

國立交通大學

管理學院(資訊管理學程)碩士班

碩士論文

以改良式 RFM 模式結合資料探勘建立客戶分群並提升行銷效益

Using Refined RFM Model and Data Mining to
Establish Customer Segmentations and Improve
Marketing Efficiency

研究生：徐火志

指導教授：劉敦仁 教授

中華民國九十五年六月

以改良式 RFM 模式結合資料探勘建立客戶分群並提升行銷效益

Using Refined RFM Model and Data Mining to Establish
Customer Segmentations and Improve Marketing Efficiency

研究生：徐火志

Student : Hou Chih Hsu

指導教授：劉敦仁 博士

Advisor : Dr. Duen Ren Liu



Submitted to Institute of Information Management
College of Management

National Chiao Tung University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

in

Information Management

June 2006

Hsinchu, Taiwan, the Republic of China

中華民國九十五年六月

以改良式 RFM 模式結合資料探勘建立客戶分群並提升行銷效益

研究生：徐火志

指導教授：劉敦仁 博士

國立交通大學資訊管理研究所

摘要

依據義大利經濟學家帕列托所提出所謂 80/20 法則(1897)，意指在原因和結果、努力和收穫之間，存在著不平衡的關係，而典型的情況是：80%的收穫，來自 20%的付出，也就是 80%的結果，歸結於 20%的原因。而在實際的情形下也是如此，20%的重要客戶貢獻了利潤的 150%；而最差的 40%客戶，使利潤縮減 50%，可見客戶區隔的重要性。

所謂「客戶分群」，係指利用完整消費者的客戶基本資料、交易分析資料、客戶的互動資料，將客戶劃分為數個不同消費行為模式的客戶群，並建立客戶的分群，以期能讓企業分離中出高價值客戶、主力型客戶、成長型客戶、無價值客戶等，有助於讓企業針對不同的使用群制訂適當的行銷策略。

本研究利用改良型 RFM 模型建立客戶分群的系統雛形，並利用資料探勘之資料分群之技術-SOM 的運算法則將不同的各客戶資料主動依客戶的交易資料歸類四群，並將此四群映射於高價值客戶、主力型客戶、成長型客戶、無價值客戶等四個客戶群組中，同時針對映射的客戶群組資料作一分析與探討，並與市場及業務部門的功能結合，期望利用此一決策系統幫助市場及業務相關人員規劃並審視現行市場及業務策略。

Using Refined RFM Model and Data Mining to Establish Customer Segmentations and Improve Marketing Efficiency

Student : Hou Chih Hsu

Advisor : Dr. Duen Ren Liu

Institute of Information Management

National Chiao Tung University

Hsinchu, Taiwan, Republic of China

Abstract

According to the Pareto's 80/20 rule, there were unbalance relationship between cause and effect. It means that 80% gain may come from 20% contribution in customers. Thus, how to obtain the maximum benefit between customers and enterprise has become a critical issue in market analysis.

Customer segmentation is one of approach which is using marketing database, including transaction data, customer data, to cluster the customers into 4 groups, : High value customers, Major Customers, Valueless Customers and Growing Customers.

In this paper, we proposed to use refined RFM model and Self-Organization Map(SOM) to build customer segmentation prototype, then to verify this prototype using transaction data of customers in semiconductor manufacturing. Further, we propose the respective marketing policy in the 4 groups in order to make great contribution to markets and sales.

誌謝

感謝神，讓我有機會繼續進修與完成一個工作、家庭與學業三方的艱難拔河競賽。

感謝我的指導教授劉敦仁教授，在劉老師的悉心指導下，讓我能對各個相關領域有了更深的認知。

最後要感謝我的太太，因為她的全力支持，我才能無後顧之憂，得以安心的完成學業。也要感謝我的女兒，雖然不能給我什麼課業上的指導，但對於我獲得知識的喜悅希望能讓他們知道。

一切喜悅歸與關心我的人，一切榮耀歸與神。



目錄

1.	緒論.....	9
1.1.	研究背景與動機.....	9
1.2.	研究目的.....	9
1.3.	論文架構.....	10
2.	文獻探討.....	11
2.1.	資料庫行銷.....	11
2.2.	RFM 分析模型(RFM ANALYSIS MODEL)	13
2.2.1.	<i>RFM 分析模型限制</i>	18
2.3.	自組織神經網路(SELF-ORGANIZING MAP, SOM)	19
2.4.	分群方法回顧.....	30
3.	改良式客戶分群系統架構概論	35
3.1.	問題描述與規劃.....	35
3.2.	研究架構.....	36
4.	改良式客戶分群系統雛形建置	37
4.1.	資料收集與描述.....	37
4.2.	資料前處理過程.....	42
4.2.1.	<i>建構RFM指標</i>	42
4.2.2.	<i>RFM指標之正規化</i>	44
4.3.	資料分群模型.....	47
5.	結果分析.....	53
5.1.	結果分析.....	53
5.2.	行銷決策支援分析.....	63
6.	結論與建議.....	65
6.1.	研究結論.....	65
6.2.	研究限制與建議.....	65
7.	參考文獻.....	67
8.	附錄.....	70
	附錄一 GNU GENERAL PUBLIC LICENSE通用公共許可證	70

附圖目錄

附圖 2.2-1 RFM 客戶分析模型.....	14
附圖 2.2-2 RFM模型最近購買日期.....	15
附圖 2.2-3 RFM模型購買頻率.....	16
附圖 2.2-4 RFM模型購買金額.....	17
附圖 2.2-5 RFM組合轉換.....	17
附圖 2.2-6 RFM指標排序結果.....	18
附圖 2.3-1 自組織特徵映射網路架構.....	20
附圖 2.3-2 優勝單元與鄰近區域內權重修正示意圖.....	22
附圖 2.3-3 由SOM輸入空間與輸出空間權重修正情形[Dittenbach, Rauber, &Merk1, 2002].....	24
附圖 2.3-4 SOM演算法學習步驟流程圖.....	25
附圖 2.3-5 常見的鄰近區域型態.....	26
附圖 2.3-6 兩種型態的鄰近函數.....	27
附圖 2.4-1 分群法分類[Jain & Dubes, 1988].....	31
附圖 3.2-1 分析架構.....	36
附圖 4.1-1 各月中的交易數量的Histogram 圖.....	38
附圖 4.1-2 依月份的營收Box Plot.....	39
附圖 4.1-3 原始交易資料的Revenue之Histogram Plot.....	40
附圖 4.1-4 依各月份的交易數量Box Plot.....	41
附圖 4.1-5 原始資料之ASP Box Plot.....	42
附圖 4.2-1 正規化後的RFM Histogram.....	45
附圖 4.2-2 正規化後的Normal Probability Plot.....	46
附圖 4.2-3 正規化後之RFM Box Plot.....	47
附圖 4.3-1 JavaSOM的系統架構.....	48
附圖 4.3-2 JavaSOM 程式運作架構.....	50
附圖 5.1-1 經JSOM所繪出網路拓樸圖 (16 * 16)	53
附圖 5.1-2 客戶分群結果分佈圖.....	57
附圖 5.1-3 平均單一客戶之R.....	58
附圖 5.1-4 平均單一客戶之F.....	59
附圖 5.1-5 各群平均單一客戶的收益貢獻.....	60
附圖 5.1-6 各群對企業收益表現比較.....	62

表格目錄

表格 2.1-1	資料庫行銷之定義.....	12
表格 2.4-1	分群方法回顧.....	34
表格 4.2-1	JSOM之資料格式定義.....	43
表格 4.2-2	RFM三項指標經正規化後的基本統計量.....	44
表格 4.3-1	JavaSOM程式結構.....	49
表格 4.3-2	JavaSOM參數設定檔.....	51
表格 4.3-3	JavaSOM相關參數設定.....	52
表格 5.1-1	客戶分群結果.....	56
表格 5.1-2	分群後R指標.....	57
表格 5.1-3	分群後F指標.....	58
表格 5.1-4	群別與收益貢獻表.....	59



1. 緒論

1.1. 研究背景與動機

對企業行銷人員而言，一直存在著二個重要課題：一是如何維持與舊顧客之間的關係，另一則是如何吸引新顧客。一般來說，企業大多著重於開發新顧客上，而忽略了與舊顧客維持長期關係。根據 Don Peppers 和 Martha Rogers 在「The One to One Future: Building Relationships with One Customer at A Time」一書中曾提到：開發一位新顧客所花費的成本是留住一位舊顧客的五倍。此外，大部分的企業每年平均有高達 25% 的顧客會流失，而若顧客流失率減少 5%，則企業的利潤將有 100% 的成長。

國際化及科技的潮流加上網路交易的發達，企業越來越容易收集到大量的顧客資料，為有效的運用這些資料，資料探勘(data mining)即成為熱門的資料分析工具，然而在做資料探勘時，集群分析是非常重要的一環。

為了長期吸引顧客，持續改善顧客滿意度、增進競爭力、維持同業領先地位，以達成企業流程再造的目標，是各服務業刻不容緩的努力方向。

David Schmittlein (1987) 認為：顧客資料庫的市場區隔能力遠大於人口統計變數，因而最佳的市場區隔基礎，應是顧客實際購買現有產品的行為。

1.2. 研究目的

在現行企業的微利營運模式中，為使行銷效益最佳化，亦即在企業實際營運中，找出最有價值，能為企業創造最大利潤的客戶群，並有效運用行銷

資源，將企業有限資源應用於最大效益的客戶群上，必可為企業創造更高之營運效能。

1.3. 論文架構

本研究分為六章。第一章緒論，包含研究背景及動機、研究目的與論文架構。第二章為文獻探討，包含資料庫行銷、RFM 分析模型、自組織神經網路、分群方法回顧圖。第三章為改良式客戶分群系統架構概論，包含了問題描述與規劃、研究架構。第四章為改良式客戶分群雛形建置，包含了資料收集與描述、資料前處理過程、資料分群模型。第五章為結果分析，包含了雛形模擬結果與行銷決策支援分析。第六章為結論與建議，包含了研究結論、研究限制與建議。



2. 文獻探討

2.1. 資料庫行銷

近幾年來，由於網際網路與資訊科技的發展，使得以資料庫行銷為主軸的關係行銷成為可能，且已逐漸取代傳統的大眾行銷，成為新的行銷思維。隨著 90 年代整體行銷環境的改變，行銷人員所面對的不再是一個均質的大眾市場，而是一個個具有個別差異的顧客。

不同於傳統的行銷活動以「交易」為核心，資料庫行銷是以「顧客」為核心所發展出來的行銷策略。資料庫行銷不只是企業對顧客的單向溝通管道，而是企業與顧客之間形成持續性互動的學習性關係。當該顧客與企業產生互動時，企業的窗口便可以立即地得到之前與該名顧客互動的所有記錄，並提出更符合顧客需求的解決方案。因此，對企業而言，與顧客產生互動的絕不會只有業務人員，諸如客服人員、技術維修人員，甚或是貨物運送人員，都算是整體資料庫行銷的一環。在進一步探討資料庫行銷之前，我們有必要對資料庫行銷的定義有更深一層的認識。在表格 2.1-1 中，我們將各學者對於顧客關係管理之定義加以彙整。

學者	定義
David Shani 和 Sujana Chalasani (1992)	「資料庫行銷」是蒐集現在或以前顧客的資料，建立起一個資料庫，來改善市場行銷的績效。
Frank Cespedes 和	「資料庫行銷」是利用顧客資料，來改善企業對行

Jeff Smith (1993)	銷資源的應用，以針對特定對象傳遞更有用的資訊，並維持與顧客之間良好的長期關係。
Arthur Hughes (1994)	「資料庫行銷」是以電腦科技來管理一套關於既有顧客與潛在顧客相關資訊的資料庫系統，使企業可以利用這些資料提供顧客較佳的產品或服務，並與顧客建立起良好的長期關係。

表格 2.1-1 資料庫行銷之定義

由上述各學者對於資料庫行銷的定義可知，資料庫行銷的執行與電腦或資訊科技的進步有極大的關聯性。這是因為當行銷人員再面對龐大的顧客資料時，唯有利用電腦方能處理大筆的資料。因此，資料庫行銷的真正意涵在於藉由行銷研究的統計模型，來分析與處理大量的顧客資料，以作為行銷決策之參考依據。

由於資料庫行銷在顧客關係管理中的角色為一項分析工具，而林慧晶（1997）提出了資料庫行銷能為企業帶來哪些行銷決策上的利益做了一個整理。

（一）顧客價值分析

資料庫行銷的最主要功能是針對顧客進行價值分析。傳統上，雖然企業可以很清楚地知道每日的銷售額有多少，但是卻很難能將個別顧客與銷售情況作一連結。而透過資料庫行銷的分析，企業可以很容易地對顧客進行價值分析，並針對不同價值的顧客進行不同的資源分配，以及採取不同的行銷策略。

（二）計算顧客終身價值

根據 Arthur Hughes (1994) 所言：「所謂的顧客終身價值 (Customer Lifetime Value) 是指在未來一段時間之內，企業或廠商可以從個別顧客獲得之利潤的淨現值。」而藉由資料庫行銷，企業可以依據資料庫中顧客的購買記錄，算出每位顧客可能貢獻於企業的終身價值。透過顧客終身價值的計算，企業除了可以預測未來的營收情況外，還可以確認出顧客價值的高低，以分配不同的企業資源於不同價值的顧客身上。

(三) 進行向上銷售 (Up-selling) 與交叉銷售 (Cross-selling)

根據 Bob Stone (1995) 所言：「所謂的向上銷售則是指企業可以針對顧客目前所購買的產品項目，推測其往後可能會需要的品項。而所謂的交叉銷售是指針對顧客目前所購買的產品項目，進行相關產品的銷售服務。」因此，針對資料庫中顧客的購買品項記錄加以分析，企業可以很輕易地達到向上銷售和交叉銷售的目的。

(四) 行銷決策支援系統

Little (1979) 認為：「所謂的行銷決策支援系統 (MDSS, Marketing Decision Support System) 是指將顧客的購買記錄透過模型分析，再利用模型分析出來的結果配合專家知識，使決策者能作出有利的決策。」由此可知，顧客資料與模型分析是資料庫行銷的兩大重要要素。因此，資料庫行銷的功能並不止於幫助企業管理其顧客，更重要的是可作為企業的行銷決策支援系統。

2.2. RFM 分析模型(RFM Analysis Model)

依據 Hughes 於 1996 年文章中描述，RFM 分析模型在直效行銷 (DirectMarketing) 領域中已使用超過 30 年。但由於電腦科技的進步及資料

庫系統的成熟應用，才使 RFM 分析模型在 1990 年以來廣泛被使用，成為資料庫行銷(Database Marketing)中很重要的應用。

RFM 分析模型是利用顧客過去的歷史交易記錄，包括最近一次的購買日期(Recency)、一段期間內的購買頻率(Ferquency)及一段期間內的購買金額，(Monetary)來進行顧客價值的衡量。如附圖 2.2-1 所示。

Customer	Recency	Frequency	Monetary
A	1	3	1
B	2	1	3
C	4	2	10
D	5	1	8
:	:	:	:
Y	7	2	7
Z	2	4	6

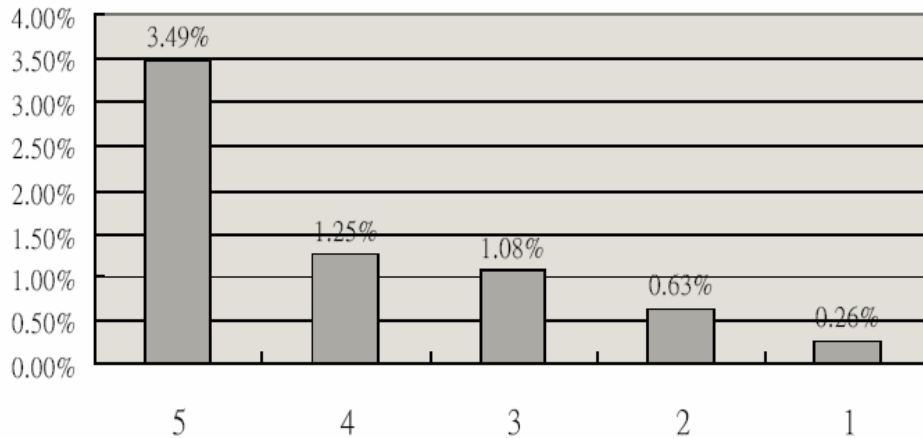
附圖 2.2-1 RFM 客戶分析模型

1. 最近購買日期

最近購買日期是衡量顧客最近一次的購買日期距現在的天數。天數越小表示顧客購買產品的時間距現在越近。首先將要分析的顧客交易資料以購買日期進行排序，依距離目前日期分為五等分(Exactly Equal Size)每一等分等於整個資料庫的 20%，最接近現今日期的 20%編為「5」、20%~40%編為「4」、40%~60%編為「3」、60%~80%編為「2」、80%~100%編為「1」。

