

蘇中信、劉俞志、劉蕙(2013),『以顧客價值為基礎之資料庫行銷架構』,
資訊管理學報,第二十卷,第三期,頁341-366。

以顧客價值為基礎之資料庫行銷架構

蘇中信

高苑科技大學企業管理系

劉俞志*

元智大學資訊管理系

劉 蕙

元智大學資訊管理系

摘要

企業的成功關鍵之一在於瞭解顧客的需求並迅速回應。資料庫行銷旨在利用資料庫探究顧客消費行為與了解銷售狀況,據此制定行銷策略以回應顧客需求。雖然資料庫行銷已廣被探討,然而,少有文獻提出一個完整與簡單的資料庫行銷架構。本研究提出一個結合資料探勘技術,並以顧客價值分群為基礎之資料庫行銷架構,此架構利用群集分析與序列型樣,擴充顧客價值計算模式,並依據顧客價值進行顧客分群,同時依交易資料進行產品分類,再整合分群與產品分類結果做為策略擬定依據。本研究以某公司之業務與資料為範例,依據顧客消費模式與分群結果提出可行之策略方案。並且,也提供以交易金額驗證策略方案有效性的方法,輔以實際案例之實驗數據,作為資料庫行銷設計與實作之參考。

關鍵詞：資料庫行銷、顧客價值、顧客分群、資料探勘

* 本文通訊作者。電子郵件信箱:imyuchih@saturn.yzu.edu.tw

2012/09/12 投稿;2012/11/28 第一次修訂;2013/02/28 第二次修訂;2013/04/15 第三次修訂;2013/06/25 接受

Su, C., Liu, J.Y.C. and Liu, H. (2013), 'A Customer Value Based Framework for Database Marketing', *Journal of Information Management*, Vol. 20, No. 3, pp. 341-366.

A Customer Value Based Framework for Database Marketing

Chung Su

Department of Business Administration, Kao Yuan University

Julie Yu-Chih Liu*

Department of Information Management, Yuan Ze University

Hui Liu

Department of Information Management, Yuan Ze University

Abstract

One of business success factors is to understand customers' needs and response their demands in time. Database marketing is to analyze customer consumption behavior and transaction data in databases, and to make marketing strategy according to the results. Although database marketing has been well discussed, very few studies provide a complete framework for it. This research provides an entire framework for database marketing, which employs data mining techniques to perform data clustering and analysis. In this framework, customer value model is extended. Customer segmentation is performed based on the extended model. The products are categorized by sequential pattern analysis of customers' transaction. Marketing strategies are decided according to the segmentation and the categorization. This work adopts a company as a case to demonstrate the framework. In addition to using customer value and discount as the parameters of customer segmentation, it provides feasible strategy to the segmentation results based on the customer consumption model. We also suggested the experiment for examining the effectiveness of the proposed strategy. The results of this work contribute to the design and practice of database marketing.

Keywords: Database marketing, Customer value, Customer segmentation, Data mining

* Corresponding author. Email: imyuchih@saturn.yzu.edu.tw

2012/09/12 received; 2012/11/28 1st revised; 2013/02/28 2nd revised; 2013/04/15 3rd revised; 2013/06/25 accepted

壹、緒論

網際網路的發達使企業的競爭由地區性轉為全球性，商品的訊息透過網路無遠弗屆，便捷的宅配打破產品購買的地域藩籬，使得消費者的接觸點和選擇性增加，企業面臨的競爭也相對提升。在此環境下，如何了解顧客消費行為與需求，以有效地執行目標行銷並降低顧客流失，是企業提升競爭力的重要方法(Malhotra & Blattberg 2004)。許多企業利用資料庫蒐集顧客消費狀況與產品銷售資料，期能藉此了解顧客需求並制定行銷策略。為此，學者提出資料庫行銷的概念。

資料庫行銷包括顧客價值的分析與顧客未來交易的預測等(Pinto et al. 2009)，顧客分群是資料庫行銷中最基本的一環(Shaw & Stone 1990)，主要是將顧客分成不同群體，使每一群體具有內部同質性與外部異質性，此意味著同一群顧客的消費行為特徵會趨向一致，並且迥異於其他群體之特徵。而顧客分群模式的行銷效果也遠較整體顧客行為考量的行銷來得有效(Andrews et al. 2010)，因此一直是行銷者提升行銷效率常用的手法。傳統分群方式多依據顧客的基本資料(如：年紀、教育、收入、居住地等等)、交易金額、或採購物品種類進行區隔(Peterson et al. 1993)，近期則有以網頁搜尋之序列(Jiang & Tuzhilin 2006)、顧客之不忠實傾向(Chen et al. 2007)、顧客之購買數量金額與成本(Markic & Tomic 2010)，或顧客購物行為(Liu & Shih 2005a)等方式來對消費者進行分群。現有的資料庫行銷研究大多強調顧客分群方法，或是探討如何利用資料庫進行顧客價值分析，雖各有其特色，但仍缺乏一個包含流程、建模、策略規劃與策略驗證之完整資料庫行銷架構。

本研究目的在結合資料探勘的技術，提出一個以顧客價值作為分群基礎之完整資料庫行銷架構，以包含上述諸要項。此架構之分群技術包含群集分析與序列型樣(sequential pattern)分析，前者是將顧客依條件分群，後者是將產品依消費關聯分群。本研究擴充現有顧客價值分析模式，做為顧客分群的基礎，以找出不同群集顧客的特徵與消費行為。此擴充模式除了採用以往顧客消費的RFM指標，亦將顧客價值趨勢與產品銷售折扣納入群集分析中；產品消費關聯分群則是以顧客在不同時期之消費關聯做分群，藉由顧客分群與產品消費關聯分群的對應，使不同分群之顧客的目標行銷範圍更明確，並據此規劃商品服務與發展行銷方案。行銷方案的規劃是以顧客與產品分類的配對為基礎，藉由專家群體討論提出可行方案，並對可行方案進行彙整，再由專家做進一步的確認。

此研究的貢獻有五：一、提出一完整的資料庫行銷流程與架構；二、此架構的分群模式結合了群集分析與序列型樣分析，以進行顧客和產品分群；三、擴充傳統顧客價值分析模式，納入顧客價值趨勢與折扣為分析因子；四、以顧客與產品分群做為行銷策略制訂依據，優於以單一分群因子做為行銷策略制訂的考量；

五、提供資料庫行銷策略之成效的驗證方式。簡言之，此研究成果不僅提供企業管理者一套資料庫行銷的執行方法，其行銷策略的制定方向與成效驗證方式亦可作為決策者的參考。

貳、文獻探討

本章所回顧的相關文獻包括：顧客價值、資料庫行銷與資料探勘技術，並且對顧客價值分析模型之參數及計算方式進行說明。

一、顧客價值分析與 RFM 的評估

在大量生產與競爭全球化的環境下，如何獲得與保留有益的顧客，幫助企業更有效率地執行市場銷售競爭，是行銷者致力的目標（Bult & Wansbeek 1995; Malthouse & Blattberg 2004）。顧客價值（customer value）即是學者針對辨識有益顧客與發展相關策略所提出的典型指標（Irvin 1994），顧客價值的定義多元，可由顧客和企業的角度分成兩種相對觀點：顧客觀點之顧客價值強調產品為顧客帶來的知覺感受或財務利益（Albrecht 1994; Woodruff 1997），或是顧客對產品的品質、服務和價格的評價（Zeithaml 1988; Han & Han 2001）。實證研究顯示，此觀點之顧客價值會影響顧客滿意度、顧客忠誠度（Tsai et al. 2010），學者強調，著重於顧客價值最大化的管理方式，能達到較佳的顧客關係（Gupta et al. 2004）。

