

# 臺鐵會員制再深化—利用資料探勘技術 訂定忠誠計畫規則<sup>1</sup>

## MINING TRA'S TRANSACTION DATA FOR LOYALTY PROGRAM RULES

陳民祐 Min-Yu Chen<sup>2</sup>

王建富 Jiana-Fu Wang<sup>3</sup>

(101年10月1日收稿，102年7月24日第一次修改，102年8月12日第二次修改，  
102年9月15日定稿)

### 摘要

積點制的會員回饋計畫，或所謂的忠誠計畫，已廣泛地被許多行業用來維繫與顧客間的關係，甚至作為刺激顧客消費與公司獲利的工具。由於會員系統需要資源的持續投入，因此有必要針對會員的加入進行篩選，甚至分級提供不同程度的服務，以激勵及鎖定會員的消費。本研究利用臺鐵路電話及網路訂票的巨量資料庫，以RFM及延伸變數進行集群分析與建立決策樹模型，並輔以忠誠度矩陣加以評估，結果發現利用三個月訂票資料所建置的顧客價值分類準則即可達到94%的預測準確率，同時本研究並運用前述準則篩選出高價值旅客及具潛力的旅客作為臺鐵路邀請加入會員系統的對象，也建議藉由降低積點兌換門檻、提供多樣化的獎品以及採取層級式的會員架構等措施來強化會員積點制度的功效。

**關鍵詞：**忠誠計畫；資料探勘；車票預約；鐵路；忠誠矩陣

- 
1. 本研究蒙國科會99年度專題研究計畫(NSC 99-2410-H-005-046)補助，並經兩位匿名審查委員提供寶貴的修正意見，特此致謝。
  2. 國立中興大學行銷學系碩士。
  3. 國立中興大學行銷學系助理教授(聯絡地址:40227 臺中市國光路250號中興大學行銷學系;電話:04-22840392 轉783; E-mail: jfwang@nchu.edu.tw)，王建富為通訊作者。

## ABSTRACT

*Point-based customer loyalty program has been extensively adopted in many industries to maintain customer relationships, even to stimulate repeat purchases from customers and to obtain more profits for companies. Due to the need to continuously invest resources in loyalty programs, companies should only allow profitable customers to join the programs. This study evaluates the ticket reservation data of Taiwan Railways Administration with RFM and extended variables using clustering and decision tree techniques and loyalty matrix concepts to identify customer values. Through this research, we are able to provide 94% classification accuracy on our decision tree model employing three-month ticket reservation data. Also, high-value and potential high-value customers are identified via the classification rules for member-recruiting. In the end, lowering the thresholds of redeeming points, offering diversified rewards and using tier membership structure are suggested to enhance the functions of the loyalty program.*

**Key Words:** *Loyalty program; Data mining; Ticket reservation; Railway; Loyalty matrix*

## 一、前言

臺灣鐵路的歷史悠久，長久以來為臺灣西部的重要交通運具之一，而其運輸角色在預期會受到高速鐵路通車的影響下，於九零年代初期即在政策導引下開始轉型為以中短程接駁轉運服務為主、長程客運為輔的經營型態<sup>[1]</sup>，因此雖然臺灣鐵路管理局（以下簡稱臺鐵局）的客運收入於高速鐵路通車（2007年）初期呈現下滑趨勢，但是在經營策略的調整之後，自2010年起已呈現緩步上升的趨勢，目前並已回復甚至高於高速鐵路通車前的客運收入，另外從近幾年來每人平均運程公里數已下降並維持在47公里左右的情形來看（詳表1），旅客們對於使用臺灣鐵路於中短程旅行的習慣已漸趨平穩。

雖然臺灣鐵路於外在競爭的情勢已較穩定，但是在內部仍然有平日客座利用率不高（2006年至2010年間平均約59.34%）以及民眾屢屢抱怨訂不到周末假日車票的情形，尤其訂票過程中有許多預訂車票後卻取消訂票或逾期未取票的情形（以本研究2010年8月至10月間西部幹線之訂票資料統計，其取票率僅50.83%），使得對傳統鐵路依賴性較強的忠實顧客常常因為前述佔票行為而無法順利訂到座位。據此，臺鐵局自2011年3月30日啟動了會員制的服務，希望藉由積點制度來鼓勵會員增加消費，並嘗試著從運具間的相互競爭走向顧客關係的經營。

依據「臺鐵會員服務系統會員服務條款」暨「臺鐵會員酬賓兌換辦法」之規定，只要是中華民國的民眾皆可申請成為會員，會員可將80.1公里以上的購票消費以每30元為一點的方式集點，於兩年內依累積點數兌換免費車票<sup>[3]</sup>，此外，臺鐵局並宣示將協助訂不到

票的會員媒合未取票或退票的車票。

表 1 臺鐵局客運經營概況

年	搭乘人次 (億)	延人公里 (億)	收入 (億元)	每人次平均運程 (公里)
2004	1.6847	93.5892	153.6388	55.6
2005	1.6956	94.9967	156.3075	56
2006	1.6899	93.3917	152.7135	55.3
2007	1.6969	89.3739	145.6118	52.7
2008	1.7866	87.1778	141.2076	48.8
2009	1.7937	83.8686	134.5868	46.8
2010	1.8976	89.9841	146.2618	47.4
2011	2.0583	97.1961	155.7684	47.2
2012	2.2030	101.8555	161.1874	46.2

資料來源：臺灣鐵路管理局<sup>[2]</sup>。

臺鐵局的會員積點為忠誠計畫 (loyalty program) 實施的方式之一，源起於 1980 年代航空業的飛行常客計畫 (frequent flyer program)，其主要目的在於藉由獎賞來鼓勵顧客忠誠的行為，以獲得穩定的顧客關係並增加公司獲利<sup>[4]</sup>，其應用的業種除了原先的航空業之外，近年來已逐漸擴展到金融業、雜貨業、電訊通信業、租賃汽車業、旅館業、書籍零售業、加油站等<sup>[5,6]</sup>。

雖然歷年來實施忠誠計畫的成功案例很多，但也不乏因不符成本效益而黯然收場的個案，例如英國的連鎖超市 Safeway 在 2000 年初即因為不敷成本而放棄了原有的忠誠計畫<sup>[7]</sup>，美國的連鎖速食 Subway 放棄了它的 Sub Club Card 計畫，澳洲的 Cole 超市也結束了它的會員折扣計畫<sup>[8]</sup>。另外 Dowling<sup>[9]</sup> 則質疑忠誠計畫並不必然能培養顧客的忠誠，也不一定符合成本效益，而 Ferguson 與 Hlavinka<sup>[10]</sup> 則在 2006 年於美國的普查中發現平均每個美國家戶竟參與了高達 12 個忠誠計畫，而其中屬於活動中會員的卻僅有 4.7 個，意謂著其它的 7.3 個會員資格平常是沒有在利用的，因此實施的企業並無法藉此與顧客建立穩定的關係。

企業期望藉由忠誠計畫的獎賞來改變顧客的行為，而這樣的行為轉變如同廣告、促銷一樣需要成本的投入，除了獎賞提供所需要的成本之外，會員電腦資料庫的紀錄與維護亦需要人力與硬體上持續性的投入，因此企業會斟酌獎賞的額度與實施的範圍；然而，計畫的實施並不保證能成功地增加顧客的消費，獎賞額度不夠吸引人以及兌換手續繁瑣或所需時程過長等都會造成會員的挫折感，所以如何設計忠誠計畫的架構以達成原訂目的遂成為企業關心的焦點。

由於積點制的會員系統需要資源持續的投入，因此此種系統首重於讓獲利性高的會員加入，實務上，一般企業僅能運用最低消費要求或者繳交會員費等方式來篩選<sup>[4]</sup>。然而其

實許多企業本身即儲存有大量的顧客歷史購物資料，可以運用科學方法制定行銷規則並主動找出值得加強關係的顧客群，此概念在最近一年來巨量資料 (big data) 的處理與運用上已成為探討的焦點<sup>[11-13]</sup>，使得如此的應用在臺鐵局顧客關係加強方面顯得益形重要。

