J , Chandrasekar Subramaniam ,Gek Woo Tan, and Michael E. Welge, "Knowledge Management and Data Mining for Marketing ,"*Journal of Decision Support Systems*,Vol.31,pp.127-137,2001.
- [45] Stone Bob, Successful Direct Marketing Methods, 4th ed.,NTC Business Books,1989.
- [46] Sung Ho Ha and Sang Chan Park, "Application of data mining tools to hotel data mart on the Intranet for database marketing,"*Journal of Expert Systems with Applications*,Vol.15,pp.1-31,1998.
- [47] U. M. Fayyad, "Data mining and knowledge discovery: making sense out of data,"*IEEE Expert*, Vol. 11, No. 5, pp.20-25,1996.
- [48] T. Fukuda, Y. Morimoto, S. Morishita and T. Tokuyama, "Mining optimized association rules for numeric attributes," *The ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pp. 182-191,1996.
- [49] Yen John and Reza Langari, Fuzzy Logic Intelligence, Control, and Information , Prentice-Hall, Inc.,,1999.
- [50] Yi-ChungHu , Ruey-ShunChen, and Gwo-HshiungTzeng, "Discovering fuzzy association rules using fuzzy partition methods", *Journal of Knowledge-Based Systems*, Vol.16,pp.137-147,2003.