企業觀點之顧客價值則強調顧客為企業帶來的利益，主要取決於顧客目前或未來可能的淨獲利，學者認為，顧客在與該企業往來的特定時間內，企業利潤方面所呈現的平均淨現值可代表顧客價值（Pepper & Rogers 1997; Hughes 2006），或是可由顧客交易的現金總額與加權平均資本成本之計算折現後的總值來表示（Kumar & Rajan 2009）。藉由對顧客終身價值的計算與運用，企業能把適當的資源集中在獲利潛力較大的顧客，提高行銷資源的效益。此企業觀點之顧客價值被視為公司市場價值的指標之一，亦為本研究分析顧客價值的指標。

分析顧客價值的有效方法之一是 RFM 評估模式（Liu & Shih 2005a），此模式以 RFM 指標代表顧客消費的近期（Recency, R）、頻率（Frequency, F）與金額（Monetary, M），行銷學者 Goodman（1992）認為 RFM 指標能用以評估和辨識出有益（profitable）顧客，讓行銷投資效益達到較佳效果。RFM 指標不僅可用於顧客分群（Bauer 1988），亦可用來分析顧客的消費特質，作為未來消費預測的基礎（Colombo & Jiang 1999）。Liu 與 Shih（2005a）藉 RFM 指標計算顧客價值來進行顧客分群，再結合資料探勘，進一步提出推薦購買的產品以支援目標行銷。

隨著產品的特質差異，RFM 會有不同的定義。例如，針對撥接上網之消費，RFM 定義為：R 是指最近一月來平均的上網間隔（天）；F 是指每月平均的撥號上

網次數（次數/月）；M 是指每月平均的撥號上網時數（小時/月），此定義涉及到頻率或次數，所以較適用於提供多樣性商品的企業，或是該產品能提供顧客多次消費的企業。本研究延伸 Hughes（2006）之做法，以 RFM 代表產品購買之最近日期（R）、購買頻率（F）與購買金額（M），定義如下：

- R（近期）：顧客最後一次消費日距離分析日的天數。
- F（頻率）：單位時間內的購買次數。
- M（金額）：單位時間內的消費金額。

R 值越高代表此顧客沉寂時間較長；R 值越低則代表顧客有較高的再購買傾向，F 值越高代表此顧客之購買頻率較高，M 值越高代表此顧客消費金額越高，越能提高公司收入。因此，顧客之 RFM 指標是值得企業重視（Liu & Shih 2005b）。Verhoef 等人（2002）透過實證研究之問卷統計發現，除了透過直覺，行銷經理最傾向使用常期性交叉列聯表（cross-tabulation）與 RFM 指標來進行顧客分群。

除了考量 RFM 指標的三個變數，張心馨與蔡獻富（2004）也加入了顧客關係長度（L）的概念，關係長度代表顧客與企業之關係長短，是以顧客第一次購買日到最後一次購買日之間的天數計之，該研究分別利用 SOM 和 K-mean 對 RFML 進行兩階段分析，其分析數據顯示顧客關係長短會影響顧客對企業之獲利貢獻，然而此結果與其他學者例如 Reinartz 與 Kumar（2000），之看法迥異。亦有文獻使用 RFMTC 指標預測顧客未來購買次數，所考量的 T（Time）代表顧客購買時間距第一次購買的時間長短，C（Churn）則是顧客每歷經一次行銷後流失的機率（葉怡成等 2008）。

相較於傳統以顧客基本屬性與背景來進行市場區隔的方式，RFM 指標更能有效地區隔顧客，因此廣被採用。除了使用 RFM 進行顧客價值分析，Jen 與 Wang（1998）採用比較最大概式估計法（Maximum Likelihood Estimation, MLE）與加權最大概式估計法（Weighted MLE, WMLE）的差異來分析顧客的價值趨勢，其計算方式是以 MLE 代表平均購買時間間隔，而 WMLE 則是加權平均購買期間，其加權方式是根據相對最近的日期排序值做為權重值，計算出 MLE 與 WMLE 差異後便據此將顧客分群，分群方法詳述於後。

二、資料庫行銷與資料探勘

資料庫行銷是一個以資料庫為基礎，探索資料庫中資料所潛藏的訊息以支援行銷活動或行銷決策的程序（Pinto et al. 2009）。Kotler 與 Keller（2008）認為資料庫行銷會成為未來行銷的基礎，業者必須掌握顧客的喜好和購物特性，才能適時提供較合適或特定的個人化服務。資料庫行銷主要是利用資料庫中所儲存之顧客資料與交易資料，分析顧客的消費特質，藉此進行顧客的分群與顧客未來交易的

預測，進而提供產品推薦或行銷決策所需之資訊 (Paas 2009; Roberts 1992)。決策者可以透過資料庫檢索查詢顧客的交易狀況、增加對顧客的了解，進而提升交易機會，而且也可以計算出各種行銷活動的直接財務報酬和非直接財務報酬 (Paas 2009)。顧客資料庫與資料分析是顧客關係管理的元件，如何利用有效工具分析顧客資料以提供行銷決策是資料庫行銷的要務之一 (Verhoef et al. 2002)。然而，若以傳統統計方法作為資料庫行銷的分析工具，所能獲得的訊息和分析效能有限，例如：分析耗時（因資料量大）、無法有效分群（缺乏分析彈性），以及無法找出隱藏在資料庫中的型樣（pattern）或關聯性，而缺乏預測所需的訊息。學者建議利用資料探勘的技術，根據所欲銷售的產品或服務進行顧客區隔，並依此規劃特定的行銷策略 (Wong et al. 2004)。

資料探勘的目的在於發掘資料庫中潛在的重要資訊 (Fayyad 1996)，資料探勘所獲取之知識可用於銷售預測或成為決策支援的重要依據 (Kleissner 1998)。主要做法是利用不同的方法或演算，從大量資料中挖掘隱藏的知識、趨勢 (trend)、型樣 (pattern) 及關聯性等。本研究利用資料探勘工具 SPSS Clementine，混合 EM 群集演算法和 Markov Chains 的序列型樣演算法 (Tang & MacLennan 2005) 對資料進行計算。

群集分析演算法是資料探勘的方法之一，作法是將所有資料根據其屬性值區分成數個群集 (cluster)，使同一群集內部的資料具有較高的相似度，而不同群集的資料則具有較低相似度，從而推論出每一個群集的主要特徵。在群集分析的方法中，K-means 演算法是一種廣為被採用的方法，它最早是由 Forgy 於 1965 年提出。K-means 演算法假設每一筆資料有 n 個屬性，則每一筆記錄可視為 n 度空間中的一點，這些點可被分成 K 個群集，在各個群集中，令每一群資料點的平均值為中心點。分群的目的在使每一資料點距離其所屬群集之中心點的距離最小，距離其他群集之中心點的距離較大，並且每一個資料點，只能隸屬於一個群集，故群集之間不相連也不重疊，此種 K-means 演算法技術被稱為 Hard 群集法。K-means 的群集步驟包括：一、決定要產生的叢聚數目 K 。二、任意選擇 K 個叢聚中心點。三、將所有資料點分配到距離最短的中心點。四、計算新的叢聚中心點（運用 Euclidean distance）。五、重覆三與四的步驟，直到所有叢聚內的記錄點不再改變 (Hartigan & Wong 1979)。

期望最大化 (Expectation Maximization, EM) 也是有效的分群方式，它採用機率的評估以決定資料點隸屬於每一個叢聚的比率 (Lauritzen 1995)，並以此進行分群。EM 演算法是根據平均值與標準差，對每一種維度建立鐘形曲線 (bell curve)，凡是資料點落在鐘形曲線內，EM 就會給予該資料點隸屬於這個叢聚的機率。本研究除了以 K-means 分群，亦採用 EM 分群，並且進行這兩個分群的效果比對。序列型樣分析的目的是找出顧客在不同時間點所採購商品之先後關係（非時間格式

資料也可以指定為序列性資料屬性，像是交易流水號)。根據交易時間對每一位顧客的所有交易進行排序所得到的結果就是一個序列，稱為「顧客序列」。一個序列 s 的支持度被定義為「包含 s 的顧客序列總數佔全部顧客總數的比例」。若 s 序列滿足使用者設定的最小支持度，則稱之為「大型序列」。大型序列 s 若不被包含在其它大型序列中，則它被稱為「最大序列」，每一個最大序列所代表的是一個序列型樣。