本研究希望運用臺鐵局顧客訂票資料庫的紀錄，藉由資料探勘的集群與分類技術，建立一套識別顧客價值的程序與方法，並從辨識結果中尋找適合制定忠誠計畫之規則。進一步地說，本研究關心的是臺鐵局實施的會員積點制中那些顧客是屬於應積極招募的，至於其他獲利性不佳或者無法藉由積點計畫鼓勵其消費行為的顧客，如果任由其加入會員系統，恐怕只是徒增系統的資源負擔。其次，我們該如何辨識這些顧客群？在現有的資訊下是否有能力做出這樣的判斷？另外，針對這些價值較高的顧客，除了以積點兌換車票之外，也探討提供其它服務來提高他們忠誠程度的作法。

為了回答以上的問題，本研究以臺鐵局西部幹線的旅客為研究範圍，首先運用資料探勘中的集群分析技術將鐵路網路與電話訂票系統資料庫中的顧客分群，初步找出不同價值群組的顧客，然後利用 Reinartz 與 Kumar<sup>[14]</sup>的忠誠度矩陣 (loyalty matrix) 評估出那些群組是高價值旅客，那些群組是有潛力的顧客，另外那些是不值得期待與投資的，藉以設定值得招募的會員範圍為何；前一階段的分群乃是基於顧客過去購票歷史所得的結果，由於網路與電話訂票的顧客消費資料不斷地在更新，亦即顧客的價值會隨著時間的演變而有所調整，也會有新成員的加入，所以建議再運用決策樹的技術找出不同群組顧客間的分類規則，並從這些規則中建議如何判斷某一顧客是否值得納入會員，會員的階層如何分級，以及進一步的會員服務項目等。

本文共分為五節論述，本節說明研究背景與動機，第二節除回顧忠誠計畫之類型與設計之相關研究外，並探討資料探勘中集群分析與決策樹演算法之相關文獻，第三節介紹本研究資料來源以及變數的整理與擷取，第四節則就集群分析與決策樹演算法的實證過程與結果加以說明，最後於第五節針對研究結果進行結論與建議，以提供臺鐵局未來修正其忠誠計畫之參考，提升其營運績效與顧客滿意。

## 二、文獻回顧

### 2.1 忠誠計畫

如前言所述，忠誠計畫已經廣泛地被許多行業所運用，根據 Liu<sup>[15]</sup> 與 Dorotic 等人<sup>[16]</sup>的定義，它是一種由零售商甚至製造商所提供，透過持續性的獎賞方式來鼓勵顧客重複性購買的誘因計畫。而透過此類計畫，企業還可以獲得更加了解顧客的偏好與習性、藉由轉換成本的提高而增加顧客的保有率、提供差異化的服務或商品、改善獲利性等好處<sup>[17]</sup>。其計畫特性可歸納為下列五大特色<sup>[16]</sup>：

1. 培養忠誠度，期望透過計畫的施行，以鼓勵顧客更常購買與增加購買金額。

2. 具有結構性，顧客必須加入會員方能享有兌換獎勵的權利，其意謂著零售商可以擁有會員的資料而與會員做更好的互動。
3. 廠商與顧客雙方長期的投入。
4. 給予會員獎勵。
5. 必須針對會員進行持續性的行銷活動。

忠誠計畫的類型可依據會員收到獎勵的方式以及參與廠商的數量等兩種方式做區分，從表 2 的分類可知臺鐵局會員制度係屬於 A3 暨 B1 的型式，在 A3 型忠誠計畫中，計畫主導廠商必須持續性地更新維護會員的積點資料，而 A4 型則是更進一步地分析會員歷史購物資料，提供客製化的獎賞及行銷方案。

表 2 忠誠計畫的類型

分類方式	忠誠計畫的類型	文獻來源
A.會員收到獎勵的方式	A1.所有顧客均能加入會員，每一位會員在結帳時均能立即得到相同折扣。 A2.所有顧客均能加入會員，會員購買 N 件產品時可得到一件免費。 A3.會員憑藉累積的購買數額可以得到回饋或點數。 A4.廠商依據會員購買歷史資料提供不同獎賞商品以及廣告郵件。	Berman <sup>[6]</sup>
B.參與廠商的數量	B1.單一公司的計畫，目的在於增加重要顧客的忠誠度。 B2.由單一公司開始並逐步擴大至數個公司夥伴，其目的為提高計畫的營收。 B3.由數家公司發展的跨公司計畫，其目的在增加交叉銷售。	Kadar 與 Kotanko <sup>[7]</sup>

企業在設計忠誠計畫時需考慮下列事項<sup>[6,8,14,16]</sup>：

1. 設定計畫目標：如提高營收、利潤、顧客忠誠度或蒐集更多的消費者資訊。
2. 決定會員的資格：以提高營收為目的的計畫可以放寬會員加入資格，但預算會相對較高；以利潤為目的的計畫，其會員宜注重顧客獲利性，可採邀請加入、要求最小購買金額或收取會員費等方式來篩選顧客。
3. 選擇獎賞：採用 A3 及 A4 型計畫時，需考慮點數與購買金額的轉換比率、可兌換的獎賞選擇性、是否對會員再細分層級以及積點兌換的難易度。透過層級式的架構則可以進一步鼓勵高層級會員增加消費，例如飯店金卡會員可以免費提供房間升等，鑽石卡會員更進一步地保證隨時都能訂到房間；但若於制度設計時讓會員累積集點過於困難時，則可能導致會員中途放棄。
4. 是否與其它廠商合作：若有其它合作夥伴，對廠商而言有機會增加會員數量、分擔會員系統運作經費、促成交叉銷售以及夥伴成員形象的溢出效果；對會員則有集點速度較快以及兌換項目較多元等優點。

其中在會員選擇的評估方面，一般文獻僅論及需選擇價值較高的顧客成為會員<sup>[6-7,10]</sup>，對於衡量的標準為何則較少提及，筆者僅發現 Marcus<sup>[18]</sup> 及 Reinartz 與 Kumar<sup>[14]</sup> 曾分別提出顧客價值矩陣與忠誠矩陣的觀念，他們各利用兩類變數將所有顧客區隔為四個象限。其中 Reinartz 與 Kumar<sup>[14]</sup> 的忠誠矩陣的方式係衡量顧客在忠誠度與獲利性兩方面的關係，以方便決策那些顧客應該加以重視，而那些應該予以忽略。他們認為單純使用 RFM 模型（詳 2.3 節）仍有不夠精確的缺點，因此以獲利性 (profitability) 與關係長久程度 (longevity) 來區別顧客，經實證發現可以將顧客分為 4 種類型：

1. 「真正的朋友」：具高獲利性與長期忠誠的顧客，為最有價值的顧客。
2. 「蝴蝶」：具高獲利潛力但為短期忠誠的顧客，這些顧客喜歡到處尋找物超所值的商品，並且避免和單一商家建立穩定的關係。
3. 「陌生人」：帶來很少獲利且沒有忠誠度的顧客。
4. 「戀棧」：雖然是屬於長期忠誠的顧客，但是他們所能帶來的獲利性不高，是屬於較難處理的一群，必須要進一步探究他們是屬於在該類商品花費很少的人，還是他們在該商家只佔其很少比例的消費。

## 2.2 集群分析

集群分析是將大量樣本基於分析時所需考慮之樣本特質之相似性，而予以分類成多個互斥小群組的方法，其方法可分為三大類，分別為階層式 (hierarchical method)、分割法 (partitioning method) 以及兩階段集群法 (two-step clustering)。企業在實務上可以藉此分析找出不同的顧客區隔 (customer segment) 而施以不同的服務或行銷方法<sup>[19]</sup>。本研究在集群分析方面係運用分割法中常用的 K-means 法以及結合自我組織映射神經網路的兩階段集群法，以下茲分別介紹。

### 2.2.1 K-means 法

K-means 演算法已經有超過 50 年的發展歷史，其方式是將各觀察樣本分到各群中，以使各群內樣本與該群中心點的總平方誤差 (sum of the squared error) 最小化。進行 K-means 時的三大步驟主要為群數 K 的決定、初始群心的選擇以及距離的運算<sup>[20]</sup>。

一般來說，初始群心的選擇對於分群結果的品質有相當程度的影響，若是初始群心（亦稱為種子點）的選擇不當，分類的結果就可能不甚理想。而初始群心的決定方法不只一種，隨不同電腦統計軟體的使用也可能會採取不同的決定方法。在運用 K-means 時較重要的課題如下<sup>[21,22]</sup>：