Hughes 對這五等分的直效行銷顧客進行實驗，發現等級越高也就是最近

購買日期越近的顧客有較高的重複購買比率。如圖 7 所示。R5 的顧客回應率 3.49% 是 R4(1.25%) 的 2.79 倍。表示最近購買過產品的顧客對公司有較高的印象，所以有較高的可能會再重複購買產品。



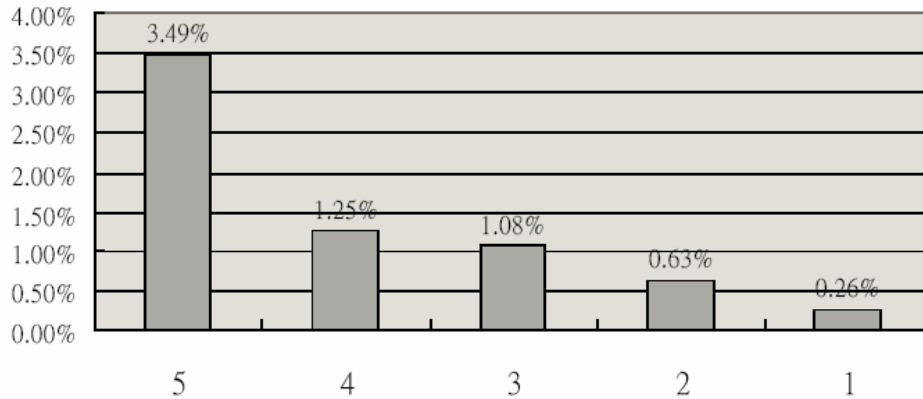
附圖 2.2-2 RFM 模型最近購買日期

2. 購買頻率

購買頻率是指顧客在一段期間內，購買產品的次數，次數越高也就是頻率越高。運作方式是首先統計一段期間內顧客的購買次數，接著依次數排序，次數最多的前 20% 編為「5」、20%~40% 編為「4」、40%~60% 編為「3」、60%~80% 編為「2」、80%~100% 編為「1」。計算購買頻率除了統計購買次數外，還有購買產品數量，電話通話次數(電信業)，支票與存提款的次數(銀行業)……等，不同產業有不同的頻率計算方式。只要能衡量顧客與企業接觸(交易)的頻率都可以做為衡量屬性。

依據 Hughes 對這五等分的直效行銷顧客進行實驗，發現等級越高回應率越高，但差異程度已沒有最近購買日期「R」的大，如圖 8 所示。購買頻率「F」也是很好衡量回應率的方式，但效果沒有最近購買日期「R」來的好。同時有一個奇特現象，F 等級「1」的顧客有部份是全新的新買主還沒有機會成為頻

繁的買主，而這些顧客在 R 的等級是最高的「5」，所以會產生 F 等級「1」的回應率(0.93%)高為等級「2」的回應率(0.92%)。

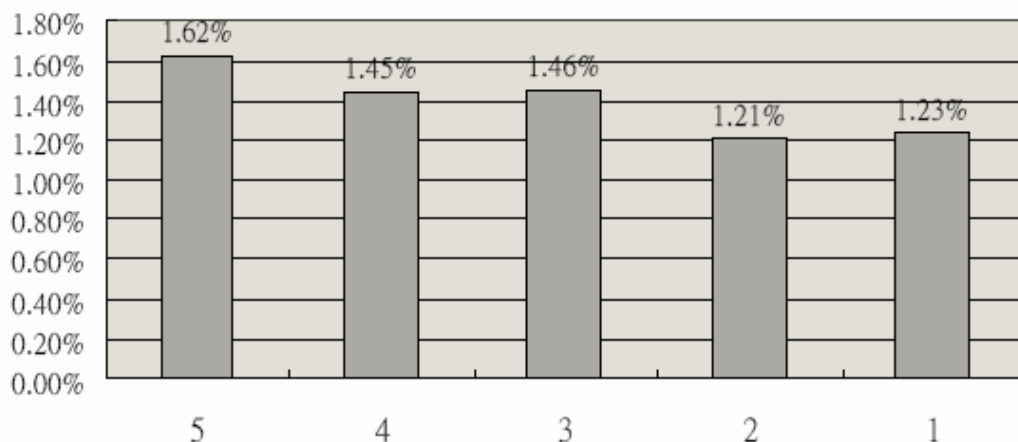


附圖 2.2-3 RFM 模型購買頻率

3. 購買金額

購買金額是指顧客在一段期間之內，購買產品的總金額。運作方式是首先統計一段期間內每位顧客的購買產品總金額，接著依金額大小排序，金額最多的前 20% 編為「5」、20%~40% 編為「4」、40%~60% 編為「3」、60%~80% 編為「2」、80%~100% 編為「1」。

依據 Hughes 對這五等分的直效行銷顧客進行實驗。在產品價格 100 元(美元)以下，屬於低單價品項，M 回應率的情況如圖 9 所示。購買金額 M 對回應率的影響顯然差很多。這是因為收入高的人，他的信件會太多或他的秘書會過濾信件。而收入低的人收到的信件少很多，將信件打開的機會也會比較大。同時收入高的人將信件打開後，若喜歡可以立刻買下信中銷售的產品，而收入低的人就必須考慮這是不是他的預算之內。因此購買金額的回應率結合這兩種相對的情況，拆信可能性及購買能力，以致圖形呈現較水平的情形。

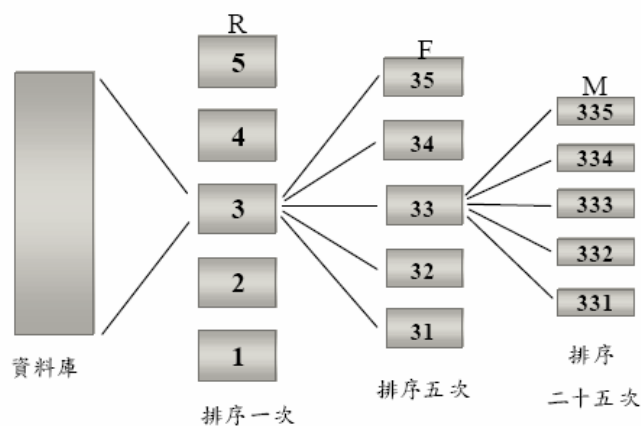


附圖 2.2-4 RFM 模型購買金額

編號完成之後，每位顧客在資料庫中的記錄會有三個數字，從 555，554，553，552，551，545，544……111，共有 125 種組合。將 RFM 分數組合轉換為三度空間向量。

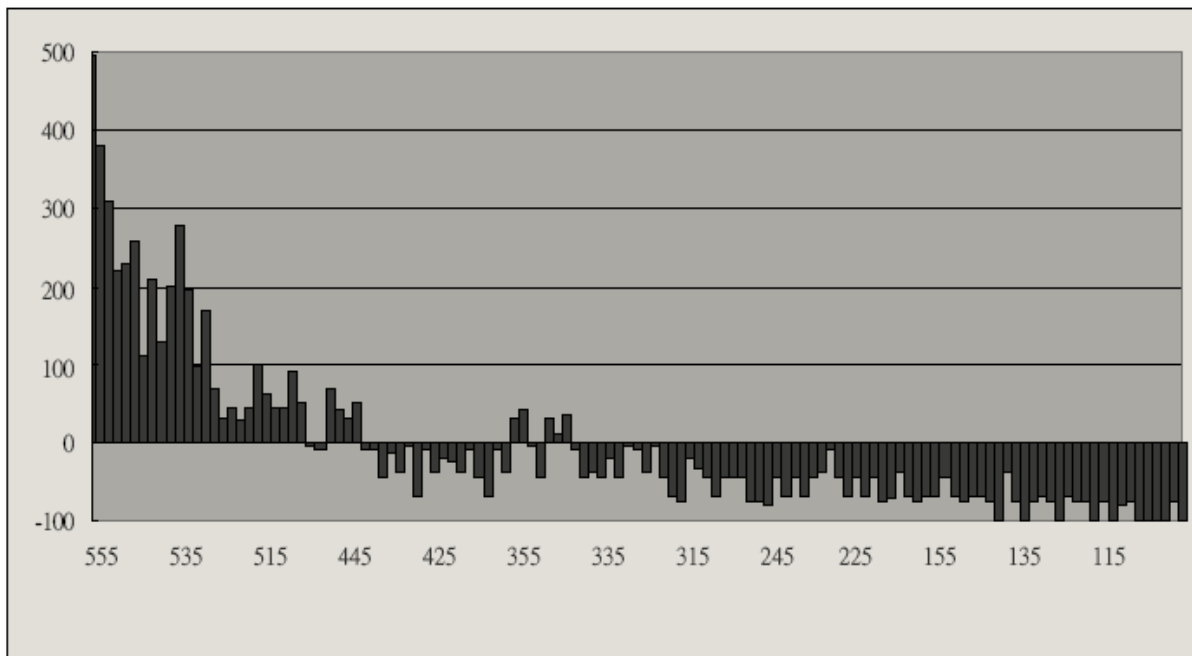
計算收支平衡指數 (Break Even Index, BEI)

$$BEI = \left(\frac{r - BE}{BE} \right) * 100$$



附圖 2.2-5 RFM 組合轉換

由圖可知大於 0 表示收益為正值，小於 0 為負值，表示該類顧客從銷售產品所得利益不足支付郵遞成本。



附圖 2.2-6 RFM 指標排序結果

2.2.1. RFM 分析模型限制

RFM 分析模型具有可適用於各個行業的特性，且計算邏輯簡單許多行銷人員在不需專屬資訊系統輔助下也可進行顧客分析。因此在企業界已使用一段很長時間。但 RFM 分析模型存在二項問題：

一是個別 RFM 屬性針對不同產業有不同的敏感性。例如某些產業對 R 屬性有很好的區隔能力，但對其他產業可能對 FM 屬性有較佳的區隔能力，而 RFM 分析模型無針對 RFM 屬性敏感性的不同，整合成單一的區隔指標。進而對於 RFM 三個指標所產生的權重也因產業的不同而有不同的分配。

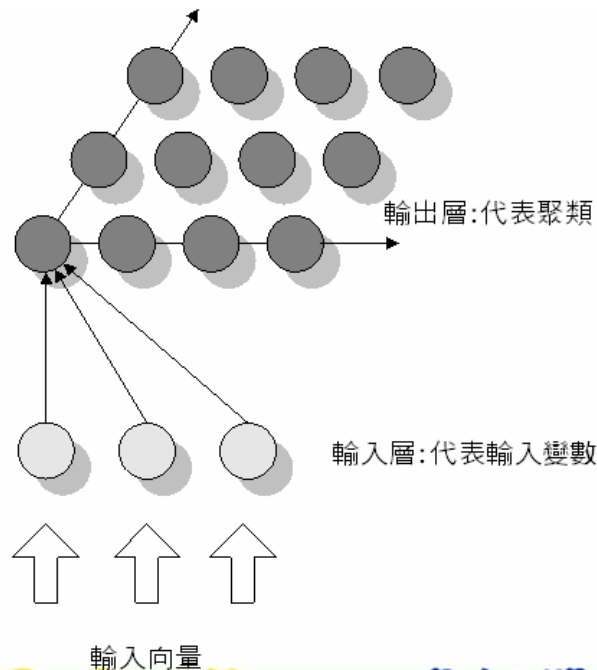
二是 RFM 的分群方法是採用等間距的區份客戶分群，但此種分群方法在

實際的情形中並不實際。

本文中即嘗試利用 Data Mining 中的資料分群運算法則，嘗試取代傳統 RFM 以等區分的評價方式，期望彌補 RFM 的缺陷。

2.3. 自組織神經網路(self-organizing map, SOM)

自組織神經網路(self-organizing map, SOM)，或又稱作自我組織特徵映射圖網路(self-organizing feature map network, SOFM)，是由芬蘭 Helsinki 大學 Teuvo Kohonen 教授所提出(Kohonen 1984; Kohonen 1982)。自組織神經網路基本為一個一維或二維空間的網路圖，輸入向量 $x_i = [x_{i,1}, x_{i,2}, \dots, x_{i,m}]^T$ 可藉由映射(mapping)投影的方式，將維度之輸入向量對映至二維或一維的空間上，此一功能可提供高維度之資料可以二維空間來呈現其彼此之相似性，以利資料分群之進行(Vesanto and Alhoniemi 2000; Simula et. al. 1998)。自組織神經網路具有臨近區域相似性高的特色，輸入層與輸出層皆同屬在一個一維或二維的空間上，以特定拓樸結構(topological structure)如距陣(grid)或多邊型之排列法來聚集網路層之神經元。



附圖 2.3-1 自組織特徵映射網路架構

SOM 的學習演算法可分為三大部分，收先定易於神經元所構成的一維或二維矩陣；其次尋找優勝神經元，其輸出值可反映出對目前輸入最有反應的神經元；最後是調整優勝神經元及其鄰近區域內神經元的權重向量，使其更接近輸入向量。

以下將 SOM 演算學習法分為四個步驟。如下分述：

【步驟一】：初始化

在進行 SOM 網路訓練前必須針對網路結構與權重向量進行初始化。在網路架構定義方面，基本上 SOM 的網路拓樸是由 i 個單元的集合所構成，並根據事先定義的固定拓樸型態來排列，最常使用的就是 $N * M$ 的二維網格。其次，每一個單元 i 都會被賦予與輸入資料相同維度的權重向量 m_i , $m_i \in \mathcal{R}^n$ ，而權重向量可有兩種方式決定，一是以隨機亂數指定，或以隨機策略如主成份分析(principle component analysis, PCA)來設定。但所有 i 個權重向量的初始值均應相異，且通常會加以正規化(normalize)成為長度為 1 的單位向

量。

【步驟二】：輸入特徵向量

針對時間/訓練循環，輸入向量 $x_t = [x_{t,1}, x_{t,2}, \dots, x_{t,m}]^T$ ，在此使用了一離散時間標記 t ，代表目前訓練重複次數。在每次訓練循環， $x(t)$ 均由輸入向量集合 \mathfrak{R}^n 中隨機抽出。

【步驟三】：尋找優勝神經元

以最小歐基里德距離的方式尋找時間/訓練循環 t 之優勝單元 c ，顯示其含有最高活動力。而對於隨機選擇的輸入向量 $x(t)$ 而言， $c(t)$ 則將顯出更高的活動力。亦即 $c(t)$ 在特定輸入向量的未來表現皆會因為呈現較高的活動力而較為合適成為優勝單元。一般而言，一個單元的活動力（作用值）是根據輸入向量(input pattern) 與該單元的權重向量（鏈結強度）之間的歐幾里得距離(Euclidian distance)來衡量。換句話說，若有一單元其權重向量 m_i 與目前輸入向量 x_t 之間的歐幾里德距離為最小時，該單元便稱為優勝單元。因此，優勝單元 c 的選擇方式可用下列算式（式 2-1）加以表示

$$c(t) = \|x(t) - m_c(t)\| = \min \{ \|x(t) - m_i(t)\| \} \quad (\text{式 2-1})$$

$c(t)$ 時間點 t 的優勝單元

$x(t)$ 時間點 t 隨機抽取的輸入向量

$m_i(t)$ 時間點 t ， i 單元的權重向量

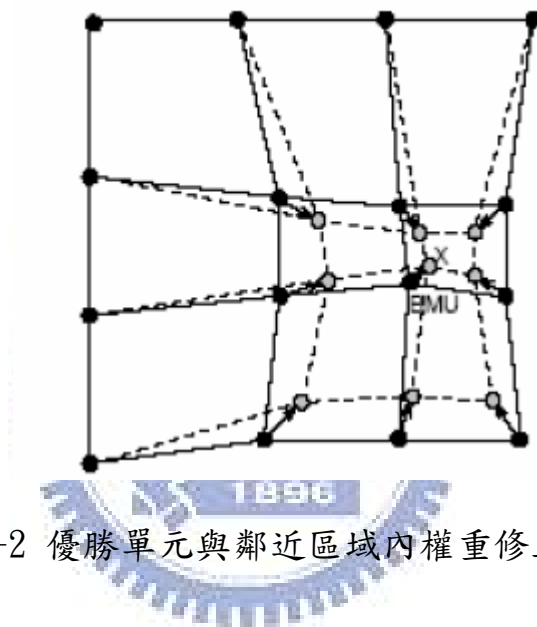
$m_c(t)$ 時間點 t ，優勝單元 c 的權重向量

$\|x(t) - m_c(t)\|$ 在時間點 t ，優勝單元 c 的活性(activity) Δm_i ，用來計算權重向量修正量

【步驟四】：調整權重向量

適應行為(adaptation)發生在每次學習重複過程中，其執行方式是根據

各別輸入向量與權重向量之差異量，朝向坡降（gradual reduction）方式進行學習。而至於適應調整的總次數是由學習速率 α 主導，其亦會隨著時間過程而逐漸遞減。這種適應(調整)強度的漸減特性使得在學習初期有大量的適應步驟，並且權重向量亦需從隨機初始值逐漸調整轉向至輸入向量的實際需求(群心)。而在學習末期，微幅的適應讓權重向量在輸入空間宛如進行微調動作(如附圖 2.3-2 所示)。