參、研究個案與研究架構

本文是以 A 公司之業務資料與業務特質做為研究個案，藉由訪談該公司資深經理人員，以了解其經營方式，並取得該公司之資料庫資料，作為本研究實驗之依據。首先採取專家訪談，以了解企業特質與經營方式，並據此提出資料庫行銷之架構與相關之流程，最後再依實驗數據，請專家小組進行實際行銷建議。

一、研究個案

A 公司的三項主要業務為：雜誌、電子商務網站和研討會，此三項業務採個別經營，採取會員優惠制。雜誌分為平面媒體與數位化內容網站，平面雜誌內容以生活領域為主，內容網站除提供有關電腦網路技術之報導，並整合前者部份平面雜誌內容，以數位化方式供人下載閱覽。電子商務網站主要銷售國內各電腦出版商的書籍及電腦週邊的輕型商品，其中的網路書店有提供會員折扣優惠。研討會則是每年舉行，透過在台灣北、中、南舉辦之資訊科技研討會概略介紹最先進的產品或技術。研討會大多為促銷商品所舉辦的活動，可免費參加。

A 公司資料庫的資料分三大類：一、顧客基本資料；二、產品資料；三、顧客交易資料。顧客交易分為四種狀態：未處理訂單、已處理訂單（向廠商取貨並取得款項）、已出貨訂單（開立發票並連絡宅配送貨）、取消訂單。所謂的有效訂單，是侷限於已出貨的訂單，其餘訂單則不納入本研究的資料範圍。

二、資料庫行銷之程序架構

根據 A 公司之業務與資料庫之資料，本研究將資料庫行銷的流程分為四個階段：定義問題、資料處理、模式建構與分析、策略發展與系統整合，列於圖 1。

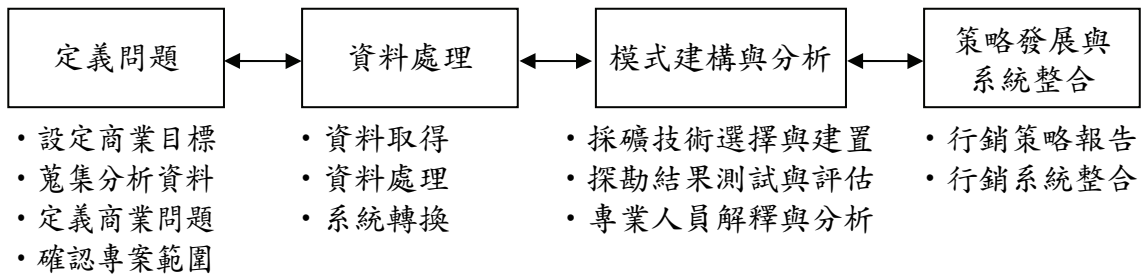


圖 1：資料庫行銷的流程建議

除了第一部份的定義問題，後續三個步驟即是藉由資料探勘輔助資料庫行銷的主要部分，其方法與細節呈現於圖 2，說明於下：

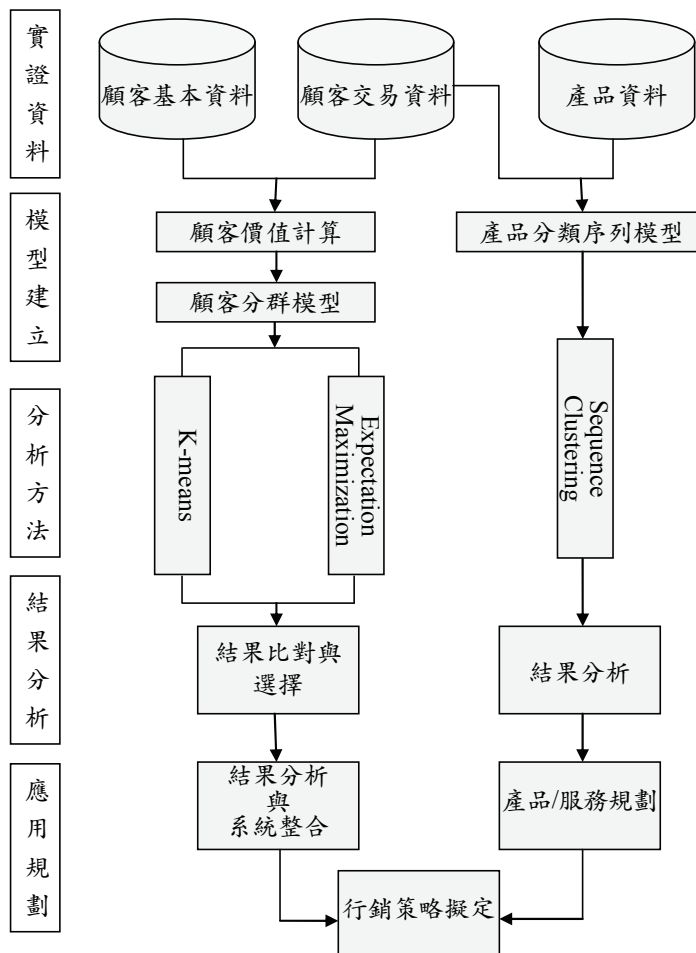


圖 2：資料探勘輔助資料庫行銷之架構圖（本研究提出）

- 實證資料：根據問題定義（如圖 1）蒐集所需之顧客基本資料、產品資料與顧客交易資料，並進行適當的資料前置處理與資料彙整。
- 模型建立：依據顧客消費資料，除建立 RFM 模型，並統計顧客的價值趨勢與交易折扣（discount）百分比外，也使用序列型樣技術找出顧客購買產品類別之先後順序，該技術所涉及的計算方法則詳述於後。
- 分析方法：本研究使用群集分析的兩種演算法 K-means 與 EM 分析顧客分群模型的數據。另使用序列型樣技術找出顧客產品類別購買之先後順序，序列中項目集合的順序是顧客消費時間的先後順序，所使用的演算法是馬可夫鍊，經由馬可夫鍊進行群集相似性計算後，可以讓落在相同群集的顧客有高度相似的序列模式，顧客採購行為更容易辨識出來。分群結果則採用統計與分類矩陣圖等方式呈現最佳的方案，然後再探討各群集內人口統計資料之分佈，以瞭解顧客屬性對於顧客採購行為是否有影響，如：性別、教育程度、年齡等。
- 應用規劃：根據分析結果擬定適當產品／服務規劃與行銷策略，並將探勘所得到的分群條件整合至應用系統模組之 SQL（查詢語言）條件式的參數中，做為目標查詢的依據。

顧客資料之分群方式可依底下描述表之：令 $C = \{C_1, C_2, \dots, C_N\}$ 代表顧客群，共含有 N 個顧客，每一顧客 $C_x \in C$ 由一組 (A, T) 表示， A 是一群含有 m 個基本屬性的集合， $A = \{A_1, A_2, \dots, A_m\}$ ，若顧客 C_x 執行過 r 個交易，則 $T(C_x) = \{T_1, T_2, \dots, T_r\}$ ，每筆顧客 C_x 所執行過的交易 $T_i (i=1, \dots, r)$ 由一組交易屬性 $\{B_1, B_2, \dots, B_k\}$ 表示。分析時，合併顧客的基本屬性資料與交易資料，所以每一位顧客每次交易的完整資訊是 $(A, T) = \{A_1, A_2, \dots, A_m, B_1, B_2, \dots, B_k\}$ ，也就是分析單位。

為了有效對顧客分群，此研究採取五種指標，包括：最近消費、消費頻率、消費金額、顧客價值趨勢、顧客折扣交易次數百分比。首先依網站產品性質與銷售屬性定義 RFM 評估模型如下：