#### 1. 集群數目的決定

K-means 需要研究者事先決定集群數，以行銷領域來說，通常不希望集群數太多，雖然集群數目的增加理論上更能針對各集群制定行銷策略，集群間的異質性或許很高，但是可能反而會增加業者的負擔—因為必須針對很多的集群去進行行銷策略的制定。集群數目

太少，又可能造成集群間的異質性不夠明顯，造成在制定行銷策略或是分析集群的消費行為不易分析的困擾。

## 2. 如何設定初始集群中心

集群中心的設定方法有許多種，選擇該用哪一種方法決定初始群心也常常是每一位研究者必須處理的問題。

## 3. 對集群幾何特性的了解

集群完畢之後，對於集群的解釋與分析才是重心所在。而研究者對於解釋結果的能力，也決定了集群效果的優劣。若研究者不知該如何解釋集群出來的結果，那麼集群出來的結果並不代表任何涵意，僅不過又是由另一堆數字所堆砌而成的資料集。

K-means 雖然有快速且方便的優點，但是同時也有些限制，例如 K-means 僅以距離為分群基礎。而隨著演算法的改良與發展，有越來越多的研究者提出不同的方法，例如 Fuzzy C-Means、Bisecting K-means、Kd-tree、X-means、K-medoid 及 Kernel K-means 等<sup>[20]</sup>。

### 2.2.2 Kohonen 自我組織映射神經網路

隨著電腦運算能力的增加，近來類神經網路也開始被應用在集群分析。Kohonen self-organized mapping network (Kohonen 自我組織映射神經網路，以下簡稱此方法為 KSOM) 採用競爭式學習演算法，輸出層的神經元會依照研究者輸入的資料特徵，以拓樸結構 (topological structure) 的方式展現在輸出空間中，又由於其所產生之拓樸結構代表不同輸入資料特徵的分類，因此又稱作自我組織映射神經網路<sup>[23]</sup>。其設計理念是模仿大腦中細胞對於記憶處理的概念而成的，人類大腦對於各種感官都有專屬於各自的處理區塊，若將多元的訊息傳送至反應神經，容易造成大腦混亂的情形，導致個體無法順利解讀並對其做出適當反應，所以人體將高維度資料 (high-dimensional data) 轉譯為二維度資料 (two-dimensional data) 或三維度的資料，甚至轉換為單維度資料的圖像訊息供大腦處理並輸出，讓使用者能更容易理解資料中隱含的訊息，做出適當反應。

過去學者利用類神經網路進行集群分析的資料量通常較小，也有結合 K-means 進行兩階段集群的研究，其結果是可行且效果不錯<sup>[24]</sup>。本研究在集群分析的部分除了以 K-means 進行外，也同時想要藉著使用臺鐵局的巨量資料來驗證類神經網路搭配 K-means 的兩階段演算法是否能提供精準的集群結果。

### 2.2.3 兩階段集群

集群有很多種方法，在運算集群分析時，雖然使用單一方法相對單純，然而用兩階段法有時不僅不會增加太多負擔，反而可以增加集群的效益。過去的研究，如 Van Laerhoven<sup>[25]</sup> 運用類神經搭配 K-means 的方法於判斷手機自動進入靜音模式的情境，林祥生與劉益豪<sup>[26]</sup> 與 Aggelis<sup>[27]</sup> 也分別運用 RFM 搭配 K-means 的做法於國際線航空旅客的線上購票行為以及金融業的客戶破產風險判斷，這些說明兩階段集群方法是個有發展性的方向。使

用兩階段法主要有兩個優點：可以降低運算成本與減少雜訊<sup>[28]</sup>。

### 1. 降低運算成本

若是以自我組織映射神經網路為一個中介的步驟，整個運算的複雜度可以由原來進行 K-means 演算法時的  $(\sum_{K=2}^{C_{\max}} NK)$  降低為  $(NM + \sum_K MK)$ ，其中  $C_{\max}$  為預先選取之最大集群數， $N$  為樣本數， $K$  為集群數， $M$  則為神經網路之數目。據 Vesanto 與 Alhoniemi<sup>[28]</sup> 的估算，當樣本點為 10,000 個時，大約可以降低六倍的運算量。以目前鐵路訂票系統每個月約一百萬筆的交易資料而言，此項降低運算成本的優點在鐵路訂票資料庫中的運用上顯得更形重要。

### 2. 降低雜訊

除了運算量的減少，雜訊的降低也是一個很重要的優點。在第一階段的分群結果是以該群鄰近樣本點的平均值作為代表，對於原始資料中變異很大時所可能產生的影響較不敏感，此特性在資料龐大且可能有極端值時特別有用。

## 2.3 RFM 模型

RFM 是應用在辨識行為分析 (behavioral analysis) 中的技術，該方法為 Hughes 於 1994 年所提出，他發現在客戶資料庫中有 3 個良好的指標可以用來判斷顧客的價值，分別為：最近一次消費 (recency)、消費頻率 (frequency)、消費金額 (monetary)<sup>[29]</sup>。RFM 模型在某種層面和 80/20 法則有點類似，也就是公司 80% 的收入是來自於公司全部 20% 的客戶。在預算不夠多，所能投入資源 (如寄發廣告) 給消費者的數量有限之情形下，選擇投資的客戶目標就顯得很重要，藉由這種方法節省下來的開銷將會相當可觀，而且成功率也比較高。

RFM 的基本模型認為 3 種變數都是同等重要，每一位顧客都分別以該 3 個變數依大到小、多到少的次序加以排列，排序完畢後會將顧客 3 個構面 (最近一次消費、消費頻率、消費金額) 各切割為 5 等份。排在最前頭的 20% 者給予 5 分、次之的給予 4 分，依此類推、最低為 1 分。分數越高，代表該顧客在這變數的重要性最高，而每位顧客會有 3 個分數，從 555 到 111，有 125 種情形，也可以視為 125 群。

爾後也有學者提出依照產業性質對 3 個構面給予不同權重的做法，此外，也有學者增加更多變數進行分析。目前該模型之應用除了最初的郵購業外，也開始運用於銀行業、航空業、飲食業、醫療業等領域。本研究中係針對鐵路訂票之特質，採取在原有的 RFM 變數之外延伸增加其他變數的方式進行。

## 2.4 決策樹

決策樹演算法是一種分類模型，其將事物或物件的分類透過文字化敘述及圖像化的呈現找出分類規則，以應用於對分類變數的預測或解釋。相對於其他資料探勘的演算法，決策樹較易於理解和實現，研究者在透過解釋後比較有能力去理解決策樹所表達的意義與內容，不像統計分析對於多數人來說並不容易了解其內涵。同時對於數據的準備往往不需要

太過注意，另外，它能夠同時處理數據型和名目型屬性的分類判斷也是一大優點，更重要的是在相對短的時間內可以對大型數據資料庫做出效果良好的結果。

分類的目的主要在於解析影響資料分類的因素為何，可以藉由資料歸納整理出分類的規則；第二個目的則是藉由分類規則來預測資料所屬的類別。一旦我們以充足的資料量建立分類模型，可預測未來其他資料所屬的類型。以下介紹常用到的 CHAID (Chi-square automatic interaction detector) 與 CART (classification and regression tree) 兩種決策樹演算法。

CHAID 應用於資料庫區隔分析中已有許多年。在某種層面來說，CHAID 跟 RFM 有些類似<sup>[30]</sup>，因為他們都利用資料庫中的顧客紀錄劃分群體，但他們最大的差別在於 RFM 可以直接進行劃分群體的行為，而 CHAID 必須在執行過前置步驟，如：寄發直銷郵件並記錄他們的反應之後，才能去進行演算。相較於 RFM，CHAID 運用統計的技術，並透過套裝軟體的幫助，使得一般使用者也可以很容易的快速使用 CHAID。

CART 由 Breiman 於 1984 年提出，是建構決策樹時常用的演算法之一。CART 適用於連續型變數與非連續型變數，隨著輸入的變數不同，亦可稱為迴歸樹 (連續型變數) 與決策樹 (非連續型變數)。CART 藉由從決策樹的根部反覆運算建立二元分支直到設定的節點層級數後，方停止運算。與 CHAID 最大的不同就在於每個節點僅有兩個分支，不會有例外的情形發生。通常來說，CART 輸出模型並不需要太長的訓練時間，且其輸出欄位可以是數值型的資料或類別型的資料，對於資料格式與尺度亦相當有彈性，對研究者來說較為方便<sup>[31-32]</sup>。