附圖 2.3-2 優勝單元與鄰近區域內權重修正示意圖

對於位在優勝單元周圍鄰近區域的單元，就輸出空間而言，可用與鄰近中心的距離 h_{ci} (neighborhood function)，也就是該單元 i 與該次學習循環的優勝單元 c 之間距離加以間接的表示。換句話說，我們以該次學習循環的優勝單元 c 當作是輸出空間中的鄰近中心，並指派由 0 到 1 的調整量（學習速率），以確保距離優勝單元愈近者，其適應調整程度也就愈大。

在結合上述 SOM 的原理原則後，我們可以寫出一如公式(式 2-2)的學習法則來表示鄰近區域內單元之權重修正。在此使用離散時間標示 t 以代表目前學習循環次數。

$$m_i(t+1) = m_i(t) + \alpha(t) \cdot h_{ci}(t) [x(t) - m_i(t)] \quad (\text{式 2-2})$$

t 表示目前學習循環次數

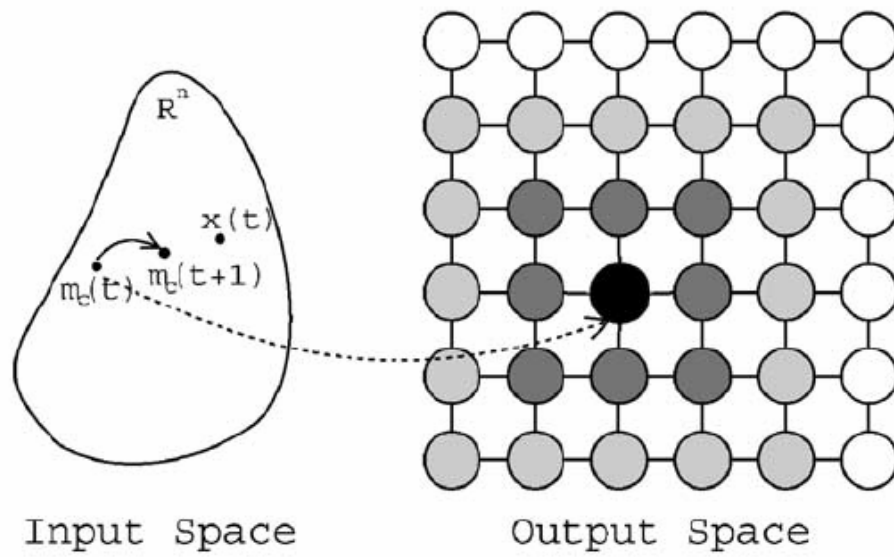
α 則代表隨時間變化的學習速率

h_{ci} 則表示隨時間變化的鄰近函數(neighborhood function)值

x 表示目前的輸入樣式(向量)

m_i 則代表分配給單元 i 的權重向量

圖 2.4-2 的簡單圖形可呈現 SOM 的結構與學習流程。在該圖中其輸出空間為 6×6 共 36 個單元的網格所構成。然後一個隨機選取的輸入向量 $x(t)$ 映射至輸出單元的網路格點上。在下一階段的學習過程中，顯示出最高活性(activation)的優勝單元 c 則被選擇。圖 2.3-3 中標繪成黑色的單元視為優勝單元，而其權重向量 $m_c(t)$ 朝向目前的輸入向量 $x(t)$ 移動，這樣的移動表現在圖 2.3-3 左側的輸入空間中。由修正(adaptation)的結果看出，單元 c 在下一次的學習循環($t = t+1$)就輸入樣式而言，將會產生更高的活性(作用值)，因為單元的權重向量 $m_c(t+1)$ 在輸入空間中更加接近輸入樣式 x 。除優勝單元外，對於鄰近單元亦修正其權重向量，受到修正調整的單元由圖中輸出空間標繪有陰影的單元可看出，其陰影的深淺即對應於該單元的權重修正量，亦可代表其鄰近函數的空間寬度。一般而言，在優勝單元的鄰近區域中愈靠近中心者所受到的修正愈強，這可由圖中單元標繪的深淺顏色說明。

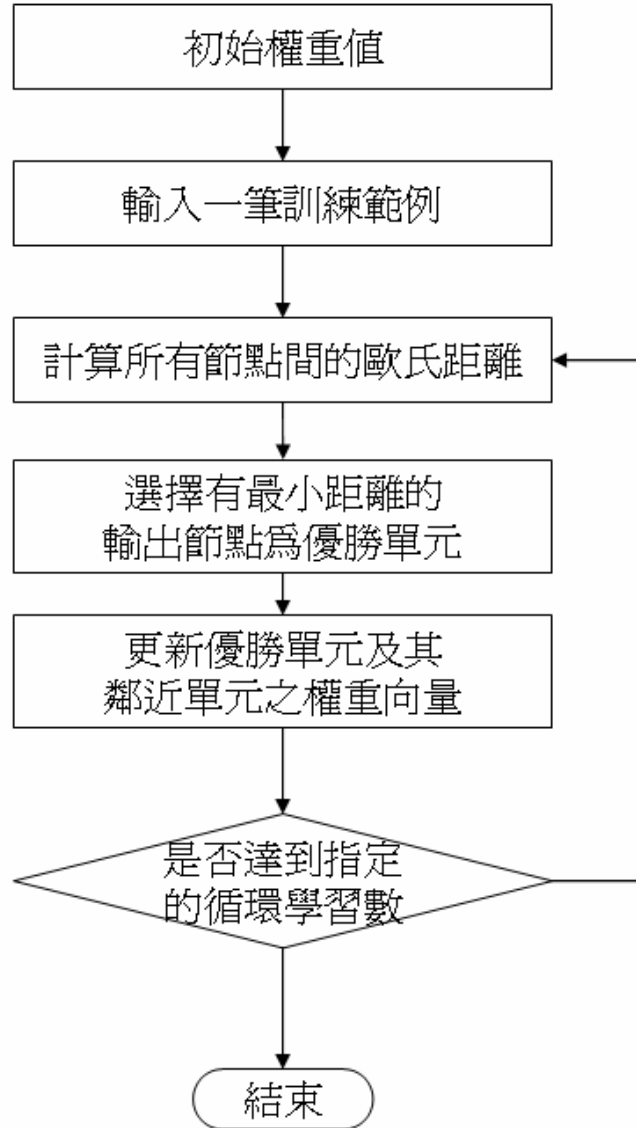


附圖 2.3-3 由 SOM 輸入空間與輸出空間權重修正情形[Dittenbach, Rauber, &Merkl, 2002]

【步驟五】：返回步驟二，直到特徵映射圖形成後終止。

歸納其上所述步驟，可以以下流程圖說明概要。





附圖 2.3-4 SOM 演算法學習步驟流程圖

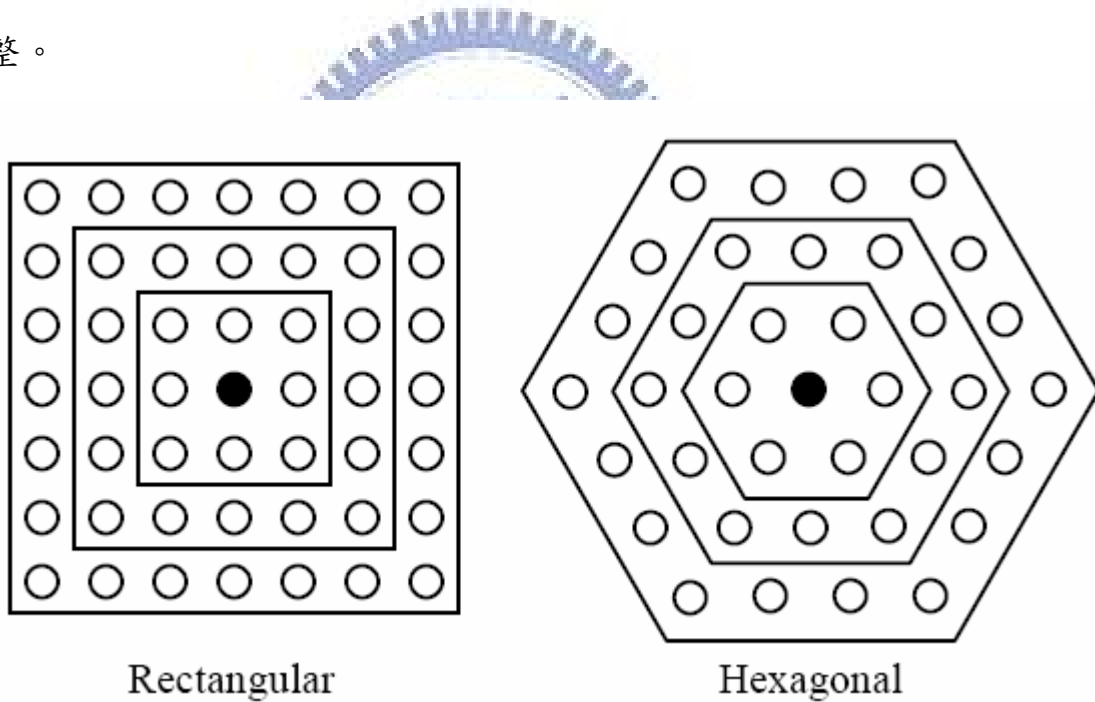
【參數設定】

1. 學習速率參數 $\alpha(t)$ ：本參數為用來調整權重向量且應隨著時間而調整，一般而言將隨時間逐漸變小。至於參數遞減的形式可以是線性遞減、指數遞減或是與時間成反比等，此時的學習循環可以視為演算法的「排列階段」

(ordering phase)。相對應於「排列階段」的是「收斂階段」(convergence phase)，即之後的學習循環的，其主要目的在於進行特徵映射圖的細部調整，此時學習循環的參數值應保持相當小的數值。

2. 鄰近區域參數 $h_{ci}(t)$:

鄰近區域函數通常定義包圍著優勝神經元 c 的正方形區域，其形式可以是矩形、六邊形或八邊形等(如圖 2.3-5 所示)。不管是何者形式都應遵守一開始先包含全部或較大範圍的神經元，然後隨時間增加而慢慢縮減鄰近區域的大小。舉例來說，在「排列階段」可以隨時間而縮減至較小範圍；而在「收斂階段」則應僅包含一個或兩個神經元，甚至只針對優勝神經元進行權重向量的調整。



附圖 2.3-5 常見的鄰近區域型態

高斯函數(Gaussian function)可以用來定義鄰近區域與核心之關係，及鄰近區域函數 $h_{ci}(t)$ ，如公式 (2-3) 所示：

$$h_{ci}(t) = \exp\left(-\frac{\|r_c - r_i\|^2}{2 \cdot \delta(t)^2}\right) \quad (\text{式 2-3})$$

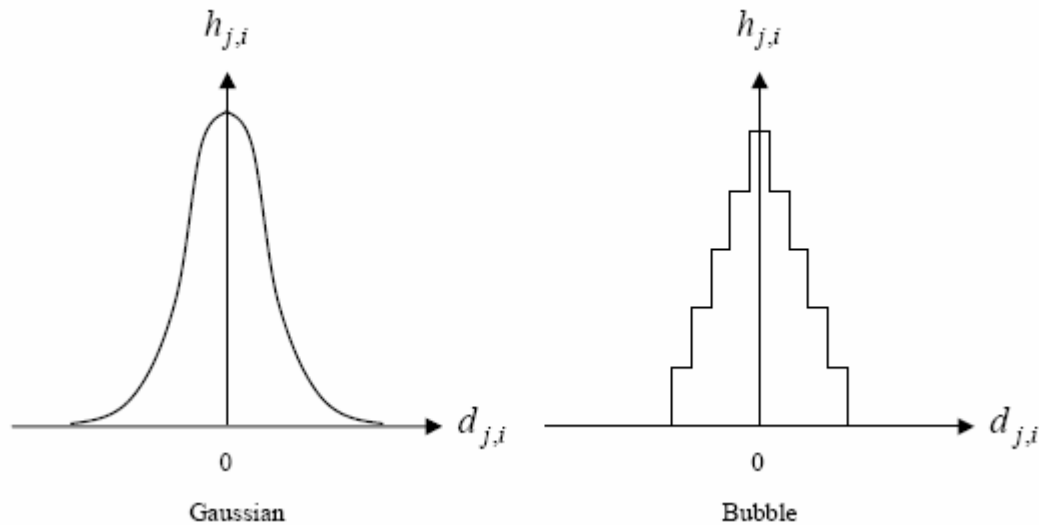
其中 $\|r_c - r_i\|$ 代表代表在輸出空間中單元 c 與單元 i 之間的距離(也就是權重向量差);換句話說, r_i 表示在輸出網格內的指向單元 i 的二維向量(也就是權重向量強度), δ 為隨著時間變化的鄰近半徑縮小因子。

相對應於高斯函數, 有另一較簡單且運算速度較快的鄰近函數定義:

Bubble Function, 如式 2-4 所示。

$$h_{j,i(\bar{x})}(t) = \begin{cases} 0 & \text{if } i \in N_j \\ 1 & \text{if } i \notin N_j \end{cases} \quad \forall n \in \mathfrak{R} \quad (\text{式 2-4})$$

然而, 高斯函數相對於 Bubble Function 有較佳的收斂速度[Z. P. Lo, B. Bavarian, 1991]。如圖 2.3-6 所示。



附圖 2.3-6 兩種型態的鄰近函數

SOM 是目前相當受到歡迎的非監督式類神經網路模型, 原本的應用領域用來處理工程問題, 但漸漸普遍應用於資料分析的領域上, 尤其近年來資料探勘的議題研究日漸普遍, 此方法也月見其廣泛應用。SOM 擁有下列優點 [Vesanto, 2000], 使其成為資料探勘的工具之一:

- 穩健性(robustness)：假設鄰近區域函數延伸至足夠遠如高斯函數，則 SOM 會擁有相當穩健特性，這是因為競爭式學習所產生的原型會受到所有資料樣本之影響。
- 局部調整(local tuning)：拓樸的次序性將在每個優勝單元的鄰近區域發揮作用，因而形成朝向資料密度特性作局部調整。
- 易現性(ease of visualization)：SOM 有規則的網路格點可使建立一個有效率且視覺化的使用者介面容易許多。

以上優點大多歸因於 SOM 鄰近區域間的關係，這也是構成 SOM 架構的基礎。

SOM 演算法的四個特性：

A. 輸入空間的趨近性(approximation of the input space)

特徵圖形經由權重向量集合展現在輸出空間，並且針對輸入空間提供一個極佳的趨近特性。換句話說，SOM 的目的為經由找尋較小的樣式集合來儲存大量的輸入向量，以便針對原始的輸入空間提供一個較佳的趨近性。這理論基礎為「向量量化」，也是資料降維或壓縮的主要原理。

B. 網路拓樸的次序性(topological ordering)

經由 SOM 演算法計算所得到的特徵圖具有一拓樸次序性，就是在輸出網路中神經元的空間位置與特定輸入樣式的特徵相對應。網路拓樸的次序性來自於將優勝單元的迫使權重向量朝向輸入向量調整；此時權重修正亦將使得最靠近優勝單元的鄰近神經元產生向中央權重修正的效果。如此便使得整個拓樸特徵圖形所形成的輸出空間產生適切的次序性，並以虛擬的拓樸網格來呈現。網格上的每個輸出單元均可以其相對應的權重向量作為輸入空間的座標值。因此，若是輸出空間中的相鄰單元，其在輸入空間相對應資料點亦為相連時，便可直接觀察到網路拓樸的次序性。

C. 機率密度的對應性(density matching)

SOM 特徵圖同時也反映出輸入分配的統計量變化情形，也就是說，樣本訓練向量中發生的機率密度較高者在輸入空間所佔的區域將映射至輸出空間的較大部分區域，因此，相較於輸入空間中機率密度較低者，擁有較佳的解釋能力(resolution)。

D. 特徵的選擇性(feature selection)

從輸入空間中給定一非線性分配，SOM 擁有掘取一組最佳特徵集合來詮釋(或趨近)資料分配之能力。這項特徵同時也是前三項特性之集合。儘管主成份分析法可藉由關聯矩陣中具有最大特徵值之向量計算，求得訓練資料中擁有最大變異的輸入維度(向度)，但仍僅限於線性或平面的輸入空間；至於曲線或曲面(surface)時，主成份分析的表現便不如 SOM 的拓樸次序特性。

然而 SOM 演算法卻也有以下缺點：【Vesanto, 2000】

- 邊際效應(border effect)：鄰近區域的定義在 SOM 拓樸圖形的邊緣為非對稱性。因此中央區域的單元，其鄰近區域函數(密度估計)必不等同於邊界單元。
- 收縮效應(contraction)：在向量量化過程中的平均分配使得變數數值的範圍縮小，並且受到鄰近函數的增強，極端值將因此去除，這在某些情況如分析者關切離散值時是不樂見的。
- 內插單元(interpolating units)：當資料群的分佈為不連續時，在資料群之間插入單元可便於資料分配的推估；然而，對於某些分析工具如單一連結分群法的例子顯示如此作法可能會提供錯誤的資料形狀線索。