- 最近購買時間間隔 (R)：距離「資料萃取日」兩年內之最近消費時間
- 總購買次數 (F)：兩年內總購買次數
- 平均購買金額 (M)：兩年內平均購買金額

此方式是先將顧客資料以「最近消費時間」過濾，最近一次消費時間若已超過兩年，則將該筆資料淘汰，以避免偏離值，其餘則按時間比例擴充次數。處理過後的資料再根據之前 RFM 的定義，逐一進行資料整合與轉換，將兩年內所有交易資料分別計算出其 RFM 值。顧客對於期刊或雜誌若有多月份訂購，則屬於一次性消費。例如：某顧客總計下了三筆訂單，經過資料處理換算後的交易明細如表 1：

表 1：某顧客交易明細

順序	交易日期	金額	(今日 - 交易日期)
1	2010/12/01	3,890	5
2	2010/11/20	2,410	16
3	2010/10/01	1,390	66

假設今日是 2010/12/06，所得之 RFM 值為：

$$R = 5 \text{ (天)}, F = 3 \text{ (次)}, M = (3890 + 2410 + 1390) / 3 = 2563 \text{ (元)}$$

關於顧客的價值趨勢計算方式如下：首先計算顧客訂單之間隔（單位：日），最多記錄 6 個有效訂單的 5 個時間間隔，若顧客 C_x 有 r 次購買記錄，即 $T(C_x) = \{T_1, T_2, \dots, T_r\}$ ，則 $r = \min(r, 5)$ ，且令 $t_i = \text{date}(T_{i+1}) - \text{date}(T_i)$ ，其中 $1 < i \leq r$ 代表訂單的順序，訂單順序的時間點是由目前的時間往前推， T_1 與 T_2 則分別代表離目前最近的第一筆和第二筆訂單， t_i 即是顧客 C_x 第 i 次購買與第 $i+1$ 次購買的時間間隔，則 MLE 與 WMLE 值的計算如下：

$$MLE = \sum_{i=1}^r t_i, \quad WMLE = \sum_{i=1}^r i * t_i$$

藉由 MLE 減去 WMLE 值呈現顧客購買產品的時間間隔變化：若 $MLE - WMLE > 0$ ，表示顧客購買產品間隔時間有愈來愈短的趨勢；若 $MLE - WMLE < 0$ ，則有愈來愈長的趨勢，以此變化表達顧客之價值趨勢。此外，本研究採用 Jen 與 Wang (1998) 所提出之顧客價值趨勢分析加上近期 (R) 與金額 (M) 平均數之列表法，可將顧客區分成不同特質，如表 2。

表 2：顧客特質 (MLE-WMLE) 分析

	$F > \text{平均數}$	$F < \text{平均數}$	$M > \text{平均數}$	$M < \text{平均數}$
$MLE - WMLE > 0$	忠誠顧客	機會顧客	忠誠顧客	機會顧客
$MLE - WMLE < 0$	消極顧客	流失顧客	消極顧客	流失顧客

本研究所提出的「顧客折扣交易次數百分比」指標，一方面探討電子商務網站，採用產品價格折扣策略對於顧客之影響力，一方面希望找出不因折扣誘惑而長期購買的忠誠顧客。在本研究中，顧客折扣交易次數百分比計算方法如下：

$$\text{顧客折扣交易次數百分比} = \text{因折扣方案而採購的總筆數} \div \text{交易總筆數} \times 100\%$$

本研究考慮兩年內個別顧客的交易總筆數，例如：某一位 B 顧客總計下了四筆訂單，經過處理換算後的資料如表 3，則 B 顧客的折扣值 $= (1/4) * 100 = 25\%$ ：

表 3：B 顧客交易明細範例

順序	交易日期	是否在折扣活動期間
1	2011/01/23	Y
2	2010/12/01	N
3	2010/11/20	N
4	2010/10/01	N

三、資料庫行銷策略發展程序

根據 A 公司業務特質所提出之資料庫行銷策略的整體程序如圖 3，簡述於下：

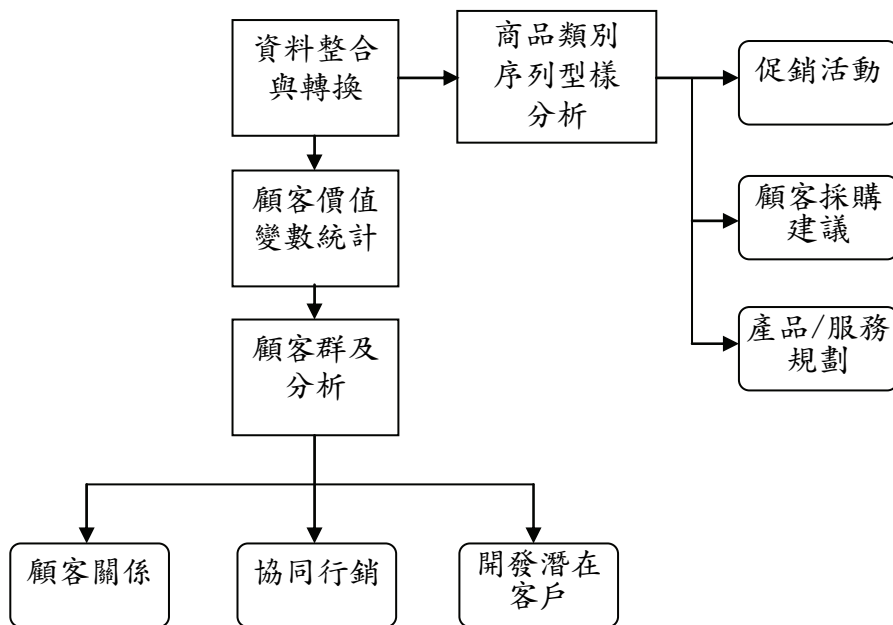


圖 3：資料庫行銷策略之發展流程

1. 資料整合與轉換：依資料探勘的目的，從交易系統中匯出需要的欄位資料。為求探勘結果的正確與效率，原始資料的前置處理是不可輕忽的；同時因應研究的需求，衍生欄位的建立，更容易增加隱藏資訊的發現。
2. 顧客價值變數統計：從顧客交易資料，計算每位顧客的顧客價值，以作為顧客分群的基準。

3. 顧客群集分析：使用顧客分群技術以瞭解各群集特徵與消費行為，發展多種行銷策略，將合適的商品與服務推薦給有意願與需要的顧客，降低行銷成本並建立顧客對企業的信賴。
4. 顧客關係：紅利積點在多種商業活動中，是維繫顧客再度光臨的方式之一；此外，提供顧客需要的相關資訊，在時間等於金錢的現代社會中，是很好的附加價值服務。「紅利積點」與「提供相關資訊」這兩種利器的使用，端視顧客的消費頻率與金額而定。
5. 開發潛在顧客：得知顧客的特質與消費行為後，即可以交叉行銷方式，開拓其它資料庫中的潛在顧客。
6. 協同行銷：積極尋找合作伙伴，增加產品與服務的廣度，吸引更多商機。
7. 商品類別序列型樣分析：對於大量產品，建立商品類別之間被採購的關聯與次序，有助於商品行銷策略的規劃。
8. 促銷活動：不同型態商品依顧客需求，進行行銷組合。例如：本期雜誌焦點為「資訊安全」，則銷售期間對於相關書籍加強介紹或進行促銷，或是針對顧客目前所購買的產品項目，預測未來可能需要的關聯產品項目。
9. 顧客採購建議：將「商品類別序列型樣分析」的結果，整合至線上交易系統中，將此分析結果作為顧客採購建議之依據，使建購之商品種類範圍縮小，以增加銷售機會。
10. 產品／服務規劃：依據資料探勘分析結果，機動調整經營的方向與產品服務的內涵，以滿足顧客需求。

肆、實證分析與討論

本實驗使用的資料庫包含顧客基本資料 23,202 筆，銷售記錄有 34,349 筆。採用兩年內資料，經過前置處理、過濾與篩選，產生適合分群分析的顧客資料共 7,344 筆，銷售記錄共 11,340 筆。顧客資料的過濾主要是去除那些有加入會員卻尚未有交易的顧客資料，以及在某些人口統計屬性有缺值的資料列，這些人口統計屬性包括：性別、年齡、教育程度、工作性質、地區。顧客分群所需的交易紀錄則篩選屬於這 7,344 位顧客的消費資料中的有效訂單，共 11,340 筆。序列型樣分析則採用兩年內的銷售記錄，共 29,586 筆。分析結果含顧客分群、交易之序列型樣分析與各群集內顧客人口統計。