CHAID 與 CART 最大的差別在於後者會嘗試先過度套用資料，爾後再加以修剪分支；CHAID 則是在過度套用資料前就盡量讓決策樹停止蔓生擴大。同時 CHAID 需限制使用於類別或名目型的變數，若是研究者原本有連續型的變數要進行演算，就必須事先轉換成幾個區段範圍<sup>[32]</sup>，例如將連續型的收入變數重新劃分為低、中、高收入等名目型變數。

## 2.5 集群指標

前面的回顧曾提到集群分析時究竟該分為幾群才是最佳解常常困擾著研究人員，為此有學者使用集群指標 Davies-Bouldin index 來幫助判斷，此外也可利用  $R^2$ 、SPRSQ 以及均方根標準差等幾種比較常見的統計指標來判斷<sup>[33,34]</sup>，本研究將使用後三者的指標來評斷 K-means 集群法及 KSOM 搭配 K-means 二階段集群法兩者之優劣，茲介紹該 3 個指標如下：

### 1. 判定係數平方值 ( $R^2$ )

$R^2$ ，代表的是群間異質性。運算式如下：

$$SS_T = SS_B + SS_W$$

其中  $SS_B$  是組間變異量， $SS_W$  是組內變異量，而：

$$R^2 = \frac{SS_B}{SS_T}$$

隨著觀察群的合併，集群數量會減少，使得集群內的異質性會越來越高，也因此組內變異量  $SS_W$  也會隨之增加。而當  $SS_T$  是一固定的數值時， $SS_W$  和  $SS_B$  的增減關係是相反的，即  $SS_W$  越大，則  $SS_B$  越小。在這種情況下，集群間的變異量 ( $SS_B$ ) 就會越小。由於集群分析最初的目的就是使集群內的變異比較小，而集群間的變異較大，所以一般來說  $R^2$  需要比較大。若是由於觀察體的合併，導致  $R^2$  突然變小，代表應該停止繼續合併。

### 2. 半偏定係數平方值 (Semipartial R-squared ; SPRSQ)

此值代表著進入新的集群數量時，比上一個階段所增加的異質性 (同時也意味著損失的同質性)。通常值越小越好，代表損失的同質性比較少。

### 3. 均方根標準差 (root-mean-square standard deviation)

此指標是集群中所有觀察值的聯合標準差，代表集群各變數的均齊性，值越小越好，公式如下：

$$\text{Root - mean - square standard deviation} = \sqrt{\frac{\sum_{i=1}^P S_i^2}{P}}$$

$P$ ：變數個數

$S_i$ ：第  $i$  個變數的各群內標準差之和

過去集群分析最令人頭痛的地方就在於最佳群數的決定，K-means 該分幾群往往是研究者相當主觀的經驗選擇分群數目。透過這 3 個指標，可以比較究竟該選多少集群與比較不同分群方法間的差異。

## 三、資料來源與處理

本研究使用的資料範圍包括 2010 年 8 月至 10 月間經由電話及網路預訂起訖點皆在西部幹線車站間 (從基隆至枋寮共 64 站) 之臺鐵局對號車票旅客取訂票資料，該 3 個月間經篩選後於該區段預訂車票的交易紀錄總計有 1,424,796 筆，來自於 332,787 位不同的旅客，經剔除非本國籍的旅客後共有 332,584 位國內旅客的交易資料。

各旅客之原始訂票紀錄包括有旅客代號、訂票時間、車次、車種、起站、迄站、開 (上) 車時間、訂票張數、與取票狀態等 9 種資訊，並基於前述資料另行計算出里程數、金額及取票機率等變數。由於在資料整理期間發現鐵路訂票過程中許多顧客會有中途退票或者逾期未取票的情形發生 (本研究資料庫中平均取票機率僅 50.83%)，因此顧客的價值性難以僅由原始的 RFM 指標評斷，例如某顧客最近一次的往來為自行取消訂票的紀錄，則其是否計入新進程度  $R$  以及訂票頻率  $F$  呢？如果計入，則該顧客與另一位同一天訂票但最後有

取票的顧客該如何區別呢？因此，在本研究中建議加入其它變數一併分析，以釐清在鐵路訂票特殊狀況下的顧客區隔問題。本研究後段將依據顧客之忠誠度與獲利程度進行忠誠度矩陣的分類，因此在集群分析與決策樹分析過程中所使用的變數乃是考量以 RFM 為基礎加以延伸而得，共有下列 6 個：

#### 1. 取票機率 (purchase possibility, PS)

各筆交易資料中的取票狀態可分為已取票、有訂票但自動取消訂票，以及訂票逾期未取票三種。某顧客於計算期間內已取票的次數除以總訂購次數所得之數值即為該顧客之取票機率。

#### 2. 平均取票狀態分數 (average status score, ASS)

該變數類似旅客信用的概念，透過該屬性衡量旅客的信用程度，拿來分辨旅客 3 種行為為特徵。分別為每次訂購後取票者給予計分 5；訂票後自行取消者，給予計分 3；訂票後逾期未取票者則給予計分 1。

平均取票狀態分數表示旅客習慣性的取票行為，透過取票機率及平均取票狀態分數兩項變數，可以分辨出旅客未取票的行為是屬於自行取消或是逾期被取消。平均取票狀態分數越接近 1 者，其習慣逾期未取票的行為對臺鐵局車票預售所產生的干擾影響也越大。

#### 3. 平均金額 (average monetary, M)

這 3 個月中的平均購票金額，除了計算已取票的金額外，其餘未能取票的交易紀錄金額會以零計算，仍計入交易次數，因此平均金額有可能因為訂購後取消或未取票而降低。

#### 4. 總里程 (total mileage, TM)

套用航空業忠程方案的概念，以總累積里程的概念計算。但此部分採取單張的里程，就算消費者訂 3 張，該次里程也不會變成 3 倍，因為本研究認為旅客的忠誠程度並不會因為幫別人代訂而有倍數性的增加。而且以此方式衡量也可避免與平均金額變數過於類似的情形，透過平均金額與總里程的比較，可以辨別出旅客的訂購是偏向於個人的旅行，或是團體式的旅行。

#### 5. 新進程度 (recency, R)

新進程度指的是顧客最近購買的日期距離計算基點日期的天數，本研究以 2010 年 11 月 15 日做為資料處理的基點日期。該變數之數值越大代表越久沒有訂票，顯示與臺鐵局的關係較為淡薄。

#### 6. 頻率 (frequency, F)

旅客在 2010 年 8 月至 10 月共 3 個月內訂票的次數。不論是否完成取票，只要有訂票交易紀錄，皆列入頻率中計算。

從表 3 所列前述 6 個變數的敘述統計資料可發現，在整個資料中，平均每位旅客透過臺鐵局訂票系統訂購的時間點距離基點日期為 53.75 天前，而這些旅客在這 3 個月的期間透過電話及網路訂票的平均次數為 4.28 次，在平均金額及總里程方面也分別花費 285.78 元及搭乘 364.03 公里，對照標準差的數據較大的情形也顯示了各個項目中個別旅客的表現

差異非常大，尤其是平均取票機率只有 50.83%，也使得一些常取票與習慣不取票的旅客在這 6 項變數中的數值差異會拉大，方便於後續的分析中對不同行為模式的旅客加以辨識。

表 3 資料庫中 6 個變數之敘述性統計

	平均數	標準差
新進程度 (R)	53.75 日	27.156 日
頻率 (F)	4.28 次	5.786 次
平均金額 (M)	285.78 元	378.70 元
總里程 (TM)	364.03 公里	539.102 公里
平均取票狀態分數 (ASS)	3.68 分	1.20 分
取票機率 (PS)	50.83%	39.52%

另外為加快並增強 KSOM 模型於類神經網路的學習效果，在 KSOM 運算的過程中特別將原始的輸入值加以正規化 (normalizing)，正規化後數值的範圍設定在 0 與 1 之間。原始數值正規化的計算方式如下：

$$S = \frac{L - L_{min}}{L_{max} - L_{min}}$$

S：正規化的數值

$L_{min}$ ：該資料集中最小值

$L_{max}$ ：該資料集中最大值

在利用資料探勘軟體 SPSS Modeler 之 KSOM 模型運算時需設定參數，其中在矩陣大小方面，係參照 Vesanto 與 Alhoniemi<sup>[28]</sup> 提到兩階段法中的矩陣設定的公式。

$$M = 5\sqrt{N}$$

M：KSOM 的矩陣總數

N：樣本總數

表 4 Kohonen self-organized mapping 參數設定

階段 \ 參數	鄰近距離	初始學習率	迴圈次數
第一階段	3	0.01	1000
第二階段	2	0.1	500

經計算後之矩陣設定為  $53 \times 53$ ，其餘第一階段及第二階段須設定的鄰近距離 (neighborhood)、初始學習率 (initial eta)、迴圈次數 (cycles) 3 個參數之設定如表 4。其中第一階段主要是用來獲得整個資料大致特徵，第二階段是調整階段，用來調整 KSOM，使其完成最終模型的資料特徵。