2.4. 分群方法回顧

集群(或稱聚類、群聚、叢集)分析(cluster analysis)是一種將眾多個體或目標對象歸類為若干未知的分群，有就是在資料中找尋群組的學問，也是將類似的目標對象歸聚成一群之行為。與分類不同的是，分群的數量及其特性必須從資料中獲取而無法事先得知。集群分析是一門所以群集分析可說是各種用來找出資料集合中相似資料的數學方法之通稱。換句話說，群集分析嘗試將資料點歸類成同質性的群集，並假設無法事先得知群集資訊。

而分析的第一步就是檢查資料點彼此之間的鄰近性(proximity)，因此亦可視為一種探索資料的分析技術。若是以分析結果而言，群集分析是一種將觀察資料結合成群類(groups)或群集(clusters)的技術，使其可達到以下兩個目標。其一是每個群類或群集之內，就某種特性而言，具有高度同質性或表現出緊密結實的分布狀態；其次是每個群組或群集之間，就某種特性而言，具有高度異質性，而不同群組或群集中的觀察值彼此相異。因此整體而言，群集分析是一種能根據資料變數之相似性與相異性，客觀地進行分類分群的邏輯程序，其目的在根據某種特性而劃分成的多個群集中，同一群集具有高度同質性(homogeneity)，而不同群集間則具有明顯的異質性(heterogeneity)。而現今群集分析一詞較常指尋找資料中隱含群集的程序之通稱。

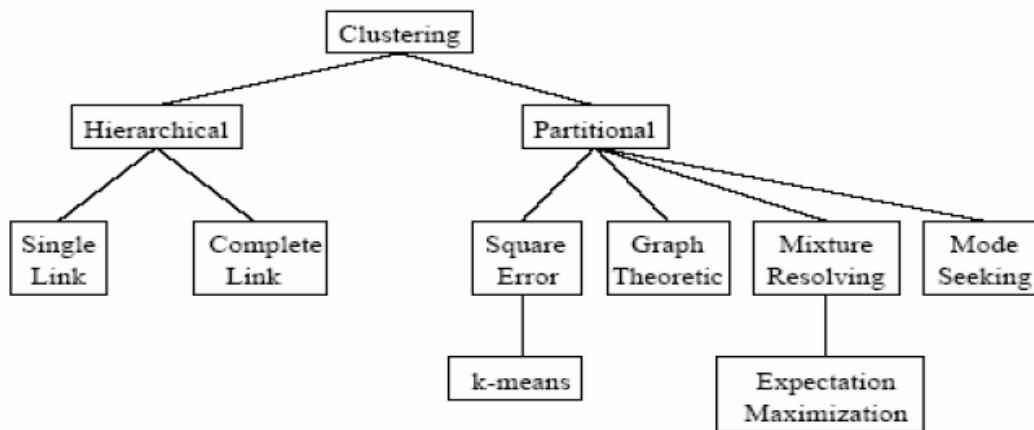
分群觀念與方法已經成功應用於許多領域中，諸如生命科學、醫藥科學與工程科學等，並且在不同背景有其不同專有名稱。例如人工智慧中樣式辨認的非監督式學習(unsupervised learning)，生物學與生態學的數值分類學(taxonomy)，社會科學的拓樸(topology)，圖論中的分割(partition)、心理學的 Q 分析與行銷研究中常提到的市場區隔(segmentation)等【Halkidi,

2001；Everitt, 2001】。

傳統分群法基本上可分為階層式(hierarchical)與分割式(partitional)兩大類，且各自分別有其多元化的演算法，如附圖 2.4-1 所示。

其中最具代表性的是階層式的凝聚分群法與分割式的 K-means 分群法【Vesanto, 2000】。隨著人工智慧的興起，遂產生模糊分群(fuzzy clustering)、類神經網路分群與演化式分群法(evolutionary clustering)，其中類神經網路中最具代表性者即為自組織映射圖網路(SOM)，而演化式分群法則為遺傳演算法【Jain et al., 1999】。

在對分群的回顧研究中，共列舉介紹 11 項分群相關技術，如附表 2.5-1【楊東昌, 2004】所示。除上述所列之外，尚包括以搜尋法為基礎(search-based)的分群法，如模擬退火法(simulated annealing)；接近直觀式的最鄰近者分群法(nearest neighbor clustering)；為處理現今大型資料型態，如 CLARANS 與 BIRCH 分群法；以及因應分群使用者之實際需求而加入局部性限制條件之分群法。



附圖 2.4-1 分群法分類[Jain & Dubes, 1988]

	分群方法	簡述
--	------	----

1	<p>階層式分群法 Hierarchical Clustering Algorithms</p>	<p>依次將較小群集合併或將較大群集分割，使得分群結果為反映資料結構的樹狀群集。藉由在樹狀圖之特定高度水準切割，可將資料分成若干數目的群集。</p>
2	<p>分割式分群法 Partitional Algorithms</p>	<p>以反覆程序直接將資料分解成若干分離群集，使某能量函數最佳化。</p>
3	<p>混合求解與模式尋找分群法 Mixture-Resolving and Mode-Seeking Algorithms</p>	<p>目的在確認從分配中產生的資料樣式，藉由估計元素密度的參數向量之最大可能性，來確定各分配所屬參數或甚至其數量。經參數評估而置於同一元素的樣式則可視為同一群集。</p>
4	<p>最近鄰近者分群法 Nearest Neighbor Clustering</p>	<p>將輸入樣式分配至與最鄰近已被指派的樣式之同一群集中，設定距離門檻值加以控制分群過程，直到所有樣式皆被分配完畢為止。</p>
5	<p>模糊分群法 Fuzzy Clustering</p>	<p>使用模糊技術進行資料分群，單一資料可以分配於超過一個群集中，並給予一個介於 0 到 1 的隸屬度(degree of membership)，使得日常生活所遭遇到具有不確定性的實際資料均可以適用。</p>
6	<p>群集表現分群法 Representation of Clusters</p>	<p>透過群心點、分類數節點或是邏輯符號的聯集等方式將來表現分群結果所產生的資料群集之分佈及其結構，以達到資</p>

		料萃取之目的。
7	類神經網路分群法 Artificial Neural Networks for Clustering	基本概念源自生物神經網路，因具備處理數量化資料，並擁有平行分散式處理架構，且透過神經元間的權重修正可正確學習樣式特徵。
8	演化式分群法 Evolutionary Approaches for Clustering	由自然界演化啟發動機而來，並將所有可行解進行染色體編碼，被合使用演化操作因子如選擇、重組與突變等，來尋找全體母體中最佳的資料分割。
9	搜尋式分群法 Search-Based Approaches	可分成固定與隨機式搜尋法以獲得準則函數之最佳值，固定式搜尋藉著徹底列舉方式保證達到最佳資料分割，而隨機搜尋則先產生一合理而接近最佳解的分割方式，且以漸進方式保證收斂至最佳分割的結果。其中常用的為模擬退火法 [Aarts & Korst, 1989]。
10	限制式分群法 Incorporating Domain Constraints in Clustering	分群具有主觀的本質，因此需要根據時空環境或使用者需求而制宜，每種演算法在執行分群時皆牽涉到使用某種直接知識，如專業意見；或是間接知識，如選擇相似度的衡量方法或分群演算法，甚至在特徵選擇與編碼時是否採用經驗值等。

11	<p>大型資料集合分群法</p> <p>Clustering Large Data Sets</p>	<p>針對數以百萬計高維度的樣式資料集合進行分群以達到資料萃取，尤其是多媒體影音型態，目前最佳求解法如以基因演算法、塔布搜尋法與模擬退火法皆僅適用小規模資料集合。因此，收斂性的 K-mean 演算法與 Kohonen 的 SOM 受到矚目 [Mao & Jain, 1996]。隨著資料探勘學門的興起，遂刺激新的分群法產生，如 CLARANS [Ng & Han, 1994] 與 BIRCH [Zhang et al., 1996]</p>
----	--	---

表格 2.4-1 分群方法回顧

本論文即是結合變形的類神經網路學習方法依據客戶交易資料所提出的客戶分群模型，並舉一實例驗證其客戶分群結果並進而對市場行銷策略做一參考建議。

3. 改良式客戶分群系統架構概論

3.1. 問題描述與規劃

經濟學家帕列托所提出所謂 80/20 法則，說明了原因和結果、努力和收穫之間，存在著不平衡的關係。而典型的情況是：80%的收穫，來自 20%的付出，也就是 80%的結果，歸結於 20%的原因。而在實際的情形下也是如此，20%的重要客戶貢獻了利潤的 150%；而最差的 40%客戶，使利潤縮減 50%，可見客戶區隔的重要性。

然而，隨著科技的潮流加上網路交易的發達，企業面對大量的顧客資料中，尤其經濟規模已由高成長高獲利的商業模式進入穩定成長與微利的商業模式，企業在尋求客戶的重要性與並將企業最重要的資源置於最重要的客戶上，並對企業營收與成長幫助不顯著的客戶從新檢討其對應的行銷策略，對於競爭激烈的成熟市場已是企業中亟需面對的問題。

對於客戶的維持與改善客戶關係一直是企業中常常被提出來的問題。為了長期吸引顧客，持續改善顧客滿意度、增進競爭力、維持同業領先地位，企業必須不斷且長期投入資本與人力改善客戶關係，然而企業資源有限，應用帕列托所提出的 80/20 原則，並研究顯示，20% 的好顧客貢獻了利潤的 150%，而最差的 40%顧客，使利潤縮減 50%。

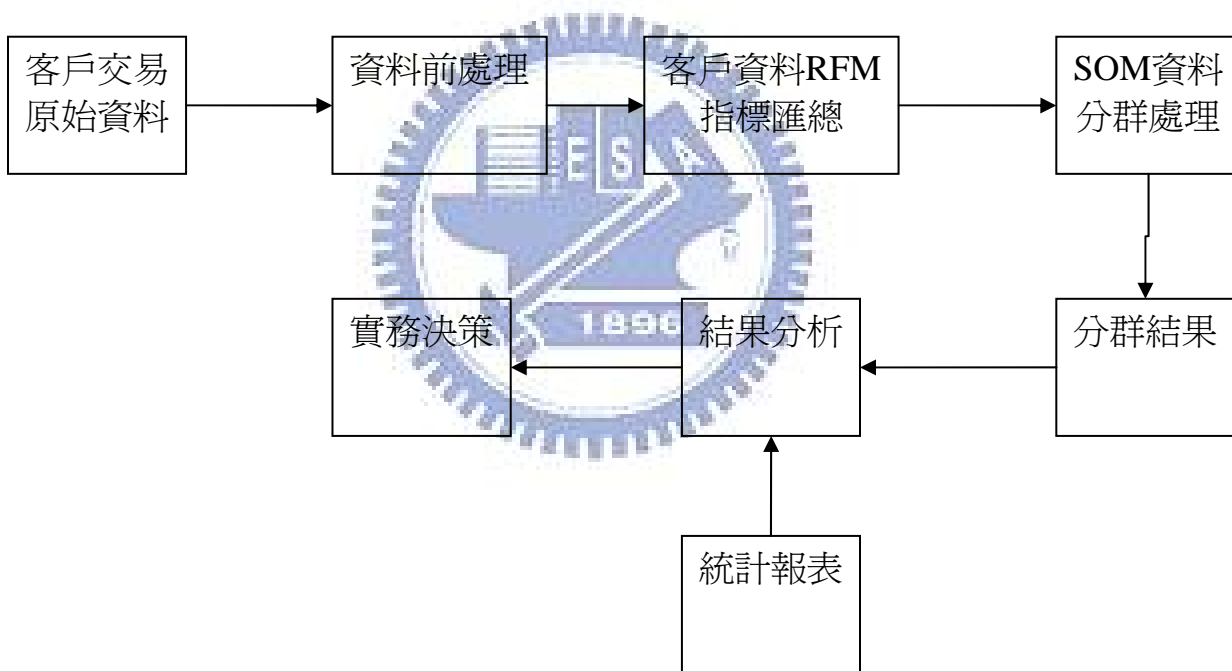
如此，我們需將最有效的行銷與市場資源投資於收益最大的客戶群中，所以如何找出最有價值的客戶群並做最有效的行銷計畫成為本論文所要解決的問題。

3.2. 研究架構

本論文提出一個系統雛形分析客戶資料，並利用 RFM(Recency, Frequency, Monetary) [Huge, 1944]，模型作為客戶分群指標，並在分群的邏輯中，應用人工智慧 Data Mining 的方法，以 SOM 演算法作為客戶分群的邏輯，並探討分群結果與實際應用的分析。

文中並以半導體產業的客戶資料實際資料模擬系統雛形，並做一結果分析研討。

歸納本文雛形架構如下：



附圖 3.2-1 分析架構

4. 改良式客戶分群系統雛形建置

在建置系統雛形上，本文以半導體公司實際的客戶交易資料作為系統雛形資料輸入，並經由 Java SOM [Tomi Suuronen, 2001] 的分群演算法，產生客戶分群結果，並以 XML 方式呈現群組結果，以方便後續的資料分析的轉檔的工作。

以下逐步介紹系統雛形的分析過程。

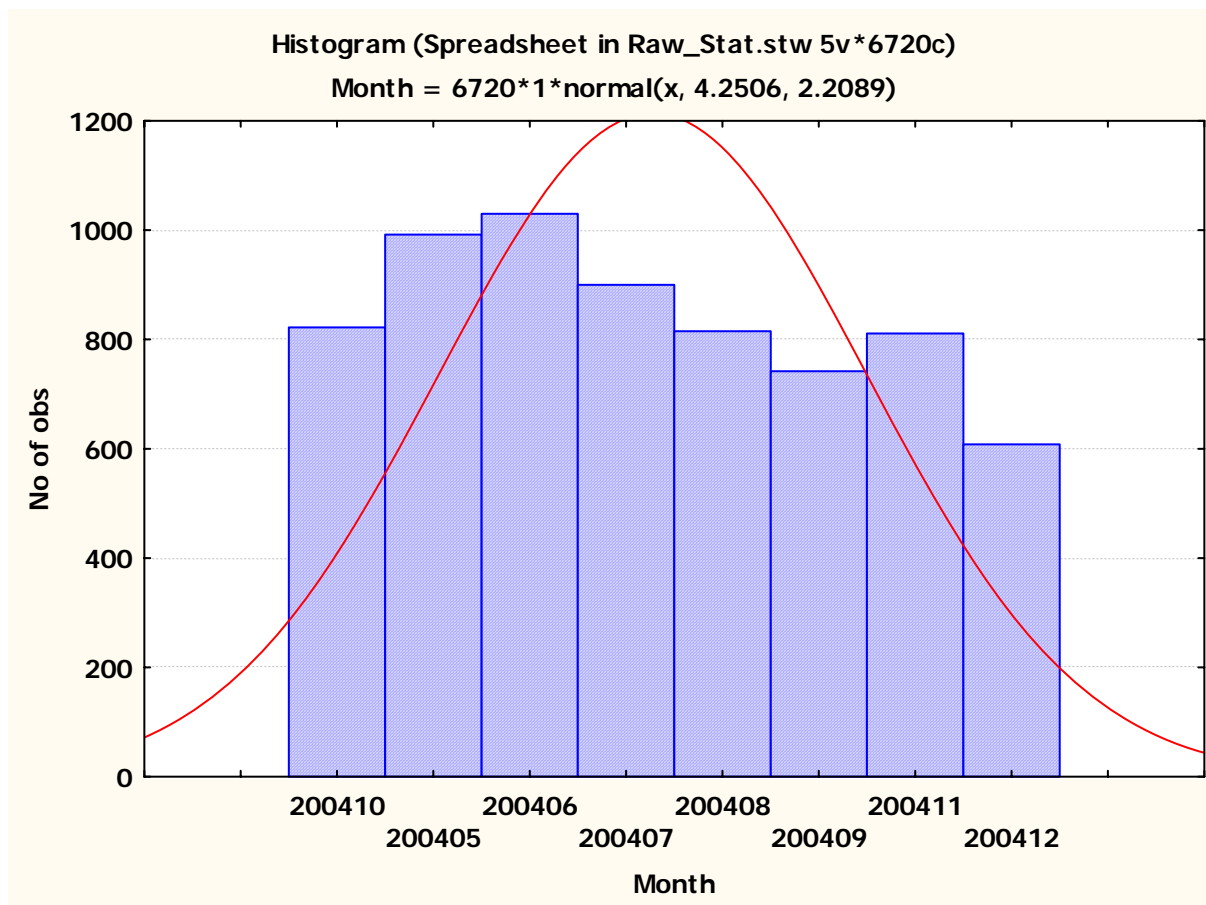
4.1. 資料收集與描述

本文所收集的客戶資料為某半導體製造公司 2004 全年的客戶交易資料，其資料共有 7987 筆的交易資料，其間包含了 586 個不同客戶，期望利用這 586 個客戶的作為分群的依據，期望找出無價值客戶、成長型客戶、主力型客戶、高價值客戶等群組，進而區分出客戶的價值指標。