一、顧客分群結果

本案例中的變數有範圍大的連續數值，因此利用內建在資料探勘工具的處理

分配調整與正規化 (normalization)¹之功能，目的是讓具有不同範圍之連續數值的各屬性在模型中的影響力均等。本研究最初依據前述五個指標以 K-means 與 EM 分別對顧客分群，並使用單因子變異數分析，探討五項指標對分群的貢獻度，結果顯示各變數對分群結果均達到顯著性水準 ($p < 0.05$)。採用 K-means 所得到之分群統計值如表 4：共分為四個群集，以群集一至群集四命名之，各指標與各群集之屬性值則描述於後。一般 K-means 作法是針對一組觀察值的數個特定維度，執行多次分析以找出這組觀察值的最佳分群數目，本研究是以 Goodman (1992) 的顧客價值作為分類之理論基礎，並藉由每次實驗中反覆納入一組不同的維度，將顧客分成理想的四群，所有考量的維度包含：R、F、M、T 與 D 五項（如表 4）。維度若只採用購買金額，雖然仍可用 K-means 時可分為四群，但執行 EM 時，卻只能分為一群。因此，同時納入五個維度，才能讓 K-means 和 EM 分群兩者皆將實驗中的觀察值分為四群。

表 4：K-means 各群集屬性值表

變數\群組	統計值	總樣本	群集一	群集二	群集三	群集四
Recency (R) 下限：7 上限：730	平均	358.60	380.85	260.61	174.24	130.91
	標準差	209.09	207.13	182.37	149.73	127.93
Frequency (F) 下限：1 上限：5.93	平均	1.54	1.16	2.79	6.31	10.87
	標準差	1.47	0.44	1.54	3.74	6.86
Monetary (M) 下限：56 上限：10356.66	平均	1799.64	1011.11	4131.88	11530.05	34275.48
	標準差	2852.34	583.51	1312.03	3330.41	12054.85
Trend (T) 下限：-26.14 上限：25.29	平均	-0.43	-0.10	-2.49	-0.85	4.60
	標準差	8.57	4.09	18.82	20.19	14.22
Discount (D) 下限：0 上限：100	平均	16.60	15.92	19.73	21.02	28.22
	標準差	34.43	35.44	29.08	24.68	28.50
	樣本數	7,344	6,152	989	180	23

然而，以分類模型檢測發現，EM 分群結果的正確率比採用 K-means 分群結果的正確率低。本研究在獲得 K-means 和 EM 的分群結果後，利用各群的结果建立分類模型，根據分類矩陣來判斷分群的正確率。實驗結果如附錄中表 A-3 和 A-4 顯示，根據 K-means 分群結果建立決策樹，其分類矩陣所顯示之數據，共誤判²了

¹ 選用的正規化演算方式採全距法：調整後的值 = (原始值 - 最小值) / (最大值 - 最小值)

² 此處的誤判定義為：某觀察項在決策樹分類矩陣之分類結果與原先的 K-means 分群結果不同。

20 筆資料，判別正確率³為 99.72%，而 EM 分群結果誤判比數有 567 筆，K-means 分群的判別正確率高於 EM 分群的正確率。相較之下，以 EM 的分群結果建立分類決策樹之效果較差，因此本研究以 K-means 的分析結果為依據。

近期 (R)：樣本之最近消費日的顯示範圍從 7 天到 730 天。R 值愈小代表顧客最近消費日愈近，所以此數值越低越佳，可代表較傾向為能獲益的 (profitable) 顧客。各群集的獲益傾向最優是群集四，群集三次之，接著是群集二和群集一，以「群集四 > 群集三 > 群集二 > 群集一」簡示之。表 A-1 中 (參附錄)，R 值被分為四個區間，並顯示 R 值在不同區間的機率。群集一的平均值是唯一超過總樣本的平均值，但其標準差也是四個群集中最大者，代表在這群集中，有少部份的顧客之最近消費日其實並不遠，群集一甚至有 27.86% 的機率，顧客的最近消費日是落在 7~217 天的區間內，這些 R 值小卻被分類到群集一的顧客，代表他們雖然最近有來消費，但可能購買金額低或消費次數少等原因，才會落在平均消費日最遠的群集中。其它群集亦有類似群集一的情況，同一群集內顧客 R 值之範圍則是分散在各區間內。

頻率 (F)：樣本顧客之平均消費次數所顯示的範圍從 1 到 5.93 次，此一變數值愈大代表顧客消費次數愈多，此數值越高越佳，代表該群集傾向為較能獲益之顧客，各群集頻率均值高低是「群集四 > 群集三 > 群集二 > 群集一」。附錄的表 A-1 中，F 值被分為三個區間，並計算所有群集內的 F 值在不同區間的機率。從表 4 可以明顯看出群集三與群集四不管是加一個標準差或減一個標準差，平均消費次數都超過總樣本平均值，代表這兩個群集內的顧客幾乎都是經常消費的顧客。然而這兩個群集內的屬性值較分散，這代表這兩個群組內顧客購買次數的差距頗大。反之，群集二內的屬性值較密集，代表群集二中，顧客消費次數較一致。群集一則大部份的顧客消費次數不超過兩次。在群集二中，超過 50% 的顧客有 2.5 次以上的平均消費次數。群集一中，大多數顧客的平均消費次數小於 3 次。

金額 (M)：樣本顧客平均消費金額的顯示範圍從 56 元到 10356.66 元，金額愈大代表顧客平均購買金額愈多，代表較能獲益 (profitable) 顧客，各群集之平均購買金額的均值高低是「群集四 > 群集三 > 群集二 > 群集一」。表 A-1 中，各群集金額分佈表是將 M 值分為三個區間，並計算所有群集內的平均 M 值在不同區間的機率。從表 4 可以發現，群集二、三、四不管是加一個標準差或減一個標準差，顧客消費平均 M 值都超過總樣本平均 M 值，代表這三個群集是高消費族群；群集一之平均值低於總樣本平均值，有 85% 以上的顧客消費金額是低於總樣本平均值。

趨勢 (T)：樣本顧客消費時間間隔變動 (百分比) 趨勢範圍從 -26.14 到 25.29，

³ 判別正確率 = (全部觀察項的個數 - 被誤判之觀察項的個數) / 全部觀察項的個數 × 100%

趨勢數值愈大代表顧客購買的時間間隔變愈短，顧客忠誠強度愈高。各群集之平均購買時間間隔的均值高低是「群集四 > 群集一 > 群集三 > 群集二」。T 值被分為四個區間，並計算所有群集內的 T 值在不同區間的機率。群集四雖是唯一平均值為正值者，但其標準差也非常大，群集四內有超過 21% 機率顧客購買產品的時間間隔是漸長的趨勢。群集一的平均值雖是 -0.1，但群集內有極少數顧客購買產品的時間間隔有漸短的趨勢。群集二與群集三內資料值甚為分散，代表這兩個群集內顧客 T 值的範圍也是分散在各區間內。在趨勢方面，各群集的忠誠強度是「群集二 < 群集三 < 群集一 < 群集四」。

折扣 (D)：此一變數值愈大代表顧客受促銷活動影響愈深，因此能藉由顧客而獲益強度愈低。表 A-1 顯示所有群集內顧客參與促銷活動的百分比之顯示範圍從 0 到 100；D 值被分為三個區間，以及所有群集內的 D 值在不同區間的機率。除了群集一之外，各群集的平均值都超過總樣本平均值，且所有群集的變異數值都很大，群集內 D 值之範圍分散在各區間內，代表這四個群集近似。數據顯示各群集顧客受到折扣活動影響的程度。在折扣方面，各群集的顧客獲益強度是「群集四 < 群集三 < 群集二 < 群集一」。