## 四、實證結果與分析

### 4.1 集群方法之比較與結果

本研究先採用兩種集群方法進行比較後，再擇取結果較優佳者進行後續的程序。第一種方法是常用的 K-means 一階層集群法，第二種方法則屬於兩階層集群法，先進行 KSOM，再進行 K-means 集群。

進行集群分析時究竟該分為幾群才是最佳解往往是令研究者感到最困擾的課題，若要達到最有效的分群，自然分群數目是越多越好，但對於資源相對有限的企業來說，分群太多反會增加困擾。為此本研究採用了 R-Squared、SPRSQ 以及均方根標準差等三種指標對集群結果進行比較。R-Squared 值的範圍為  $0 \leq R^2 \leq 1$ ，表示分群中，群集與群集之間的差異程度，所以其值越大，代表集群內的同質性也比較高。SPRSQ 表示形成新集群時一致性的損失，故 SPRSQ 越小越好。最後一種指標，則是均方根標準差，表示新集群內各變數的均齊性，所以其值是越小越好，若是值太大，就不適合形成該集群，但值的大小並沒有一定的準則<sup>[30-31]</sup>。

本研究比較兩種集群方法從 15 群至兩群時三種集群指標的增減情形，由於 SPRSQ 的值是不同集群之間所損失的 R-Squared (集群間相異的程度)，因此參考 Sharma<sup>[33]</sup> 的建議，最佳群數標準以 SPRSQ 為主要決定標準，然後比較兩個集群方法在該最佳群數時，另外兩種指標的差異性，結果如表 5 所列。

表 5 最佳群數與集群指標之比較

方法	K-means	KSOM+K-means
最佳群數	11	11
R-Squared	0.776808981	0.844663677
SPRSQ	0.000428661	0.000490703
均方根標準差	8.376465843	0.989164294

這兩種集群方法在集群指標的表現有所不同：其中單純使用 K-means 在 R-Squared 的數值，比起採用兩階段集群方法的 KSOM 搭配 K-means 的數值低大約一成，顯示出雖然採用 K-means 集群有較為單純與不需要正規化前置步驟的優點，然集群之間的相異程度

(R-Squared 值) 卻劣於採用兩階段方法。而在均方根標準差方面，該指標衡量的是集群內各變數的均齊性，該指標則是越小代表越好，也可使用來衡量集群內同質性 (homogeneity) 的一種方式，從該指標可以發現 KSOM 搭配 K-means 的二階段集群方法表現也是優於 K-means 集群法。

基於前述的比較結果，後續的分析將以 KSOM 搭配 K-means 的二階段集群結果為主，依據該方法劃分出的 11 個群組之敘述性統計資料如表 6 所示，可發現各群組的特性都非常不同，茲將各群組之特徵進一步地說明於表 7。

表 6 KSOM 搭配 K-means 時各群組之敘述性統計

群別 屬性	1	2	3	4	5	6	7	8	9	10	11
R	92.46	61.28	68.19	29.61	56.14	26.86	27.14	34.32	60.14	90.62	79.58
F	2.17	1.84	5.78	4.34	5.17	8.71	7.73	3.71	3.17	3.16	2.02
M	34.56	1.61	394.15	40.22	253.86	190.68	301.13	514.05	39.14	330.75	602.30
TM	28.78	3.42	622.12	91.97	407.14	531.24	788.92	572.57	69.96	250.41	322.90
ASS	2.49	1.18	4.26	2.52	3.60	3.35	4.11	4.95	2.90	3.65	4.96
PS	3.7%	0.2%	70.2%	6.8%	45.0%	37.9%	63.2%	97.9%	5.9%	44.2%	98.5%
旅客數	25192	21281	15064	32977	25001	25905	37933	55968	25455	23710	44098
百分比	7.57%	6.40%	4.53%	9.92%	7.52%	7.79%	11.41%	16.83%	7.65%	7.13%	13.26%

表 7 各群組主要特徵說明

群組別	主要特徵
群組 1	<ul style="list-style-type: none"> <li>◆ 距離上次訂票有很長的天數 (92 天)</li> <li>◆ 訂票次數不多 (三個月內共 2 次)</li> <li>◆ 平均每次金額不高 (34 元)，也容易未取票 (取票率 3.7%)</li> </ul>
群組 2	<ul style="list-style-type: none"> <li>◆ 鮮少上臺鐵局預售系統訂票 (1.8 次)</li> <li>◆ 總里程非常低 (3.42 公里)</li> <li>◆ 平均金額很低，常常訂了又取消 (取票率 0.2%)，同時平均訂票分數偏低，發現旅客信用極低</li> </ul>
群組 3	<ul style="list-style-type: none"> <li>◆ 使用預售系統的頻率較高 (5.78 次)</li> <li>◆ 每次平均訂票金額在所有群組中第三 (394.15 元)</li> <li>◆ 相當守信的旅客 (取票率 70.2%)，通常是有計畫性訂票 (4.26 分)</li> </ul>
群組 4	<ul style="list-style-type: none"> <li>◆ 離進行資料分析之基點日期近，新進程度高 (29.61 天)</li> <li>◆ 平均訂票金額不高 (40.22 元)</li> <li>◆ 總里程普通 (91.97 公里)</li> <li>◆ 真正取票機率非常低 (6.8%)</li> </ul>

表 7 各群組主要特徵說明(續)

群組別	主要特徵
群組 5	<ul style="list-style-type: none"> <li>◆ 取票率 45%，但是平均金額不低 (253.86 元)，對於臺鐵局來說是相對可增加營收但是有風險的旅客</li> <li>◆ 此集群的旅客大致上若不需要會自動取消訂票</li> </ul>
群組 6	<ul style="list-style-type: none"> <li>◆ 最常預訂車票的族群 (8.71 次)</li> <li>◆ 總里程在所有集群中第三高，依賴鐵路，但同時有反反覆覆訂了票又取消的情形 (總里程 531.24 公里；取票分數 3.35 分)</li> </ul>
群組 7	<ul style="list-style-type: none"> <li>◆ 總里程非常的突出 (788.92 公里)，最近訂票時間距離基點日期很接近 (27 天)，推估有長期需要鐵路作為交通工具的需求</li> <li>◆ 訂了票比較不會去取消 (4.11 分；取票率 63.20%)</li> </ul>
群組 8	<ul style="list-style-type: none"> <li>◆ 不常使用訂票系統 (3.71 次)</li> <li>◆ 通常是長距離移動，平均每次訂票金額也高 (訂票金額 514.05 元；總里程 572.57 公里)</li> <li>◆ 非常守信用的旅客，計畫性訂票，不取消 (4.95 分；取票率 97.70%)</li> </ul>
群組 9	<ul style="list-style-type: none"> <li>◆ 平均訂票金額 (39.14 元) 與總里程 (69.96 公里) 累積皆不高；</li> <li>◆ 雖然不需要會主動取消 (取票分數 3.65 分)，但是真正取票的機率不高 (44.20%)</li> </ul>
群組 10	<ul style="list-style-type: none"> <li>◆ 有很長的時間沒有使用鐵路車票預售系統 (90.62 天)</li> <li>◆ 平均訂票金額在所有群組屬中等 (330.75 元)</li> <li>◆ 不需要車票時傾向於主動取消訂票 (3.65 分)</li> <li>◆ 取票率 44.2%，可能是同時仰賴其他交通工具的旅客</li> </ul>
群組 11	<ul style="list-style-type: none"> <li>◆ 很少使用訂票系統 (2.02 次)</li> <li>◆ 平均每次訂票的金額都非常高，可能常常幫朋友或家人一起訂票 (訂票金額 602.3 元；總里程為 322.9 公里)</li> <li>◆ 有訂票幾乎不會取消 (取票分數 4.96 分)，同時只要預訂，將近百分之百的可能性會搭該班火車 (取票率 98.5%)</li> </ul>