原始資料以下列欄位分別收集：

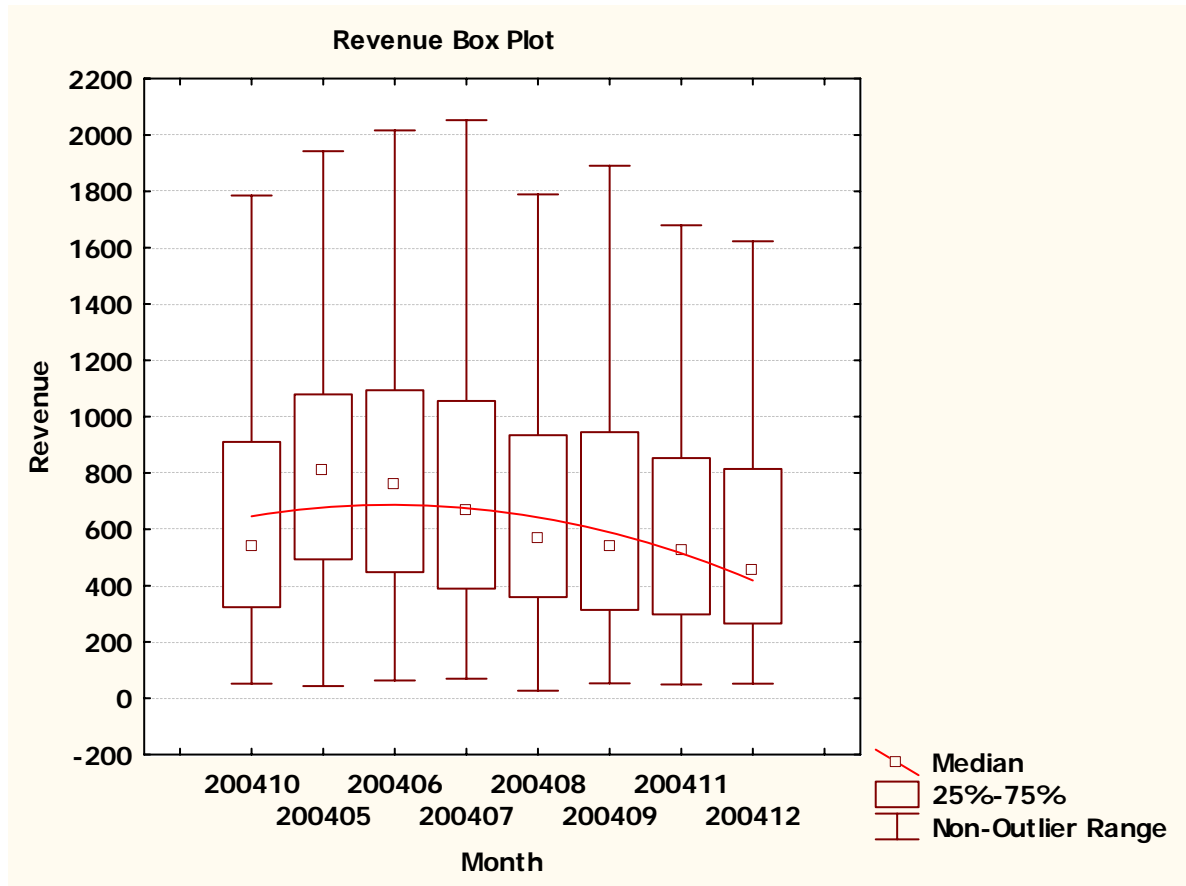
客戶代號 Customer	年份 Year	月份	產品名 Product	產品規格 Density	地區 Region	數量 Quantity	單價 ASP
------------------	---------	----	----------------	-----------------	--------------	----------------	-----------

以下就針對所收集的原始資料作敘述統計分析。



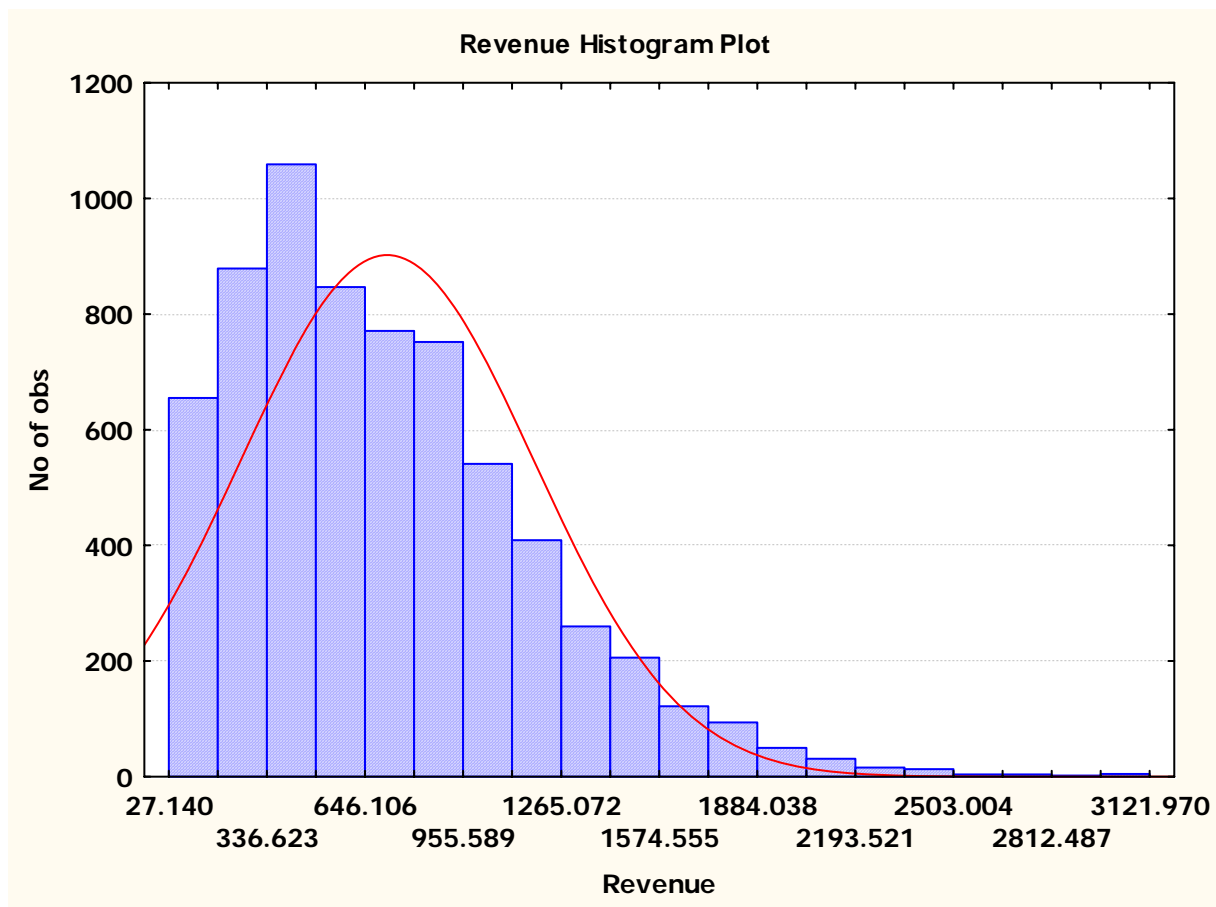
附圖 4.1-1 各月中的交易數量的 Histogram 圖

附圖 4.1-1 所代表之意義為原始資料中各月的交易資料總數，以此項原始資料當成 Recency 的指標，亦即將選定一特定時間（現選定為 2004/00 為參考月份），而相對於該交易資料的交易日期與其特並時間之時間差即當成 Recency 指標，及越接近現在交易日期，其 R 指標貢獻就越大，也符合實際狀況的模擬。



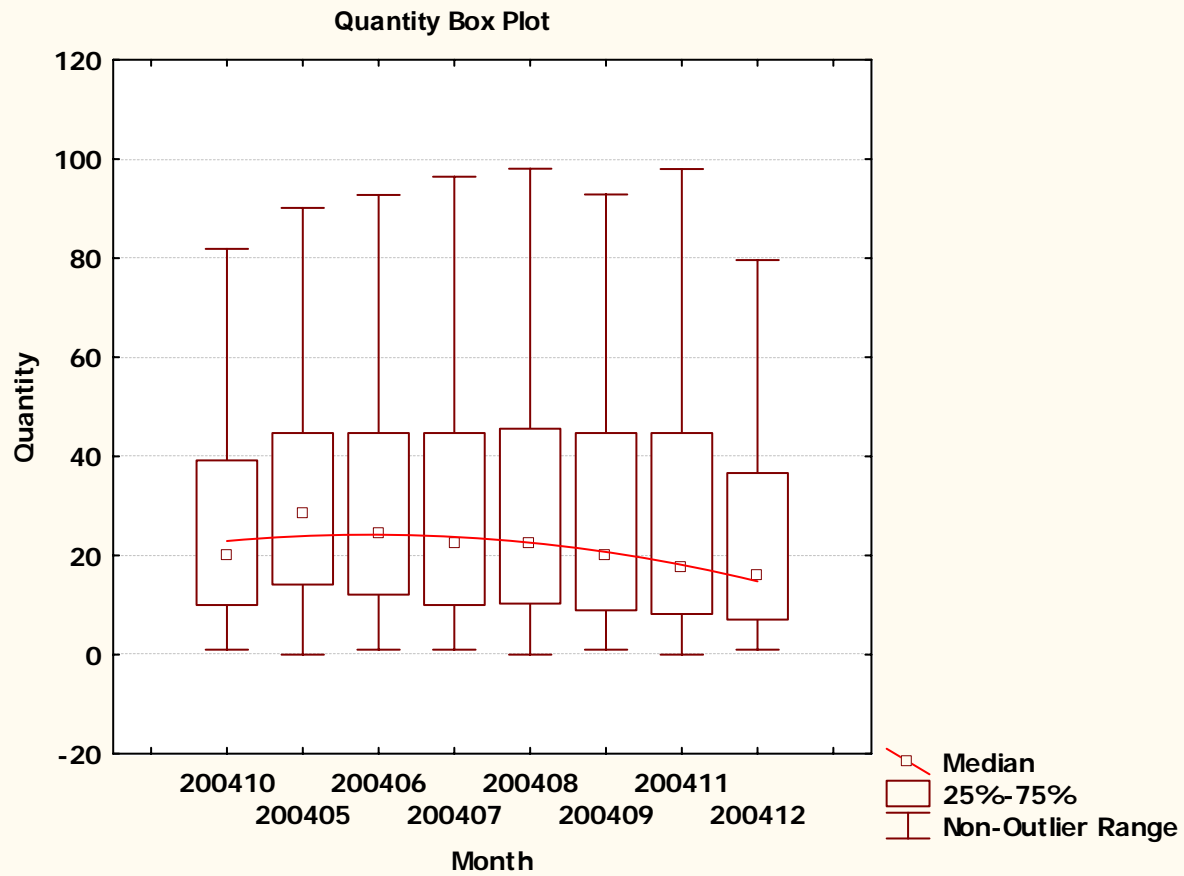
附圖 4.1-2 依月份的營收 Box Plot

附圖 4.1-2 所代表之意義為依據各月分中交易資料的營收 Box Plot。由此圖中可以各月份中的營收差異有一定的程度，且各月份均有相同的情形，亦即各代表客戶中交易的金額變化幅度相當大。

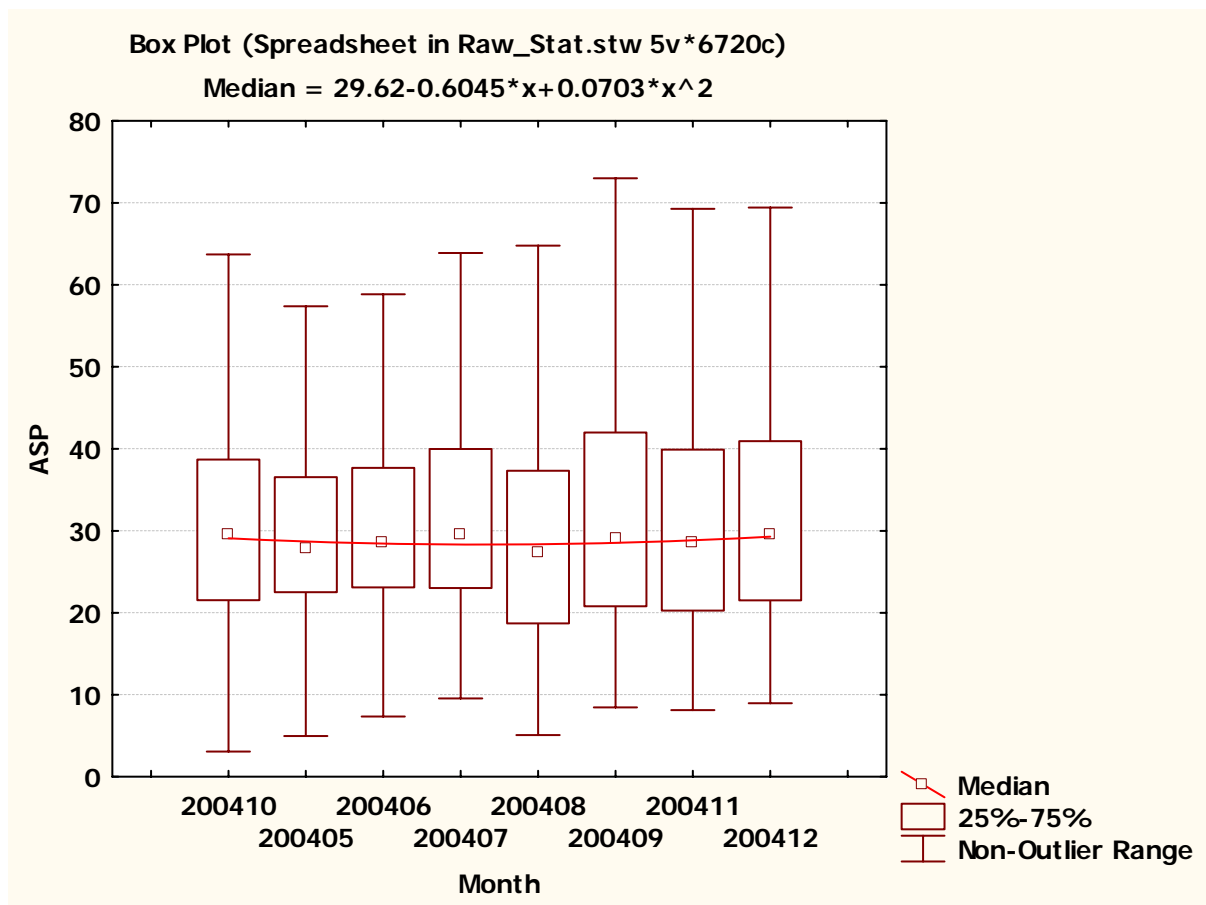


附圖 4.1-3 原始交易資料的 Revenue 之 Histogram Plot

附圖 4.1-3 所代表之意義為交易資料的營收 Histogram Plot。由此圖中可知其營收的分配並非為常態分佈，由圖形顯示其分佈情形偏左的分佈。



附圖 4.1-4 依各月份的交易數量 Box Plot



附圖 4.1-5 原始資料之 ASP Box Plot

4.2. 資料前處理過程

4.2.1. 建構 RFM 指標

經由敘述統計的基本資料分析，接下利用此原始資料，找出對應客戶資料中的 RFM 三種指標，三種指標對應於原始資料所代表意義如下：

Recency：指最近一次的交易時間。在原始資料中所呈現的即是最後一筆交易資料的時間。本文中為了量化R的指標，採用與特定時間(訂於2004年00月)與最後一次的交易的時間差作為R指標定義。

Frequency：指的是交易頻率。在原始資料中每一個客戶的一交易資料即為一

筆資料。故定義每一客戶在特定時間的交易筆數作為F的指標。

Monetary：交易金額。本文則利用每一客戶的交易資料的金額（數量*單價）的總和當成M的指標定義。

根據上述的指標定義對原始資料做處理，產生處理過的客戶 M 指標輸入檔，並將 RFM 客戶指標輸入檔當成 Java SOM 分群引擎的輸入。由於 Java SOM 所定義的輸入檔為 XML 格式，故利用上述 RFM 的指標資料轉成 Java SOM 的輸入格式。輸入格式以 DTD 方式定義，以下列格式存在（jsom.dtd）：

```
<!ELEMENT jsom (pref,data)>
<!ELEMENT pref (meta?,dimension)>
<!ELEMENT meta (author*,project?)>
<!ELEMENT author (name,organization)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT project (name,code?)>
<!ELEMENT code (#PCDATA)>
<!ELEMENT dimension (dim_type+)>
<!ELEMENT dim_type (#PCDATA)>
<!ELEMENT data (node+)>
<!ELEMENT node (dim+)>
<!ELEMENT dim (#PCDATA)>
<!ATTLIST meta code CDATA #IMPLIED>
<!ATTLIST meta date CDATA #IMPLIED>
<!ATTLIST node label CDATA #REQUIRED>
<!ATTLIST dim type CDATA #REQUIRED>
```