附錄表 A-2 整理了 K-means 分群之各項變數的條件範圍，分群特徵值即是群集分群條件，這些範圍值，可利用表格方式自動儲存，做為模組參數的輸入，並依據顧客的新增採購而定期重估與分群，再根據分群結果更新表格，以便有效查詢所需的行銷目標。依據前述的顧客分群定義，將群集四歸為忠誠顧客、群集一是流失顧客、群集二與群集三均應認定為消極顧客。但因群集三中，有將近 35% 顧客的 T 值接近正值，群集三的 R 值平均值是總樣本 R 值平均值的一半，其 F 值平均值與 M 值平均值均分別超過總樣本的 F 值平均值與 M 值平均值，故將群集三歸屬於機會顧客，需提供不同誘因以增加其購買機會。

二、序列型樣分析

本實驗採用時序群集 (sequence cluster) 分析，從同一群集的顧客行為中找出有意義的產品類型與購買規則。主要方法是以交易流水號做為序列性資料屬性，找出顧客在不同時間點所採購商品之先後關係。每一位顧客所有交易所構成的序列稱為顧客序列，一個序列 A 的支持度定義為「包含 A 的顧客序列總數佔全部顧客總數的比例」，滿足使用者設定的最小支持度稱之為大型序列，每一個最大序列代表一個序列型樣。本研究所使用的演算法是內建於探勘工具中的馬可夫鍊 (Markov Chain)。

所有資料整理如表 5，分析之資料為兩年內所有顧客 (不侷限於用以分群的顧客) 的 29,586 筆交易資料，不預設分群個數，由系統自動偵測可能的群集數量。

序列型樣的分析結果顯示，所有交易資料被分成 17 個群集，群集 3 到群集 17 大多只聚焦在單一產品類型上（故省略），但從群集 1 與群集 2 中則找出了 13 項規則（關聯性較強的連結）。表 5 中的人數，代表滿足該序列群集中任一規則之顧客序列的顧客人數，也就是有此消費規則的顧客人數。序列型樣分析結果目的有三：一、支援產品與服務項目推薦之電子化作業，藉由分群結果編排特定相關產品，以利顧客在更短時間，瀏覽到較有興趣的項目。二、提供管理者檢核產品進貨與庫存量之參考；三、作為雜誌與研討會相關產品介紹內容以及邀請對象之規劃參考。

表 5：序列型樣分析結果

群集名稱	產品類別與類別轉換規則	信心%	人數
群集 1	Unix → Linux 電腦輔助教學 → 工程製圖 作業系統 → 遊戲設計 文書處理 → 試算表 企業電腦化 → 電子商務 HTML → 網頁設計	21% 19% 29% 17% 18% 14%	4886
群集 2	採購指南 → 硬體設備/組裝/維修 解毒軟體 → 硬體設備/組裝/維修 作業系統 → 影像處理 作業系統 → 繪圖 ORACLE/INFORMIX → JAVA Unix → Linux 應用管理/理論/方法 → SQL	54% 30% 31% 33% 22% 25% 27%	1009

三、顧客人口統計特徵

從 7,344 筆顧客資料與 29,586 交易資料中，群集分析初步將顧客分成四類：忠誠顧客（23 位）、機會顧客（180 位）、消極顧客（989 位）、流失顧客（6,152 位）。從各個群集的樣本特徵可以發現，四個群集內顧客分佈與總樣本性質相近，男性顧客佔 7 成以上；年齡在 26~45 歲之間高達 7 成 5；教育程度以專科和大學為首；工作性質除了「其他」，以學生與工程師最多；分佈區域以北部地區為主（請參附錄中表 A-5）。簡而言之，此四大顧客族群之人口特徵沒有明顯的差異性，若想從這些人口統計特徵發展出獨特的行銷策略，著力點不大，因此，以顧客之消費行為分類來發展整合行銷策略才是最佳的方式。分析結果亦顯示，交易在四次以上顧客有 444 人，不受到促銷活動影響的顧客總共只有 155 人（34.90%）。

綜合前述之分析，為了針對不同群集之特徵提出行銷策略，本研究對專家進行群體訪談，經由專家的討論與提出應有之執行方案，最後再彙整出執行方案並由專家確認如表 6。

表 6：各群集之行銷策略對應表

顧客分群	行銷策略	執行方案
流失顧客	需告知有關公司產品與服務的新資訊，無須提供過多的行銷資源。	以電子報定期傳送所提供的產品與服務訊息。
消極顧客	除告知公司產品與服務的新資訊，應瞭解當下顧客的需求與之前消費經驗，整合顧客意見，在能力範圍內滿足顧客新需求。	除了定期的產品與服務訊息傳遞，請顧客填寫有關需求與建議的問卷調查。並在顧客同意下只要「One click」，就直接成為雜誌普通會員，定期將半年內最新的數位雜誌內容標題與摘要依時間排序，以電子報方式推薦給顧客。
機會顧客	除告知公司產品與服務的新資訊，並規劃轉換這群顧客消費習性，使其朝向忠誠顧客推進。	除了定期的產品與服務訊息傳遞和「One click」雜誌會員加入，並以各項增值服務取代產品折扣。廠商所提供的特價商品儘量搭配附加贈品，改變顧客直接因為價格因素才消費的行為。
忠誠顧客	公司應隨時掌握顧客新需求與改變，適時提供顧客需要的資訊。並建立互動的管道，採納顧客適當的建議。	發行「終身學習護照」，顧客可隨時查詢曾經購買的產品、閱讀的文章與參加的研討會。有適合顧客的產品提前通知，並保留所有研討會優先參與和最優惠折扣的權力。

為驗證各群集行銷策略的有效性，此個案採獨立樣本與隨機分派方式，將各個群集區分為甲、乙兩組，每組人數大約相同。實驗分為兩階段：第一階段甲組為實驗組，乙組為對照組，第二階段則反之，每階段實施半年。第一階段是對每一顧客群之甲組執行特定促銷，促銷方式是根據表 6 之方案執行；而對乙組之每一顧客，則以亂數任選一不同於對甲組顧客所採用之方案，換句話說，就是對同一群顧客分組，但對於不同組之顧客則採用不同方案。然後以方案實施半年的交易紀錄做為分析數據。結果發現，甲組全體顧客的交易總金額有大幅提升，乙組全體顧客的交易總金額則小幅度提升，甲組的交易總金額亦明顯高於乙組。第二階段則將甲組改為對照組，乙組為實驗組，此階段的乙組的交易總金額則高於甲組。然而此個案作法有底下的不足：一、甲乙兩組的分群後，應該先以獨立樣本 T

檢定分析其平均金額與總金額是否有顯著性差異，若無顯著性差異，再進行實驗。二、促銷方案實施後，必須進行甲乙兩組後測之差異性探討，檢定實驗組的交易金額高於對照組，並且有顯著性差異。三、檢定實驗組在促銷前後的交易金額是否有顯著性增加。如此，才能顯示行銷策略的有效性。

四、綜合歸納與建議

分析結果與資料庫策略大致歸納如下：

1. 群集分析：從本案例的群集分析結果發現，大部份的顧客屬於流失顧客，消費金額少，對於企業的利益很小。分佈區域以北部縣市為主，有八成以上消費與價格優惠有關。可見單純以購物網站的方式經營，較難有明顯成果。針對這個問題，本研究建議整合另外兩項業務，包裝成一個複合商品，並整合多種增值服務的行銷策略。
2. 序列型樣分析：從本案例的序列型樣分析中，可以觀察到一些趨勢轉移。但是受資料量影響，能發現的產品類別序列關聯並不多。
3. 行銷系統整合：行銷策略對應表的目的是推動不同群集朝向提高消費金額。實證分析中所採用的群集分群條件，可以整合到行銷系統模組中，對消費的行為提供自動化的回應與調整。也就是隨時因應消費者採購次數或金額的變化自動進行等級歸類與服務調整。並且，定期檢視顧客的需求與交易狀況變動，以調整行銷策略之執行方案。