## 4.2 影響變數的縮減—主成分分析

在統計中，以最佳分群數來分群雖然效果最為優異，但在實務中，市場區隔太多對於資源有限的業者來說反倒相當麻煩，會希望能將市場區隔劃分的群組少一些，鎖定特定屬性相對較高的區隔。因此本研究進一步透過主成分分析方法，將前述六個變數縮減為較少的構面。

經以 SPSS 檢測 6 項變數，得到 KMO (Kaiser-Meyer-Olkin) 值為 0.62，Barlett's 球形檢定之  $p$  值為 0.000，結果顯示適合進行主成分分析。經主成分分析暨因素轉軸後之結果，發現原先的 6 個變數可以萃取為兩大構面，且足以解釋原先變數 72.98% 的變異量 (詳表 8)，表示儘可能在減少損失解釋能力的情況下，使用這兩種主成份能夠解釋大部分的變異。

表 8 解釋總變異量表

成份	初始特徵值			平方和負荷量萃取		
	總和	變異數百分比	累積變異數百分比	總和	變異數百分比	累積變異數百分比
1	2.576	42.929	42.929	2.576	42.929	42.929
2	1.803	30.049	72.978	1.803	30.049	72.978
3	0.785	13.082	86.060			
4	0.519	8.643	94.703			
5	0.223	3.724	98.427			
6	0.094	1.573	100.000			

兩大構面所涵括的變數如表 9 所示，其中第一個構面包括平均金額、平均取票狀態分數以及取票機率 3 個變數，平均金額越高之消費者代表對鐵路的消費越多；平均取票狀態分數越高者越傾向於取票，或者至少會自行取消訂票，比較不會空佔住預訂座位整整兩天，影響臺鐵局對其他顧客的售票；而取票機率高旅客代表其確定度較高，對臺鐵局的營收貢獻自然較大。第二個構面則為新進程度、頻率以及總里程，在以往的 RFM 分析中，新進程度與頻率大多數都被劃分為「忠誠」的概念，這點與本研究有相似性。新進程度代表上次訂票的日期距離執行資料分析的所相差的天數，距離越近通常比較有機會在最近消費，也代表對鐵路車票的需要程度。頻率越高，代表旅客越常登入臺鐵局訂票系統，需要鐵路車票的程度比較高。總里程則是類似於航空業中的「忠誠方案」概念，理論上整體里程越高，代表使用鐵路的需求性也比較高。由前述說明可發現兩個構面所顯示的意義與 Reinartz 與 Kumar<sup>[14]</sup> 運用的忠誠度矩陣非常貼近，所以將第一個構面命名為營收貢獻度，第二個構面則稱為忠誠度。

表 9 主成分所包含之變數及命名

主成分編號	主成份一 (Y 軸)	主成份二 (X 軸)
原變數名稱	M 平均金額	R 新進程度
	ASS 平均取票狀態分數	F 頻率
	PS 取票機率	TM 總里程
命名	營收貢獻度	忠誠度

KSOM 搭配 K-means 所劃分出來的 11 個群組，可以依照各群組內顧客的六個變數數值，換算為兩大構面的主成分分數，得到 11 個群組之平均值在忠誠度矩陣中所處之象限位置，如圖 1 所示。其中第 3 群、第 7 群及第 8 群位於第一象限，即為忠誠矩陣的「真正的朋友」區，這些旅客具有最高的潛在獲利，他們常常利用臺鐵局系統訂票，訂票後的取票機率也相當高，利用鐵路旅行的總里程數在四群中是最高的，不過金額方面僅為第二高

(詳如表 10 所示)，因此推估此類的旅客大部分屬於個人行程，例如出差或假期往返工作就學地點與住家的旅客，這些顧客約占全部旅客的 32.76%；第 10 群以及第 11 群在第二象限的「蝴蝶」區，這些旅客預訂鐵路車票的平均次數不多，但其訂票金額卻是最高，訂票後的取票機率也是很高，是值得爭取其增加消費次數的一個族群，從訂票金額來推估，此族群大部分屬於群體性且有計畫性的訂票，例如家庭式的旅行，而從新近程度及頻率指標來

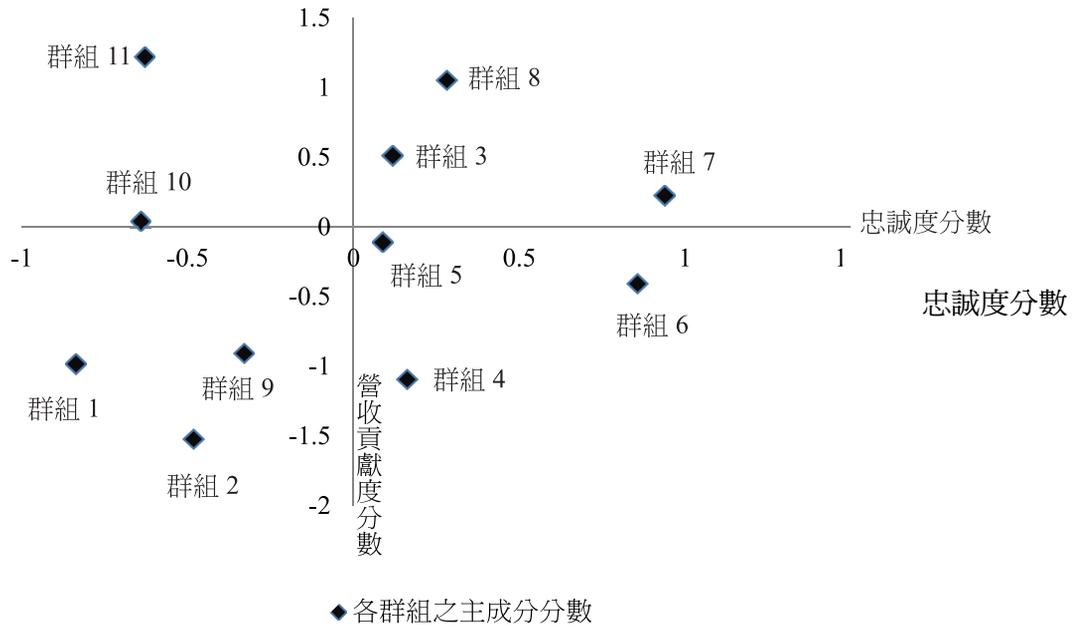


圖 1 11 個群組在忠誠矩陣之位置

表 10 忠誠矩陣中各象限旅客之特性

變數	象限一 「真正的朋友」	象限二「蝴蝶」	象限三「陌生人」	象限四「戀棧」
R	36.50 天	83.44 天	71.80 天	36.67 天
F	5.39 次	2.42 次	2.43 次	5.94 次
M	423.35 元	507.35 元	26.43 元	150.36 元
TM	654.74 公里	297.55 公里	35.85 公里	321.56 公里
ASS	4.56 分	4.50 分	2.25 分	3.10 分
PS	82.99%	79.52%	3.45%	27.78%
顧客數	108965 人	67808 人	71928 人	83883 人
百分比	32.76%	20.39%	21.63%	25.22%
總購買金額 (百分比)	1.97 億元 (56%)	0.68 億元 (19%)	0.09 億元 (3%)	0.78 億元 (22%)

看，則顯示這些旅客可能是旅行次數不多，或者有時會利用其他運輸工具，他們是屬於精打細算的顧客，會比較不同運具的效益後再作決定，因此和臺鐵局的關係仍不穩定，若能提高這些顧客的滿意程度，則可進一步增加臺鐵局的收益；第 1 群、第 2 群及第 9 群處於第三象限的「陌生人」區，這些旅客不常訂票，取票的機率也非常低，而且有較多的情形是逾期未取票（從  $ASS < 3$  可看出），他們佔據了太多的訂位資源，是不值得爭取的一群，約有 21.63% 的顧客屬於這個群組；第 4 群、第 5 群、第 6 群則位在第四象限的「戀棧」區，這些是屬於謹慎的旅客，可能會預先保留行程或同時安排其他運輸工具，待旅行計畫較明朗時再做決定，所以其真正的取票機率不高，不過他們是屬於守信的旅客，會自動取消不需要的行程，由於這些旅客具有游移的特性，並不值得投資額外的資源去吸引。由前面的分析可知，透過主成分萃取的方式，我們可以更簡便的方式了解不同特性的旅客。