表格 4.2-1 JSOM 之資料格式定義

4.2.2. RFM 指標之正規化

根據原始資料建立RFM指標後，我們可以觀察到R, F, M此三個指標的數值分佈截然不同。然而，現今所用的RFM分析模型並未將RFM三個指標的權重做適當的處理[陳彤生, 2002]，以現行資料而言，Recency的Order約在兩位數，但Monetary的order約在五位數，相形之下M的指標權重即被放大約 10^3 倍，故在此將三個指標的分佈做正規化。

首先，利用列出正規化後的三個指標的基本統計量，如表

	R	F	M
Mean	186.06872	13.260513	11.430085
STD	82.127302	27.622438	24.513304
Min	0	0	0.2
Max	255	255	254
25th%	145.7	1	1
75th%	255	12.6	10.5

表格 4.2-2 RFM 三項指標經正規化後的基本統計量

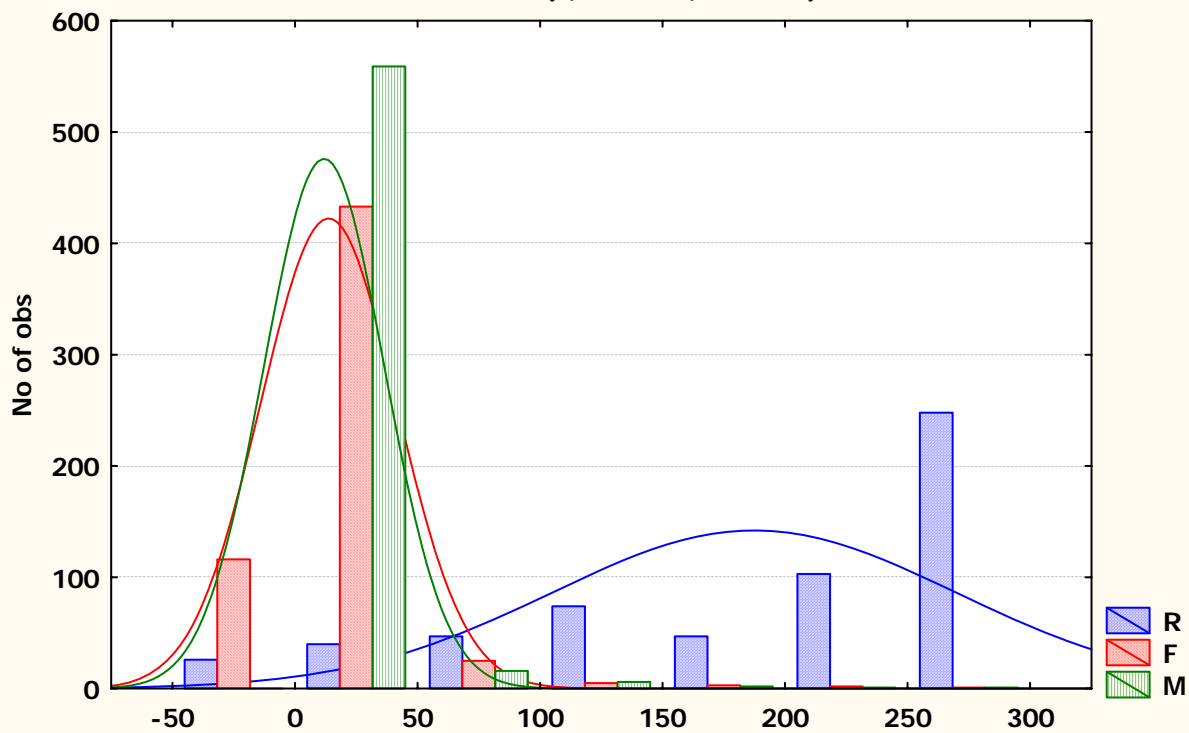
其次，將各個指標的分佈情形依比例對應到 0 ~ 255 的數值當中。圖 X 所表示的的是經正規化後 RFM 三個指標的分佈情形。

Histogram (RFMdata_normal.sta 4v*585c)

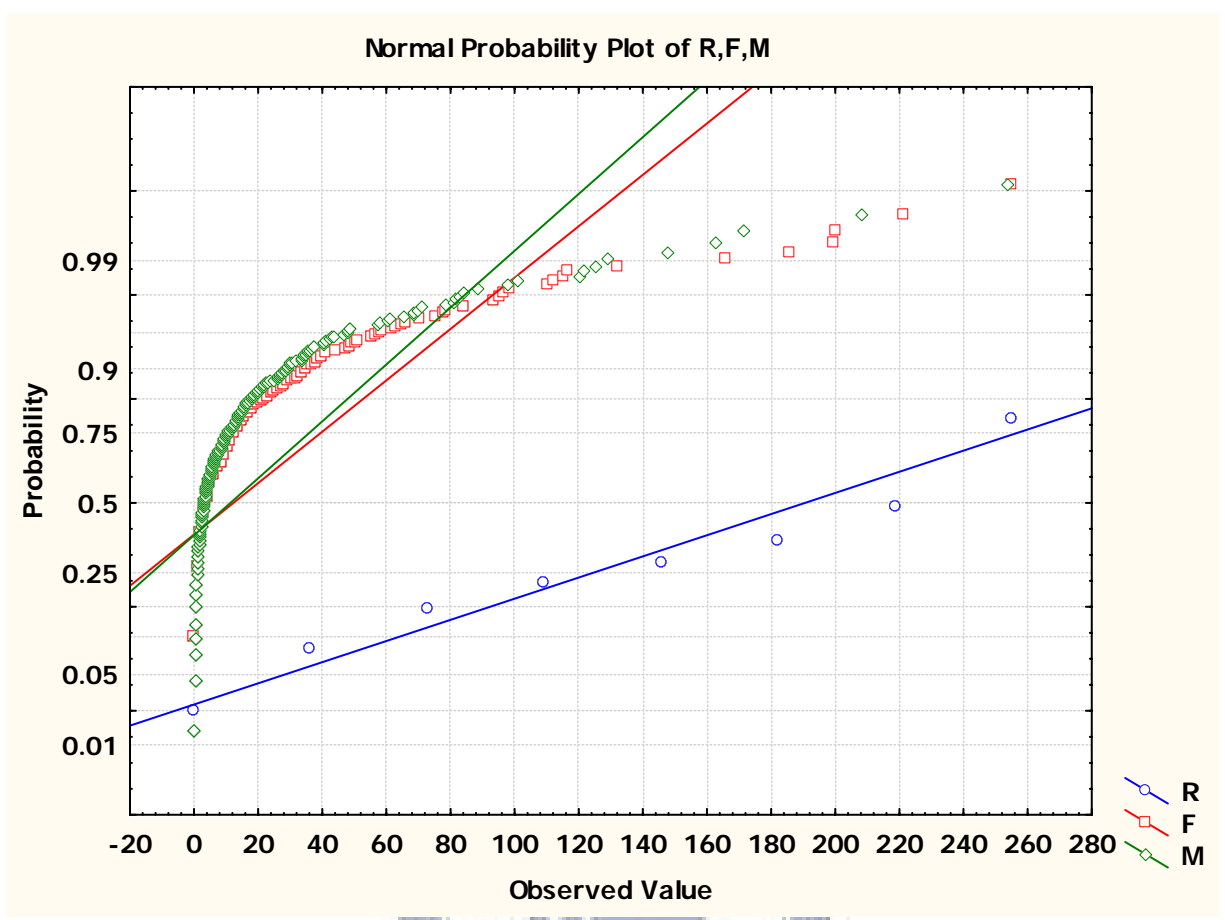
$R = 585 \cdot 50 \cdot \text{normal}(x, 186.0687, 82.1273)$

$F = 585 \cdot 50 \cdot \text{normal}(x, 13.2605, 27.6224)$

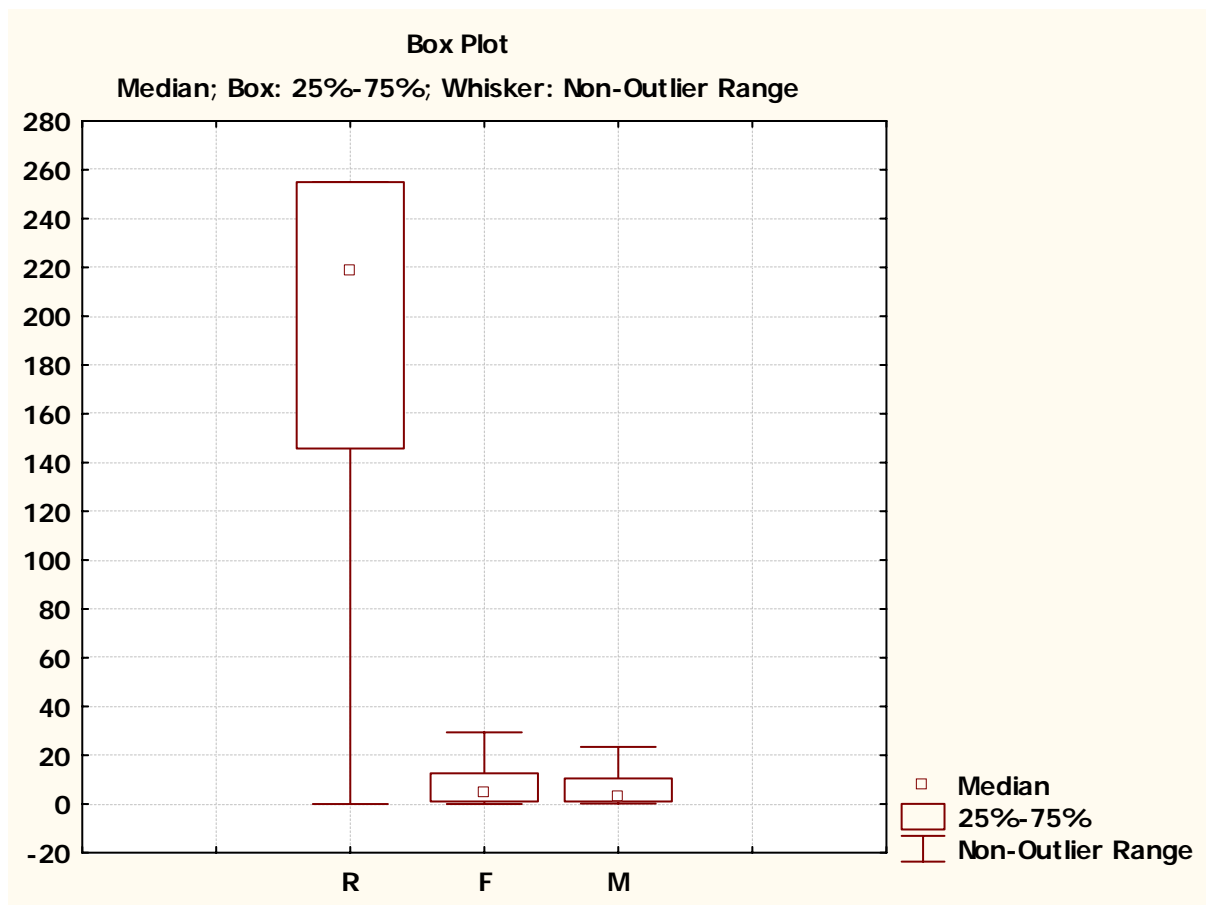
$M = 585 \cdot 50 \cdot \text{normal}(x, 11.4301, 24.5133)$



附圖 4.2-1 正規化後的RFM Histogram



附圖 4.2-2 正規化後的 Normal Probability Plot



附圖 4.2-3 正規化後之RFM Box Plot

由圖可知 Recency 指標因取樣以單月為基本，故放大了其指標的意義，而分配類似於矮胖型。Frequency 及 Monetary 則多集中於中央偏右。

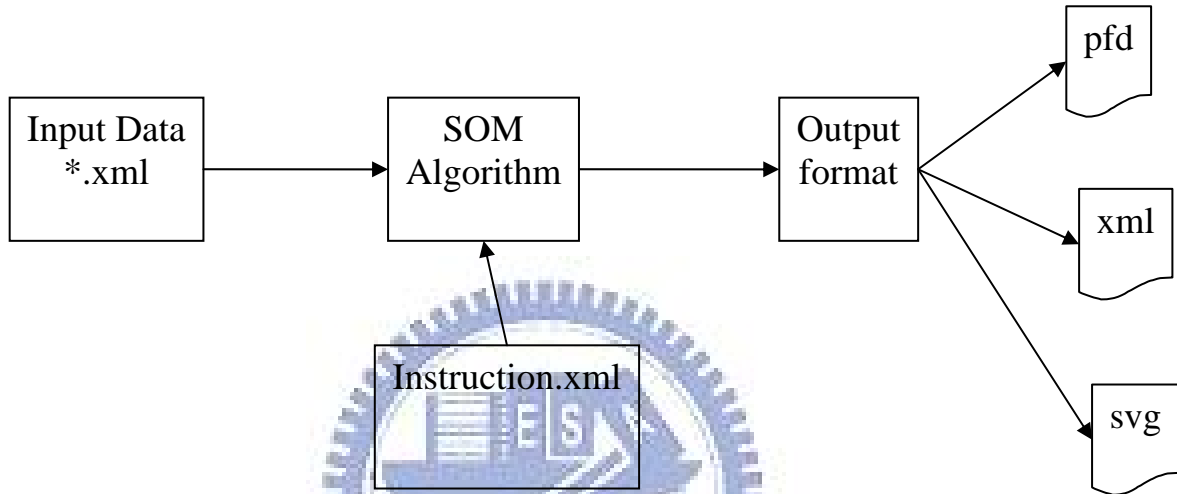
在此，資料呈現非常態的分佈。由於我墳所採取的分群模型是利用類神經網路的 SOM 模型演算法，而此演算法不侷限於常態與線性模型，故此資料也可驗證此模型演算法的強健性。

4.3. 資料分群模型

在原始RFM模型中，各指標中採取以等分的方式來做分群的方法，但此方法並不考慮到原始資料的分散程度與敏感性，導致於在實際應用於分析客戶

資料時並不充分表現出客戶的區隔性。因此，本論文以自組織神經網路 (self-organizing map, SOM) 的手法嘗試區分RFM三個指標，並將指標映射於已定義之四個象限之中，分別代表無價值客戶、成長型客戶、主力型客戶、高價值客戶，進而區分出客戶的價值指標。

附圖 4.3-1 說明了 JavaSOM 分析的系統架構。



附圖 4.3-1 JavaSOM 的系統架構

JavaSOM 一個利用 Java 所開發出來的自組織映射圖網路演算法 (SOM) 之套裝軟體，其基於 General Public License 的開放原始碼所發展(參考附錄)。JavaSOM 分成兩的部分，其一是 SOM 的核心演算法部分：javasom.jar，另一個部分為由 Apache Software Foundation 所提供的週邊程式所組成，如表格 4.3-1 所示。

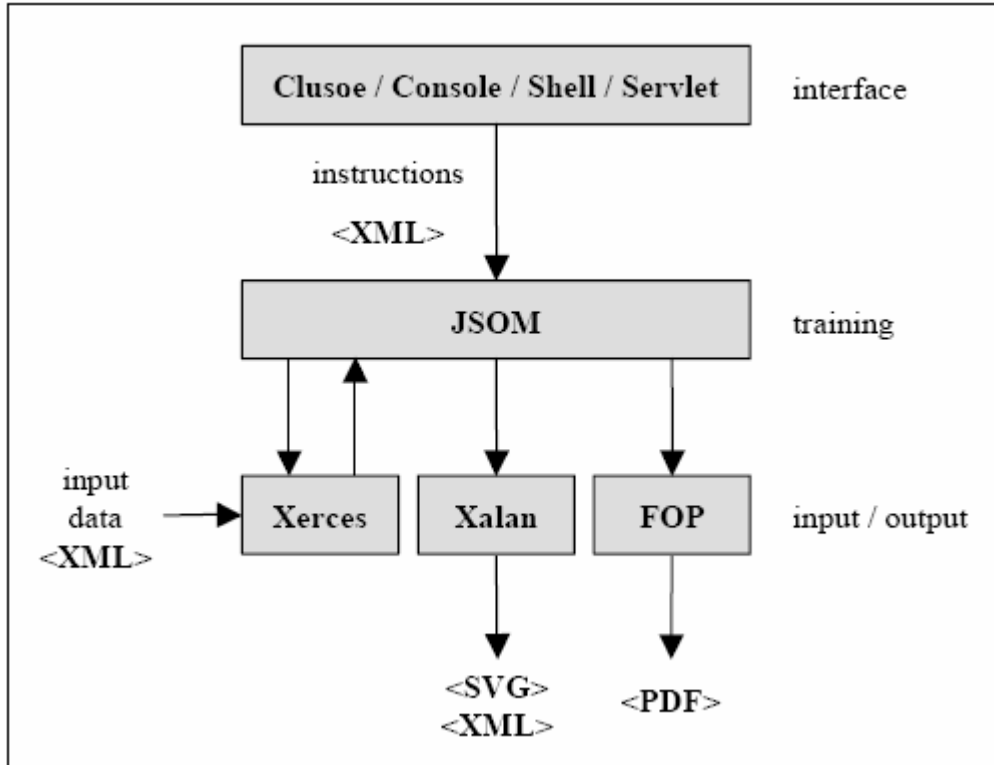
Javasom.jar	Apache Software Foundation
JSOM	Xerces
SOM 核心演算法	一種 XML 解析器
Clusoe	Xalan