伍、結論

本研究的主要貢獻如下：一、提出一個以顧客價值作為分群基礎之完整資料庫行銷流程與架構；二、混合對顧客之群集分析與對購買產品之序列型樣分析，做為行銷策略的考量，不侷限以往只對顧客分群之方式；三、擴充現有顧客價值分析模式，作為顧客群集分析之依據；四、針對各別群集提出行銷策略意涵，並發展不同的執行方案；五、以實際案例說明此資料庫行銷架構之分群模型建構與分析方法，並提供驗證行銷策略方案執行實驗的交易結果，以兩階段各半年之交易金額總數，評估行銷方案策略之執行成效。

本研究提出折扣指標，使得分群結果更具有解釋消費特質的意義，因此，未來研究在顧客分群或產品薦購的研究上，可將顧客價值趨勢與折扣指標一併納入考量。本文所述之行銷策略發展流程和資料庫行銷架構，包含詳細的執行細節和分群判別說明，以及策略成效的實驗法，均可作為企業與學術在資料庫行銷設計時之參考。另外，本研究顯示，RFM 指標比顧客的基本屬性與社會背景更能有效地區隔顧客，此結果也印證了先前文獻的論述。

一、研究限制與未來研究

本研究有下列幾點不足與限制：一、本個案中流失顧客之百分比很高。雖然文獻中並未指出，各類價值顧客合理的佔有率，但是在本個案中流失顧客比率超過百分之八十，不是很一般，可能原因是，本個案之資料庫所含之資料，大多為兩年內的新顧客之交易資料，這些新顧客的交易資料少，F 值相對較低，會被歸類為流失顧客。二、本研究之顧客資料庫的資料沒有記錄顧客是企業或散戶，所以不考量企業顧客是否會造成離群值。三、此研究藉由分群法將顧客分為忠誠、機會、消極、流失等群，並藉由專家討論依顧客的特質或需求擬定策略，但並未致力於建立策略擬定之知識本體與規則，未來可考慮建立此知識本體以利專家系統擬定策略。四、本研究以資料中近兩年之交易資料找出序列樣式，其信心度落在 14%~54% 之間，相較一般常用於實驗比較的 20%~60%，略為偏低。由於忠實顧客群的人數相對較少，若將門檻設太高，可能會忽略掉對忠實顧客群有意義的型樣。此研究探究之序列樣式，並未以分群之顧客的交易做區隔，未來可探討加上分群，以檢視是否能更有效地反映出群集顧客交易的序列樣式。

參考文獻

- 張心馨、蔡獻富 (2004)，『以 Data Mining 技術結合 SOM 和 K-Mean 的消費者分群方法於顧客關係管理和績效獲利性評估之實證研究』，*資訊管理學報*，第十一卷，第四期，頁 161-203。
- 葉怡成、楊錦章、連立川 (2008)，『RFMTC 行銷模式之理論解析』，*電子商務學報*，第十卷，第三期，頁 625-642。
- Albrecht, K. (1994), 'Customer Value', *Executive Excellence*, Vol. 11, No. 3, pp.14-15.
- Andrews, R.L., Brusco, M.J. and Currim, I.S. (2010), 'Amalgamation of partitions from multiple segmentation bases: a comparison of non-model-based and model-based methods', *European Journal of Operational Research*, Vol. 201, No. 2, pp. 608-618.
- Bauer, C.L. (1998), 'A direct mail customer purchase model', *Journal of Direct Marketing*, Vol. 2, No. 3, pp. 16-24.
- Bult, J.R. and Wansbeek, T. (1995), 'Optimal selection for direct mail', *Marketing Science*, Vol. 14, No. 4, pp. 378-381.
- Chen, Y., Zhang, G.Z., Hu, D. and Fu, C. (2007), 'Customer segmentation based on survival character', *Journal of Intelligent Manufacturing*, Vol. 18, No. 4, pp. 513-517.
- Colombo, R., Jiang, W. (1999), 'A stochastic RFM model', *Journal of Interactive*

- Marketing*, Vol. 13, No. 3, pp. 2-12.
- Fayyad, U. (1996), 'The KDD process for extracting useful knowledge from volumes of data', *Communications of the ACM*, Vol. 39, No. 11, pp. 27-34.
- Forgy, E.F. (1965), 'Cluster analysis of multivariate data: efficiency versus interpretability of classifications', *Biometrics*, Vol. 21, pp. 768-769.
- Goodman, J. (1992), 'Leveraging the customer database to your competitive advantage', *Direct Marketing*, Vol. 55, No. 8, pp. 26-27.
- Gupta, S., Lehmann, D.R. and Jennifer, A.S. (2004), 'Valuing customers', *Journal of Marketing Research*, Vol. 41, No. 1, pp. 7-18.
- Han, J. and Han, D. (2001), 'A framework for analyzing customer value of internet business', *The Journal of Information Technology Theory and Application*, Vol. 2, No. 5, pp. 25-38.
- Hartigan, J.A. and Wong, M.A. (1979), 'A K-means clustering algorithm', *Applied Statistics*, Vol. 28, No. 1, pp. 100-108.
- Hughes, A.M. (2006), *Strategic Database Marketing* (3rd Ed.). McGraw-Hill Enterprises Inc, New York, NY.
- Irvin, S. (1994), 'Using lifetime value analysis for selecting new customers', *Credit World*, Vol. 82, No 3, pp. 37-40.
- Jen, L. and Wang, S-J. (1998), 'Incorporating heterogeneity in customer valuation: an empirical study of health care direct marketing in Taiwan', *International Journal of Operations Quantitative Management*, Vol. 4, No. 3, pp. 217-228.
- Jiang, T. and Tuzhilin, A. (2006), 'Segmenting customers from population to individuals: does 1-to-1 keep your customers forever?', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, No. 10, pp. 1297-1311.
- Kleissner, C. (1998), 'Data mining for the enterprise', *Proceedings of the 31st Annual Hawaii International Conference On System Sciences*, Kohala, Hawaii, January 6-9, Vol. 7, pp. 295-304.
- Kotler, P. and Keller, K.L. (2008), *Marketing Management* (13th Ed.). Prentice Hall Inc., NJ.
- Kumar, V. and Rajan, B. (2009), 'Profitable customer management: measuring and maximizing customer lifetime value', *Management Accounting Quarterly*, Vol. 10, No. 3, pp. 1-18.
- Lauritzen, S.L. (1995), 'The EM algorithm for graphical association models with missing data', *Computational Statistics and Data Analysis*, Vol. 19, No. 1, pp. 191-201.

- Liu, D.R. and Shih, Y.Y. (2005a), 'Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences', *Journal of Systems and Software*, Vol. 77, No. 2, pp. 181-191.
- Liu, D.R. and Shih, Y.Y. (2005b), 'Integrating AHP and data mining for product recommendation based on customer lifetime value', *Information and Management*, Vol. 42, No. 3, pp. 387-400.
- Malthouse, E.C. and Blattberg, R.C. (2004), 'Can we predict customer lifetime value?', *Journal of Interactive Marketing*, Vol. 19, No. 1, pp. 2-16.
- Markic, B. and Tomic, D. (2010), 'Marketing intelligent system for customer segmentation', *Studies in Fuzziness and Soft Computing*, Vol. 258, No. 1, pp. 79-111.
- Paas, L.J. (2009), 'Database marketing practices and opportunities in a newly emerging african market', *The Journal of Database Marketing & Customer Strategy Management*, Vol. 16, No. 2, pp. 1741-2439.
- Peppers, D. and Rogers, M. (1997), 'Don't resist marketing automation', *Journal of Sales and Marketing Management*, Vol. 150, No. 10, pp. 32-33.
- Peterson, L.A., Blattberg, R.C. and Wang, P. (1993), 'Database marketing past, present and future', *Journal of Direct Marketing*, Vol. 7, No. 3, pp. 27-43.
- Pinto, F.M., Marques, A. and Santos, M.F. (2009), 'Ontology-supported database marketing', *Journal of Database Marketing & Customer Strategy Management*, Vol. 16, No. 2, pp. 76-91.
- Reinartz, W. and Kumar, V. (2000), 'On the profitability of long-life customers in a non-contractual setting: an empirical phase and implications for marketing', *Journal of Marketing*, Vol. 64, No. 4, pp. 17-36.
- Roberts, M.L. (1992), 'Expanding the role of the direct marketing database', *Journal of Direct Marketing*, Vol. 2, No. 2, pp. 51-60.
- Shaw, R. and Stone, M. (1990), 'Competitive superiority through database marketing', *Long Range Planning*, Vol. 21, No. 5, pp. 24-40.
- Tang, Z. and MacLennan, J. (2005), *Data Mining with SQL Server 2005*, Wiley Publishing, Inc., Indianapolis, indian.
- Tsai, M., Tsai, C. and Chang, H. (2010), 'The effect of customer value, customer satisfaction, and switching costs on customer loyalty: An Empirical Study of Hypermarkets in Taiwan', *Social Behavior and Personality*, Vol. 38, No. 6, pp. 729-740.
- Verhoef, P.C., Spring, P.N., Hoekstra, J.C. and Leeflang, P.S.H. (2002), 'The