### 4.3 顧客群組的判斷—決策樹演算法

前面的集群分析及忠誠矩陣揭露了不同旅客類型的樣態，使得經營者了解那些顧客應該持續維繫，那些顧客可以努力爭取，以及那些顧客不值得投注額外資源，甚至於某些顧客可以適當地縮減原來給予的便利性，以增進其他顧客的福利。不過，之前的分析乃是基於「過去」的訂購情形將舊有顧客加以分類，而在訂票系統隨著時間不斷更新交易資料以及新顧客陸續加入的情形下，顧客的價值可能隨之改變，因此有必要建立一套規則來判斷顧客所處的矩陣象限位置，而決策樹演算法即是一種適合的方法。

本研究以決策樹中的 CART 演算法進行規則的探勘。為驗證所得的規則能適用於未來的分類推估，因此特別在 33 萬筆的旅客資料中，隨機挑選 70% 的資料作為訓練組來歸納出分類規則，其餘 30% 的資料則作為測試組，以測試所得分類規則是否具有效度。表 11 顯示訓練組與測試組的分類預測準確率分別達到 94.14% 與 94.22%，代表透過 CART 演算出的準則有不錯且一致的分類效果。

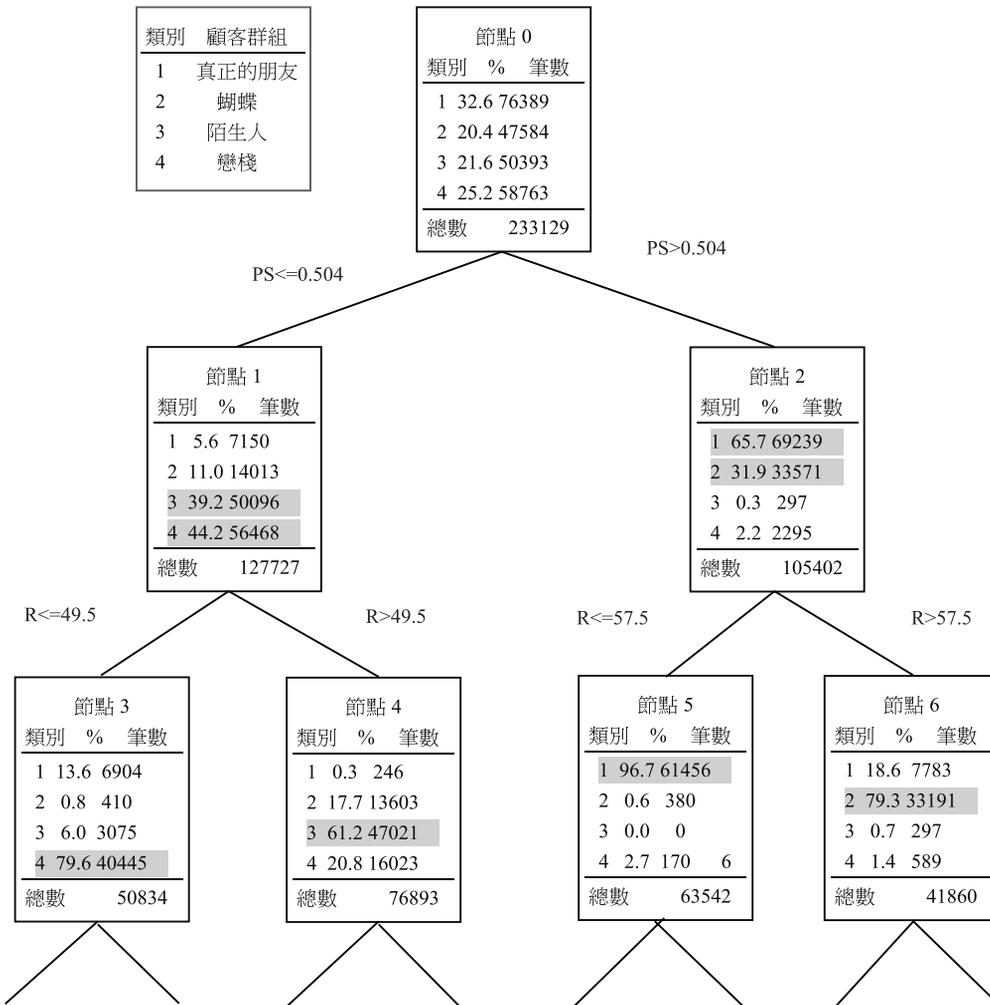
由於經過決策樹演算法所歸納出來的分類規則共有 30 條，限於篇幅，僅以樹狀圖列出所得前面兩階層的分枝情形，如圖 2 所示，此部分之分類規則將於第五節結論與建議中詳細討論。

表 11 決策樹規則之預測準確度

樣本		預測結果			
		象限一	象限二	象限三	象限四
訓練組	象限一	73190	1177	224	1798
	象限二	1014	55690	1390	669
	象限三	33	748	47889	1723
	象限四	1041	539	3297	42707
	正確率	95.81%	94.77%	95.03%	89.75%
	整體準確率	94.14%			

表 11 決策樹規則之預測準確度(續)

樣本		預測結果			
		象限一	象限二	象限三	象限四
測試組	象限一	31216	519	88	753
	象限二	469	23785	597	269
	象限三	7	298	20500	730
	象限四	440	238	1344	18202
	正確率	95.83%	94.69%	95.19%	90.00%
	整體準確率	94.22%			



(註：因篇幅關係，僅列示至第二層的分枝)

圖 2 顧客象限群組決策樹分類結果 (訓練組)

#### 4.4 資料探勘結果對臺鐵局會員制之評析

##### 1. 會員範圍的決定

臺鐵局會員制度適合讓顧客自行選擇加入嗎？目前臺鐵局會員係由消費者自行申請，並沒有任何消費額度的限制，而在現行的禮品兌換機制，一位顧客兩年內至少要消費 15,000 元才能換取一張 100 公里以內的車票（最小的兌換禮品），亦即在 3 個月內平均須購買 1,875 元的車票，而此目標值在本研究所調查搭乘西幹線鐵路的 33 萬餘名旅客中約有 57,002 人達到此額度（詳表 12），其他 27 萬餘人則預估大部分在期限內都無法兌換到最小的禮品，如果全部顧客都加入會員，除了增加臺鐵局會員系統的作業負擔外，也會因為兌換難度過高而使會員產生挫折。

表 12 本研究資料庫中預估能達到酬賓禮之顧客人數

臺鐵局會員酬賓禮項目	平均 3 個月所需消費金額 (元)	資料中符合之顧客數 (占總數百分比)
100 公里以內車票一張 (最小禮)	1875 元或以上	57002 (17.14%)
200 公里以內車票一張	3750 元或以上	17120 (5.15%)
400 公里以內車票一張	7500 元或以上	2911 (0.88%)
600 公里以內車票一張	9375 元或以上	1478 (0.44%)
TR Pass 三日券一張 (最大禮)	15000 元或以上	244 (0.07%)

註：本表僅列出部分酬賓禮項目；資料來源：臺灣鐵路管理局<sup>[3]</sup>。

忠誠計畫會員的招募應著重於顧客的獲利性<sup>[6-7,10]</sup>。從 4.3 節的分析已得知位於「真正的朋友」區的旅客是值得繼續經營顧客關係的主顧，而在「蝴蝶」區的旅客雖然獲利性亦高，但是他們習慣於在鐵路與其它運具間尋找適合的旅行方式，因此，若能利用會員制度將這些蝴蝶追捕進來，吸引他們多加利用傳統鐵路，轉化為使用率較穩定的「真正的朋友」，方能達到運用會員計畫提高重複購買次數之目的。

Tyler<sup>[35]</sup> 曾建議忠誠計畫會員的資格門檻可以訂在消費金額高於「蝴蝶」區顧客的平均數，不過，本研究並非單以消費金額來衡量顧客的獲利性，各象限間的平均消費金額並無絕對的大小關係，因此建議目標會員的範圍可以納入「真正的朋友」區的全部以及「蝴蝶」區的部分顧客。「真正的朋友」區是忠誠與營收貢獻兼備的顧客，必須繼續維繫，而「蝴蝶」區包括第 10 及 11 群組，兩者在忠誠度方面的指標相近，但第 11 群組的營收貢獻度則是明顯優於第 10 群組（詳圖 1），第 11 群組顧客的平均訂票頻率不高，有可能參雜其他運具交替使用，或者本身的旅行需求不高，但平均訂票金額及取票機率都很高，如果能利用忠誠計畫的積點活動「鎖定」他們持續利用鐵路系統，促使他們成為穩定的使用者，進而晉升至「真正的朋友」區。