使用介面	XSL 的轉譯處理器，轉換已訓練之圖形為 XML 格式
	FOP Formatting Object Processor 為利用 XSL 將資料轉換為如 PDF, PCL, PS, SVG, XML (area tree representation), Print, AWT, MIF and TXT 等格式的工具軟體

表格 4.3-1 JavaSOM 程式結構

JSOM 為 JavaSOM 中的核心演算法，並將 SOM 的演算結果利用 Xerces、Xalan、FOP 輸出為相關的輸出格式。其間的關係如附圖 4.3-2。





附圖 4.3-2 JavaSOM 程式運作架構

JSOM 所提供之輸出格式有以下三種：XML、pdf 及 SVG。XML 的檔案不具有視覺化的資料顯示功能，通常利用其儲存為運算結果，並利用其運算結果之 XML 轉換為其他視覺化的格式，如 JSOM 所提供的 SVG 及 PDF 相關格式。SVG(Scalable Vector Graphics)為 XML 語言的延伸應用，用來描述二維的向量圖形，其由 W3C 所發展並以廣為商業大廠所應用，如 Adobe、Corel 和 Macromedia。

本論文中的分群引擎利用 Open source 中的 JavaSOM，其利用 XML 的格式定義其原始資料，並經由其 SOM 演算法產生 xml，pdf，與 svg 相關格式的輸出。表格 4.3-1 列出 JavaSOM 的各種參數設定檔。

```

<instructions>
  <input>
    <file>D:\\MyTest\\javasom\\demo.xml</file>
  </input>
  <initialization>
    <normalization used="true" />
    <x.dimension>14</x.dimension>
    <y.dimension>14</y.dimension>
    <lattice type="hexagonal" />
    <neighbourhood type="gaussian" />
  </initialization>
  <training>
    <steps>1000</steps>
    <rate type="linear">0.1</rate>
    <radius>8</radius>
  </training>
  <training>
    <steps>10000</steps>
    <rate type="linear">0.02</rate>
    <radius>4</radius>
  </training>
  <output paper="a4" orientation="portrait">
    <folder>D:\\MyTest\\javasom\\output</folder>
    <identifier>animals</identifier>
    <type format="svg" />
  </output>
</instructions>

```

表格 4.3-2 JavaSOM 參數設定檔

列出以下相關參數設定：

參數名稱	參數設定
------	------

鄰近區域型態	六邊形 Hexagonal
學習速率	0.1
鄰近區域函數	高斯函數 Gaussian
學習步驟	10000
鄰近半徑	4

表格 4.3-3 JavaSOM 相關參數設定



5. 結果分析

5.1. 結果分析

經由程式分析模擬結果如下：



附圖 5.1-1 經 JSOM 所繪出網路拓樸圖 (16 * 16)

圖 5.1-1 由 16*16 個節點所構成的網路拓樸圖，由圖形可以判斷哪一些客戶可以分成一群，並群與群的相似度可以由群與群之間距離來判斷各群的差異程度。由圖 5.1-1 之中，我們可以將圖區分為四個象限，以及我們所定義的高價值客戶、主力型客戶、成長型客戶、無價值客戶等四個象限，並做進一的市場區隔分析。

將客戶資料分為四群的結果可以如表格 5.1-1 顯示。

Group 1 高價值客戶 (350)	003, 004, 005, 007, 008, 012, 013, 016, 018, 020, 022, 023, 024, 025, 027, 028, 029, 032, 035, 038, 040, 041, 045, 047, 052, 053, 054, 055, 056, 057,

059, 061, 063, 064, 065, 066, 067, 068, 069, 070,
074, 077, 078, 082, 084, 086, 088, 089, 090, 091,
092, 093, 095, 097, 098, 099, 101, 103, 104, 106,
107, 108, 109, 111, 113, 115, 116, 117, 118, 119,
120, 121, 122, 125, 126, 127, 130, 131, 132, 133,
134, 136, 141, 142, 145, 146, 147, 148, 149, 151,
152, 155, 156, 157, 158, 159, 160, 161, 162, 164,
165, 167, 168, 170, 171, 172, 173, 175, 177, 178,
181, 182, 183, 184, 185, 188, 189, 190, 191, 192,
194, 196, 197, 198, 199, 201, 203, 204, 206, 208,
210, 211, 216, 218, 222, 223, 224, 229, 232, 234,
237, 238, 240, 242, 243, 244, 247, 248, 249, 251,
252, 253, 254, 255, 256, 260, 261, 262, 263, 266,
269, 273, 274, 275, 278, 280, 281, 282, 284, 286,
288, 289, 292, 294, 295, 296, 297, 299, 300, 302,
303, 305, 308, 310, 312, 313, 317, 318, 319, 321,
322, 324, 325, 327, 328, 329, 330, 331, 332, 333,
334, 338, 339, 340, 341, 342, 344, 345, 348, 352,
353, 355, 357, 359, 360, 361, 362, 363, 364, 367,
368, 369, 370, 371, 372, 374, 377, 379, 380, 381,
383, 384, 385, 389, 393, 394, 395, 397, 398, 399,
400, 401, 403, 404, 406, 408, 409, 412, 414, 415,
417, 418, 419, 420, 422, 423, 424, 425, 427, 428,

	429, 435, 436, 437, 438, 440, 441, 442, 444, 446, 448, 450, 452, 454, 457, 459, 460, 462, 463, 465, 468, 473, 474, 475, 476, 478, 479, 480, 482, 483, 484, 487, 488, 489, 491, 492, 494, 495, 496, 497, 498, 499, 500, 501, 502, 504, 505, 506, 508, 511, 512, 513, 514, 515, 516, 521, 522, 523, 525, 527, 529, 533, 537, 538, 539, 541, 546, 547, 549, 551, 554, 555, 556, 557, 558, 559, 560, 562, 565, 566, 569, 570, 572, 573, 575, 577, 579, 581, 583, 585
Group 2 主力型客戶 (5)	202, 226, 293, 375, 528
Group 3 無價值客戶 (42)	002, 034, 043, 048, 050, 051, 058, 076, 085, 137, 140, 150, 166, 186, 212, 213, 221, 227, 228, 241, 245, 250, 276, 343, 351, 378, 388, 405, 407, 445, 461, 486, 493, 510, 520, 526, 530, 531, 563, 564, 568, 576
Group 4 成長型客戶 (66)	001, 006, 009, 010, 011, 014, 015, 017, 019, 021, 026, 030, 031, 033, 036, 037, 039, 042, 044, 046, 049, 060, 062, 071, 072, 073, 075, 079, 080, 081, 083, 087, 094, 096, 100, 102, 105, 110, 112, 114, 123, 124, 128, 129, 135, 138, 139, 143, 144, 153, 154, 163, 169, 174, 176, 179, 180, 187, 193, 195,

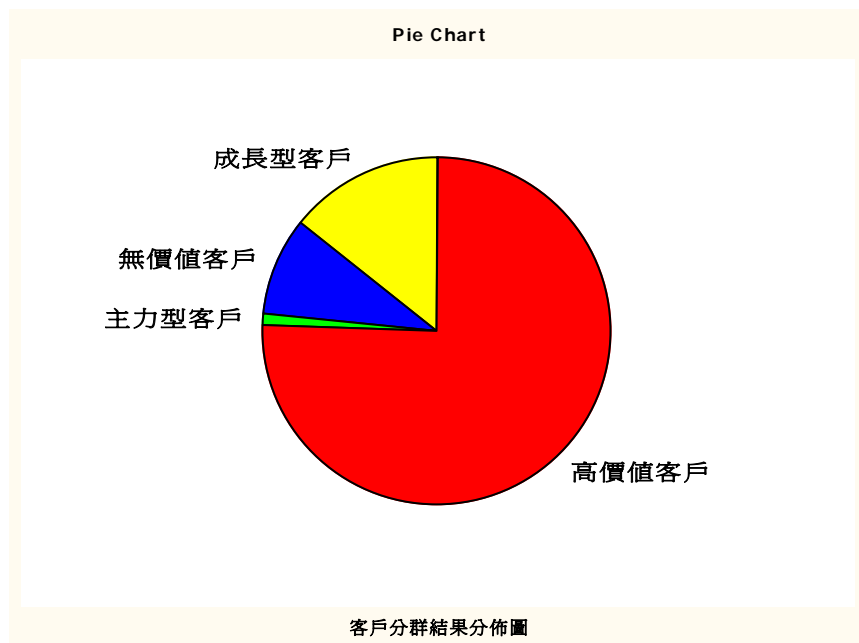
<p>200, 205, 207, 209, 214, 215, 217, 219, 220, 225, 230, 231, 233, 235, 236, 239, 246, 257, 258, 259, 264, 265, 267, 268, 270, 271, 272, 277, 279, 283, 285, 287, 290, 291, 298, 301, 304, 306, 307, 309, 311, 314, 315, 316, 320, 323, 326, 335, 336, 337, 346, 347, 349, 350, 354, 356, 358, 365, 366, 373, 376, 382, 386, 387, 390, 391, 392, 396, 402, 410, 411, 413, 416, 421, 426, 430, 431, 432, 433, 434, 439, 443, 447, 449, 451, 453, 455, 456, 458, 464, 466, 467, 469, 470, 471, 472, 477, 481, 485, 490, 503, 507, 509, 517, 518, 519, 524, 532, 534, 535, 536, 540, 542, 543, 544, 545, 548, 550, 552, 553, 561, 567, 571, 574, 578, 580, 582, 584</p>

表格 5.1-1 客戶分群結果

由表格 5.1-1 的分群結果，可以將所分析的結果映射至我們所定義的四群當中，而各群的客戶數目分別如下：

- 高價值客戶：客戶編號 003…等 350 個。
- 主力型客戶：客戶編號 202…等 5 個。
- 無價值客戶：客戶編號 002…等 42 個。
- 成長型客戶：客戶編號 001…等 66 個。

而將此四類型客戶的分佈圖繪於附圖 4.3-3 中。



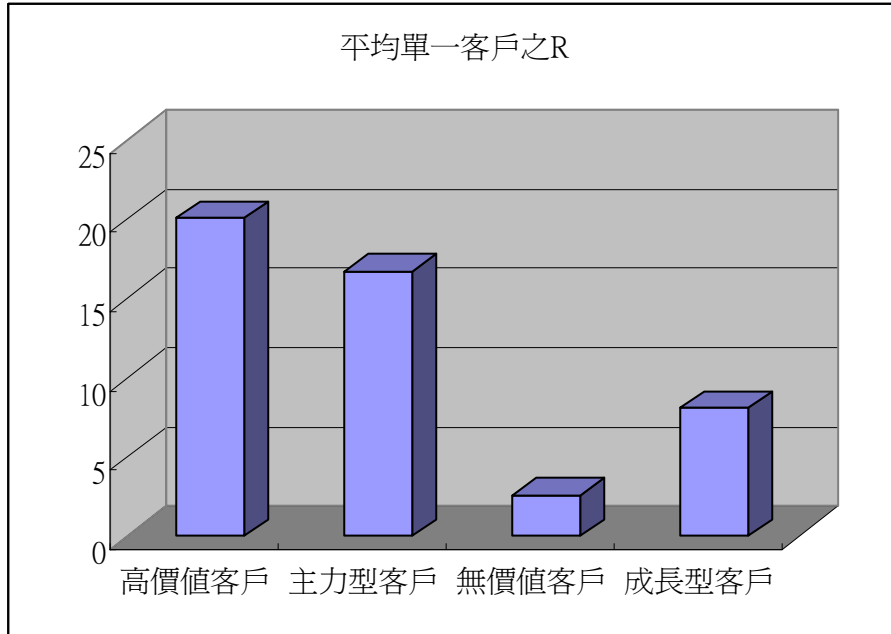
附圖 5.1-2 客戶分群結果分佈圖

首先，由已分群的結果觀察各 RFM 指標的效益貢獻。

將各群的 RFM 指標分別加總，並求出各群中單一客戶的平均 RFM 指標，並比較各群中 RFM 指標效益。分別依據 R、F、M 列出各群的指標總和與單一客戶的平均值。表格 5.1-2 列出分群後 R 的指標分佈情形。附圖 5-1.3 則表示分群後平均單一客戶的 R 指標。

各群總和	Recency 總和	平均單一客戶的 R
高價值客戶	7028.40	20.08
主力型客戶	83.80	16.76
無價值客戶	110.70	2.64
成長型客戶	534.50	8.10

表格 5.1-2 分群後 R 指標

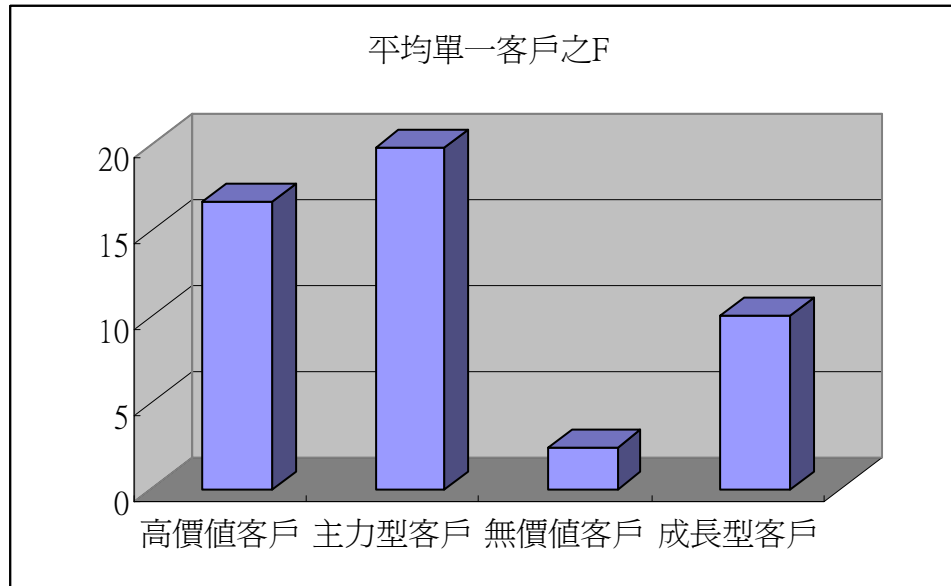


附圖 5.1-3 平均單一客戶之R

其次，再來觀察 F 指標在各群中的分佈情形。表格 5.1-3 列出分群後 F 的指標分佈情形。附圖 5-1.4 則表示分群後平均單一客戶的 F 指標。

各群總和	Frequency 總和	平均單一客戶的 F
高價值客戶	5822.10	16.63
主力型客戶	99.20	19.84
無價值客戶	99.20	2.36
成長型客戶	666.10	10.09

表格 5.1-3 分群後 F 指標

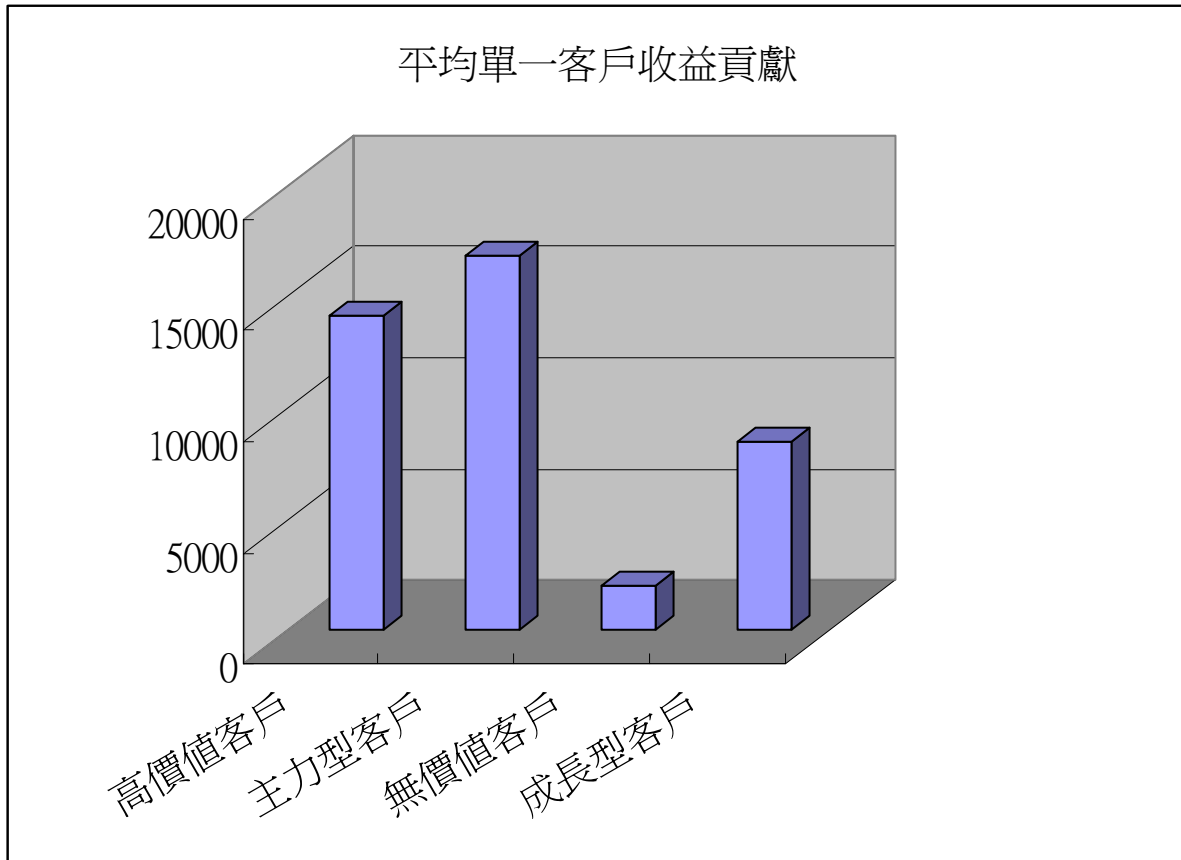


附圖 5.1-4 平均單一客戶之F

接下來，針對M指標來歸納分析比較各群對企業獲利的貢獻程度。將此四群所對應客戶資料對應回原始交易資料，將各群交易資料作加總分析，可得表 5.1-4：

群別	收益貢獻	平均單一客戶收益貢獻
高價值客戶(350)	NT\$4,927,862.79	NT\$14,079.60797
主力型客戶(5)	NT\$84,047.22	NT\$16,809.444
無價值客戶(42)	NT\$81,438.10	NT\$1,939.002381
成長型客戶(66)	NT\$553,809.65	NT\$8,391.055303