- commercial use of segmentation and predictive modeling technique for database marketing in the Netherlands', *Decision Support Systems*, Vol. 34, No. 1, pp. 471-481.
- Wong, K.W., Fung, C.C., Tamas, G. and Douglas, C. (2004), 'Intelligent data mining and personalization for customer relationship management', *Proceedings of the 8th International Conference on Control, Automation, Robotics and Vision (ICARCV 2004)*, Kunming, China, December 6-9, pp. 1796-1801.
- Woodruff, R.B. (1997), 'Customer value: the next source for competitive advantage', *Journal of the Academy of Marketing Science*, Vol. 25, No. 2, pp. 139-153.
- Zeithaml, V.A. (1988), 'Consumer perceptions of price, quality and value: a means-end model and synthesis of evidence', *Journal of Marketing*, Vol. 52, No. 1, pp. 2-22.

附錄 A

表 A-1：各群集在各指標之分佈概況表

指標值	群集一	群集二	群集三	群集四
Recency 值				
7.0 ~ 217.6	27.86%	49.24%	72.78%	78.26%
217.6 ~ 358.6	18.71%	20.42%	15.56%	17.39%
358.6 ~ 499.6	19.08%	15.87%	6.67%	0.00%
499.6 ~ 730.0	34.35%	14.46%	5.00%	4.3%
Frequency 值				
1.0 ~ 1.5	85.97%	19.21%	9.44%	8.70%
1.5 ~ 2.5	12.13%	28.61%	8.89%	4.30%
2.5 以上	1.90%	52.17%	81.67%	86.96%
Monetary				
56.0 ~ 1799.6	88.22%	0.00%	0.00%	0.00%
1799.6 ~ 3723.5	11.78%	47.83%	0.00%	0.00%
3723.5 以上	0.00%	52.17%	100.00%	100.00%
Trend 值				
-6.2 以下	0.81%	22.14%	30.56%	13.04%
-6.2 ~ -0.4	0.26%	7.48%	11.11%	8.70%
-0.4 ~ 5.4	98.39%	54.80%	23.89%	34.78%
5.4 以上	0.54%	15.57%	34.44%	43.48%
Discount 值				
0.0 ~ 16.6	82.28%	60.57%	55.56%	39.13%
16.6 ~ 39.8	0.44%	18.91%	22.78%	39.13%
39.8 ~ 100.0	17.28%	20.53%	21.67%	21.74%

表 A-2：K-means 之分群特徵值

各群集	各指標特徵值
群集一 6152 人	56<=Money Amount<=2178.0255449275, 1<=Frequency<=2.0581979667, -8.27326211<=Trend<=8.0795434516, Recency >730, Discount<=0
群集二 989 人	Money Amount >6754.6120675174, Frequency >5.863684854, Trend<=-40.122148835, 7<=Recency<=625.1769222321, 0<=Discount<=77.8496889195
群集三 180 人	Money Amount >18172.5603400087, Frequency >13.7715528278, 7<=Recency<=472.8765774153, Trend >39.4188862967, 0<=Discount<=70.2383294994
群集四 23 人	Money Amount >57855.2284300632, Frequency >22, 7<=Recency<=381.15738532, Trend >32.414390157, 0<=Discount<=83.9776125058

表 A-3：K-means 分群之決策樹分類矩陣

實際預測	群集二	群集三	群集四	群集一
群集二	984	5	0	0
群集三	0	165	0	0
群集四	0	10	23	0
群集一	5	0	0	6152

表 A-4：EM 分群之決策樹分類矩陣

實際預測	群集二	群集三	群集四	群集一
群集二	1164	0	23	0
群集三	0	610	34	202
群集四	1	0	757	0
群集一	0	295	0	4246

表 A-5：各群集顧客之人口統計分析

		流失顧客	消極顧客	機會顧客	忠誠顧客	總數
性別	男	4,531 (73.65%)	783 (79.17%)	149 (82.78%)	17 (73.91%)	5,480 (74.62%)
	女	1,621 (26.35%)	206 (20.83%)	31 (17.22%)	6 (26.09%)	1,864 (25.38%)
年齡	16~25	1,031 (16.76%)	127 (12.84%)	16 (8.89%)	6 (26.09%)	1,180 (16.07%)
	26~35	3,035 (49.33%)	466 (47.12%)	79 (43.89%)	7 (30.43%)	3,587 (48.84%)
	36~45	1,597 (25.96%)	310 (31.34%)	63 (35%)	7 (30.43%)	1,977 (26.92%)
	46~55	430 (6.99%)	77 (7.79%)	20 (11.11%)	3 (13.04%)	530 (7.22%)
	56~60	59 (0.96%)	9 (0.91%)	2 (1.11%)	0 (0%)	70 (0.95%)
教育	國小	45 (0.73%)	5 (0.51%)	2 (1.11%)	0 (0%)	52 (0.71%)
	國中	51 (0.83%)	6 (0.61%)	1 (0.56%)	0 (0%)	58 (0.79%)
	高中職	629 (19.22%)	78 (7.89%)	11 (6.11%)	1 (4.35%)	719 (9.79%)
	專科	1,560 (25.36%)	241 (24.37%)	47 (26.11%)	2 (8.70%)	1,850 (25.19%)
	大學	2,879 (46.80%)	490 (49.54%)	83 (46.11%)	16 (69.57%)	3,468 (47.22%)
	碩博士	988 (16.06%)	169 (17.09%)	36 (0.2%)	4 (17.39%)	1,197 (16.30%)
職業	其他	2,714 (44.12%)	374 (37.82%)	72 (40%)	4 (17.39%)	3,164 (43.08%)
	學生	860 (13.98%)	138 (13.95%)	--	7 (30.43%)	1,016 (13.83%)
	教師	--	--	17 (9.44%)	--	312 (4.25%)
	職員	847 (13.77%)	126 (12.74%)	24 (13.33%)	3 (13.04%)	1,000 (13.62%)
	工程師	879 (14.29%)	186 (18.81%)	27 (15%)	2 (8.70%)	1,094 (14.90%)
地區	台北縣	1,269 (20.63%)	205 (20.73%)	31 (17.22%)	6 (26.09%)	1,511 (20.57%)
	台北市	1,117 (18.16%)	177 (17.90%)	34 (18.89%)	5 (21.74%)	1,333 (18.15%)
	桃園縣	478 (7.77%)	64 (6.47%)	12 (6.67%)	--	555 (7.56%)
	高雄市	349 (5.67%)	48 (4.85%)	14 (7.78%)	3 (13.04%)	411 (5.60%)
	台中縣	285 (4.63%)	50 (5.06%)	--	--	337 (4.59%)
各群集人數		6,152	989	180	23	7344

註：職業和地區只列數目前五大，表中以--代表該群集內前五名不包含此項