表格 5.1-4 群別與收益貢獻表



附圖 5.1-5 各群平均單一客戶的收益貢獻

由附圖 5.1-6 可之各群對應 M 指標的貢獻情形。由此圖可以發現高價值客戶與主力型客戶相較於其他兩群的平均客戶收益有極大差別，其中高價值客戶群的平均收益貢獻是無價值客戶群的 7.26 倍。

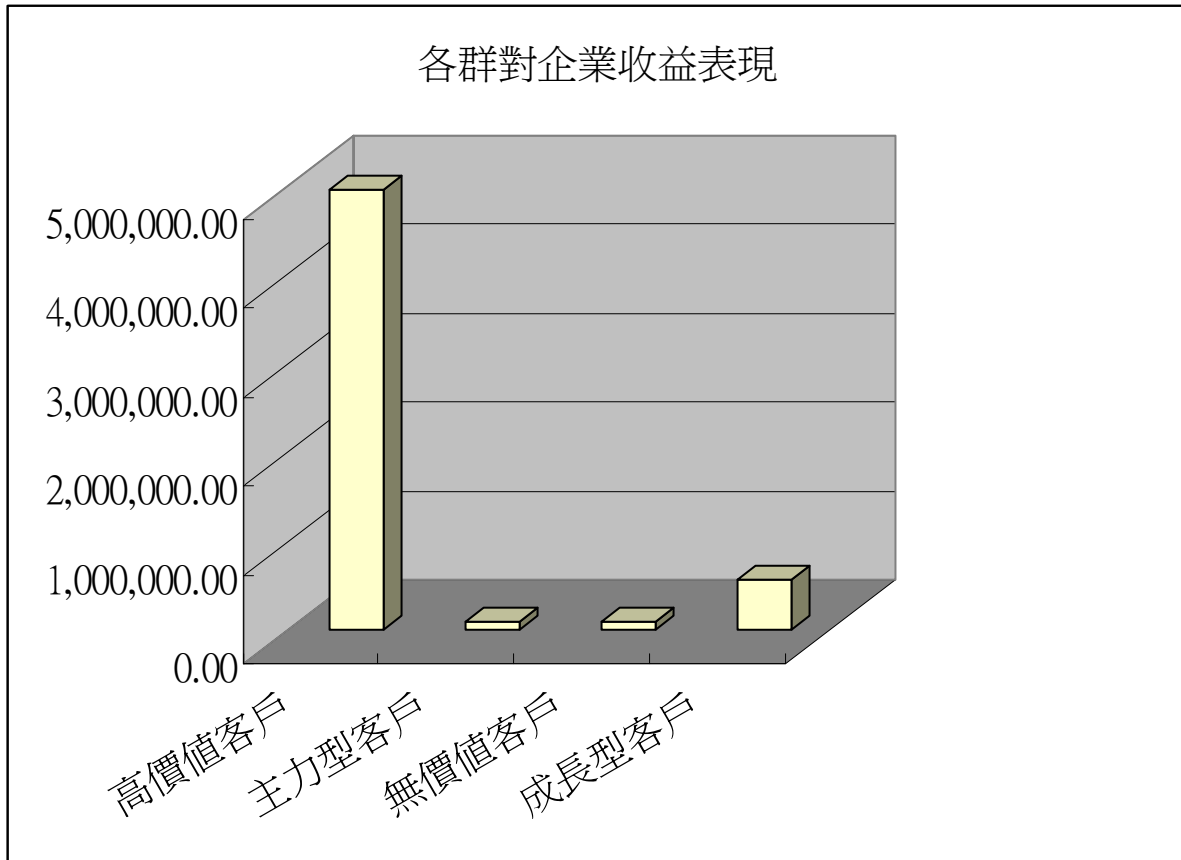
我們由此可以歸納各群的 RFM 指標的特性：

- 高價值客戶：R、F 指標普遍均較其餘三群高，但 M 指標(意指平均單一客戶的購買金額)不見得比其他指標為高。其在市場行銷上的也有其實際意義。若是一個客戶被定義成高價值客戶，其交易次數與頻率必定相對高，而企業對這樣的客戶的單筆交易金額(即此所指的 M 指標)也有相對的折扣，以爭取對客戶的忠誠度。

- 主力型客戶：M 指標普遍均較其餘三群高，即代表此群的客戶不見得交易次數如高價值客戶頻繁，但每一筆交易所帶來的收益均對企業有顯著的幫助。
- 無價值客戶：R、F、M 三者指標均相對低。對於企業來說是明顯可以忽略的一群。
- 成長型客戶：在此群客戶中，R、F、M 並不如高價值客戶與主力型客戶對企業具有顯著性的影響，但尤其是 F 及 M 指標，相較於高價值客戶與主力型客戶之下，即使現階段仍無法看出對企業的貢獻影響程度，但在長期的規劃上仍然對企業的貢獻有幫助。

以上是針對 RFM 三種指標經由 SOM 分群結果做的分析，現我們將做一匯總，即利用分群的結果，將原始中的各交易次數與金額做一加總，實際反應出各群對企業的貢獻程度比較。





附圖 5.1-6 各群對企業收益表現比較

由附圖 5.1-5 可知，高價值客戶群中對的企業收益的表現遠大於其餘三群客戶的收益表現總和，由於此表乃是針對交易資料來加總，故所代表之意義為各群實際為企業的貢獻程度。由此表可知，高價值客戶所帶來的貢獻大於其餘三群貢獻的總和，而可推論只要針對高價值客戶群投入最大的心力，即可相對貢獻於企業的收益當中。然而對其餘三群的客戶，可採取不同的客戶管理策略。

有了此四群的客戶資料，即可針對此四群的客戶採取不同的市場策略。如高價值客戶即是企業長時間的獲利來源，企業必須盡全力來保持這些客戶

的滿意度與忠誠度。主力型客戶所扮演的角色是交易頻繁程度及交易次數可能不及高價值客戶所給企業帶來的利益，但可能在單一指標上，如交易金額有相當可觀的利益，企業也必須對此類型的客戶加以提升其客戶滿意度，以期望在往後能給企業帶來更多的利益。無價值客戶所代表的的是無論從哪一個指標來看，此類型的客戶並無法帶給企業顯著的利益，故此類型在行銷策略上是採取放縱的策略，並不特意經營此類型各戶。對於成長型客戶而言，此類型的客戶對企業在短期無法為企業帶來顯著的利益，但相對具有潛力能提供企業未來利潤，可以說是企業值得培養的對象，相對而此類型的客戶，建議企業可以採取行有餘力才關注的可客戶群，可先針對高價值客戶及主力型客戶投入較大的心力與資源。

5.2. 行銷決策支援分析

分析企業客戶的類型，會發現有許多的客戶對公司來說是沒有利潤的，這些客戶應該要逐步的捨棄，但是有幾類的客戶，我們會把它們保留下來：

- 參考指標客戶
- 推薦型客戶
- 可學習的客戶

有些客戶可以提高我們的聲望，它們通常是深具影響力的大型客戶，這便是所謂的參考指標客戶，像是供貨給 Wal-Mart 或是 HP，雖然賺不到什麼錢，但是其它的客戶聽到我們跟這些公司合作，會覺得我們公司是值得信賴的公司，因此可以帶來更多的客戶，進而帶來更多的利潤，這類型的客戶就像是我們的廣告招牌，不可以輕易割捨。某些客戶會幫我們介紹新的客戶，或是在其它場合推薦本公司，我們稱它們為推薦型客戶。和它們交易可能不賺錢，

但是卻可能是最有產值的不支薪業務。

某些客戶擁有了值得學習的營運機制，或是它們在創新上經常有嶄新的突破，跟它們做生意，就像是聘顧了一位免費經營顧問，從它們身上我們不止可以學習到科技的創新、生產程序上的創新，更重要的是可以學到管理方法上的創新，以及行銷服務上的創新，有了這樣的良師益友，可以強健公司的體質，讓我們在未來可以擁有更強的競爭力，當然不可以輕易捨棄。



6. 結論與建議

6.1. 研究結論

此客戶分群的模型中，應用了類神經網路的自組織特徵映射圖網路 (self-organizing map network, SOM) 作為客戶分群的演算法，並結合 RFM 指標來評估客戶對企業的貢獻程度，並將分群結果進一步做客戶管理的決策資訊。

由以上的分析結果，可以推論此分群方法可以獲得相當良好的客戶分群結果，即使我們於資料前處理時所發現客戶 RFM 之指標分佈並非常態分佈，也由此可驗證 SOM 方法所具有的資料分析強健性，並不被限制於資料的常態性。

另外我們只針對這些客戶分成無價值客戶、成長型客戶、主力型客戶、高價值客戶四群，當然 SOM 的分群能力可將客戶分成多群，但以現有應用反而增加爾後分析的負擔，且分成四群對於決策者能做更有效的判斷。

6.2. 研究限制與建議

本文中所提出的客戶分群模型固然具有強健與效率性，但基於取樣資料的限制，在往後的研究中仍可朝以下方向繼續深究：

- RFM 分群指標的強化

在原始的 RFM 指標中，其三個指標並無權重的關係，亦即何者的指標在特定的產業當中有較重要的地位。舉例來說，以直銷業來說，每次交易

的金額可能差異均不大（因著產品的價格維持在一個水準），但此時分別客戶重要程度的應是購買頻率（Frequency）較其餘兩指標更為重要。是否存在對於不同產業的特性，因而有不同的 RFM 權重的研究也相對重要。[2002, 陳彤生]在他所提的改良式 RFM 模型中利用決策樹歸納法加權 RFM 模式、吉尼指標法加權 RFM 模式來評估顧客效益，並結合 SOM 是一個未來的研究方向。

- SOM 分群模型的強化

SOM 分群演算法雖然具有強健性，且不用理會原始的資料分佈情形（如在本文中所舉的例子即為偏左的分佈），且輸出結果可有多樣性（XML、SVG、pdf 等），但若客戶的交易中相似性（Similiarity）過高，則易導致分群的結果中同一群的客戶數目過高（在此例中，為了區分出無價值客戶、成長型客戶、主力型客戶、高價值客戶等四群客戶分類，故指定 SOM 產生的群數為四，且此例的結果產生了高價值客戶的數目為 350 個、主力型客戶共 5 個、無價值客戶共 42 個、成長型客戶共 66 個）。而在此分析結果中客戶的數目過多反而造成市場決策的困難（依此例而言，企業反而投入大部分的精力與時間在高價值客戶上，而無法區別出就有影響力的 20% 的客戶群）。因此如何在特定群中決定出適當的客戶數是一個研究課題，並期望對應於 Preto 所提的 80/20 法則，提供企業一個利用最少的資源投入於對企業最有幫助的客戶群。

7. 參考文獻

- [1] Dittenbach, M., Rauber, A. & Merkl, D. (2002). Uncovering hierarchical structure in data using the growing hierarchical self-organizing map, *Neurocomputing*, 48(1-4), pp199-216.
- [2] Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster analysis*. 4th ed., London: Edward Arnold.
- [3] Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17(2/3), pp107-145.
- [4] Jain, A. K. & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall advanced reference series. Upper Saddle River, NJ: Prentice-Hall.
- [5] Jain, A. K., Murty, M. N., & Flynn P. J. (1999). Data clustering: a review. *ACM Computing Surveys*, 31(3), pp264-323.
- [6] Robert Zeithammer(2004), "Customer analysis: Segmentation", *BUS 37000: Marketing Strategy*.
- [7] Hughes, Arthur M. (1994) , *Strategic Database Marketing*, Chicago : Probus Publishing.
- [8] Stone, Bob (1995) , *Successful Direct Marketing Methods*, Lincolnwood, IL : NTC Business Books, pp37-59.
- [9] Ron Kohavi, Rajesh Parekh (2004), "Visualizing RFM

- Segmentation", *Proceedings of the 2004 SIAM International Conference on Data Mining (SDM'2004)*, Orlando, FL, pp391–399.
- [10] Vesanto, J. (2000). Using SOM in data mining, Licentiate's thesis, Helsinki University of Technology, Finland.
- [11] Vesanto J. & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3), pp586–600.
- [12] Z.-P. Lo, M. Fujita and B. Bavarian (1991): Analysis of neighborhood interaction in Kohonen neural networks. 6th International Parallel Processing Symposium Proceedings, pp 247–249.
- [13] 李章偉 (2000), 「資料庫行銷之顧客價值分析：以 3C 流通業為例」, 國立臺灣大學國際企業學系研究所碩士論文。
- [14] 謝依真 (2000), 「不同分群方法與不同資料來源之比較」, 國立成功大學工業管理學系研究所碩士論文。
- [15] 劉世琪 (2002), 「應用資料挖掘探討顧客價值：以汽車維修業為例」, 朝陽科技大學工業工程與管理系研究所碩士論文。
- [16] 連惟謙 (2003), 「應用資料分析技術進行顧客流失與顧客價值之研究」, 中原大學資訊管理研究所碩士論文。
- [17] 趙景明 (2003), 「應用資料探勘技術於顧客價值分析之研究」, 東吳大學資訊科學系碩士論文。
- [18] 陳彤生(2002), 「運用改良 RFM 提升行銷效益的實證研究」, 朝陽科技大學主辦第七屆人工智慧與應用研討會。
- [19] 林慧晶(1997), 「資料庫行銷之客戶價值分析與行銷策略應用」,

國立臺灣大學國際企業學系研究所碩士論文。

- [20] 楊東昌(2004)，「自組織映射圖神經網路改善模式與分群應用之回顧研究」，華梵大學工業管理學系碩士論文。



8. 附錄

附錄一 GNU GENERAL PUBLIC LICENSE 通用公共許可證

Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc.
51 Franklin Street, Fifth Floor, Boston, MA 02110-1301, USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Lesser General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by

someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.

TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The "Program", below, refers to any such program or work, and a "work based on the Program" means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term "modification".) Each licensee is addressed as "you".

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

- a)** You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.
- b)** You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a

whole at no charge to all third parties under the terms of this License.

c) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

a) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

b) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the

scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy, distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM

PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

END OF TERMS AND CONDITIONS

[How to Apply These Terms to Your New Programs](#)

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

one line to give the program's name and an idea of what it does.

Copyright (C) yyyy name of author

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301, USA.

Also add information on how to contact you by electronic and paper mail.

If the program is interactive, make it output a short notice like this when it starts in an interactive mode:

```
Gnomovision version 69, Copyright (C) year name of author
Gnomovision comes with ABSOLUTELY NO WARRANTY; for details
type `show w'. This is free software, and you are welcome
to redistribute it under certain conditions; type `show c'
for details.
```

The hypothetical commands ``show w'` and ``show c'` should show the appropriate parts of the General Public License. Of course, the commands you use may be called something other than ``show w'` and ``show c'`; they could even be mouse-clicks or menu items--whatever suits your program.

You should also get your employer (if you work as a programmer) or your school, if any, to sign a "copyright disclaimer" for the program, if necessary. Here is a sample; alter the names:

```
Yoyodyne, Inc., hereby disclaims all copyright
interest in the program `Gnomovision'
(which makes passes at compilers) written
by James Hacker.
```

```
signature of Ty Coon, 1 April 1989
Ty Coon, President of Vice
```

This General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the [GNU Lesser General Public License](#) instead of this License.