從 4.3 節的決策樹分析結果可知，利用 3 個月的訂票資料已能使顧客象限群組的預測準確率達到約 94%，另外從圖 2 中決策樹的第一個分枝，發現透過取票率 50.4% 的規則可以大致預測出位於第一、二象限與第三、四象限的顧客，尤其是經過取票率大於 50.4% 的篩選後，97.6% 於該範圍的顧客都是屬於較有價值的第一、二象限（「真正的朋友」及「蝴蝶」區）；若再經過新進程度小於或等於 57.5 天的篩選，則可更進一步地找出位於第一象限的顧客。經過前述兩個簡單的規則，即可分辨出 80.45% 的第一象限顧客，有利於實務上的運用；當然，若要達到整體 94% 的準確率，則還需要更多層的規則來篩選。

至於如何篩選出第二象限中較佳的第 11 群組顧客，則須經過前述取票率大於 50.4% 及新進程度大於 57.5 天的篩選後，再從營收貢獻度的相關變數（亦即平均金額、平均取票狀態分數及取票機率等 3 項）來選擇，表 6 顯示第 10 及 11 群組間在平均金額及平均取票機率有較大的差異，因此建議以所屬第二象限之平均數來訂定篩選標準，以平均金額 507.35 元或取票機率 79.52% 來評定。至於平均取票狀態分數因為需要較多訂票次數才容易判斷，由於此兩群組顧客的訂票次數通常不高，所以不建議以該項變數篩選。

經過前面的規則，可篩選出第一象限之高價值顧客數有 90,697 人及第 11 群組之有潛力顧客數 24,895 人（依平均金額篩選）或 49,508 人（依取票機率篩選）；這些篩選出來的人數約為本研究旅客資料庫的 35% 或 42%，若臺鐵局僅優先針對這些顧客進行會員招募，將可有效地服務會員並節省會員系統資源的投入。

## 2. 酬賓禮的多元化以及會員的分級

依據前面的規則，共篩選出值得加入會員的顧客共約 115,592 人或 140,305 人，對照表 11 所示估計有可能獲得酬賓禮的顧客數 57,002 人尚有段差距。對此，建議可以擴大給獎的範圍來吸引顧客持續的積點行為，除了車票外，臺鐵局還有許多熱門的商品受到旅客的喜愛，例如排骨便當、紀念便當盒、道釘紀念品等，這些多元化的酬賓禮可以滿足不同偏好的顧客並減少顧客無法集滿點數所受到的挫折感。

此外，車票酬賓禮的價值在不同的列車時段是不一樣的。每列鐵路車次所能供給的座位數是固定的，在搭乘的旅客數尚未超出該車次之容量前，每增加一位新的顧客所需之變動成本是相當低的，因此以鐵路車票作為會員兌換的酬賓禮除了能密切地結合會員平日之需要外，對於臺鐵局禮品提供的成本上也是符合經濟效益的。不過，以自強號來說，雖然其平均客座利用率於民國 101 年間僅為 76.8%<sup>[2]</sup>，但是在周末及國定假日期間許多車次的需求實際上是超出乘載容量的，所以若考慮旅客訂不到票時產生的不滿意感甚至顧客的流失，則實際上車票酬賓禮的成本在尖峰車次與離峰車次是不同的，於尖峰車次的車票酬賓禮會增加旅客訂票時的競爭性，亦即影響其他付費旅客訂票成功的機率，因此有必要將尖峰車次與離峰車次的車票酬賓禮分開看待，區分為兩種不同的商品。

面對尖峰車次座位供給有限並且會影響顧客訂票結果之考慮下，有必要將尖峰車次車票的兌換權利保留給等級較高而且取票率較高的會員，以避免萬一該會員退票所造成先前干擾訂票結果之效應。基於前述，建議於「真正的朋友」區以平均取票率 82.99% 以上或者更高之取票率劃分為高級會員，針對這些更忠誠及營收貢獻度更高的會員提供更便捷甚至

客製化的服務，例如旅遊套票、固定行程車票預留等，而這種客製化的服務在運用顧客歷史訂票資料的分析之下，已經變得更為可行與精確。尤其是協助訂不到票的會員煤合未取票或退票的服務，更應該只優先針對高級會員提供。

此外，高級會員的積點換算可以給予較優惠的點數，例如每消費 25 元為一點，此類兌換的差距將可誘使一般會員努力消費以達到高級會員的標準。當然，高級會員在一段時間後也應該再檢驗其是否仍符合要求，而此檢驗則可促使許多高級會員為了繼續維持其資格而持續消費。

## 五、結論與建議

### 5.1 結論

1. 本研究以臺鐵局訂票系統既有的西部幹線 3 個月的顧客訂票資料為基礎，運用資料探勘的技術將訂票旅客加以分群，透過各群組旅客的變數特徵值，可以讓臺鐵局更為了解不同旅客群組間的行為特質，俾能有效地經營顧客關係。此外，在運算中也發現，透過 KSOM 搭配 K-means 的兩階段集群方式能較單純的 K-means 方法達到分群更為精確以及降低運算成本的效果。
2. 集群分析所得的 11 個群組旅客，經轉化為 Reinartz 與 Kumar<sup>[14]</sup> 忠誠矩陣的四大顧客類型後，則更進一步地顯現出旅客中兼具高忠誠度與高營收貢獻的「真正的朋友」旅客有 32.76%；「蝴蝶」區的旅客 (佔 20.39%) 雖然與臺鐵局的關係仍不穩定，但值得盡力爭取其滿意；劃歸於低忠誠度與低營收貢獻的「陌生人」旅客則有 21.63%，這些旅客不常訂票、取票機率低且傾向於逾期未取票，不值得主動爭取；而「戀棧」區的旅客具有游移的特性，會在不同運輸工具間視情況決定，所以也不需要特別投資資源來吸引他們。
3. 由於訂票系統中會有新旅客的加入，舊有旅客的訂票行為也可能會隨著時間而更新，因此本研究運用了決策樹分析中的 CART 演算法建置了旅客類型的判斷準則，這些準則無論在訓練組及測試組的資料中，均能達到準確率 94% 以上的預測水準，顯示出這些規則未來運用在旅客分類時的有效性與可靠性。
4. 臺鐵局的會員系統可以利用本研究所完成的旅客分類判斷準則，篩選出「真正的朋友」區的旅客以及「蝴蝶」區中對營收貢獻度較佳的旅客，作為邀請參加會員的對象，所篩選出來的人數約為本研究旅客資料庫的 35% (若蝴蝶區依平均金額篩選) 或 42% (若蝴蝶區依取票機率篩選)。運用會員系統的忠誠積點活動，除了可以穩定「真正的朋友」區旅客的持續惠顧，以及刺激與鎖定「蝴蝶」區旅客多多利用鐵路系統以集滿點數兌換贈品外，集中經營這些較有價值的旅客也能減少臺鐵局會員系統的資源負擔。

### 5.2 建議

1. 依據臺鐵局現行的會員積點兌換條件，本研究用資料庫的旅客中僅有 17% 的人能達到

兌換最小酬賓禮之資格，為加強會員參加積點兌換的動力，建議臺鐵局能再增加一些小獎品，降低積點達成的門檻，另外所兌換的獎品除了車票之外，也可以採取多樣化的方式，例如便當、道釘紀念品等，增加會員蒐集兌換的興趣。

2. 層級式的會員架構具有鼓勵較高級會員增加消費的效果，臺鐵局可針對「真正的朋友區」平均取票率高於 82.99%者給予較優惠的積點換算以及提供進階的服務，尤其是媒合車票的服務。由於臺鐵局訂票系統中未取票的比例仍高，協助媒合車票的服務應該限定於較有價值的會員。
3. 本研究使用的資料範圍僅涵括基隆至枋寮間旅行的對號列車顧客，限於軟體處理最大筆數的限制，並未使用全部旅客的資料，因此，若加入東部地區及南迴線旅客之訂票資料，也許在結果上會略微不同。不過，本研究所涵括地區之運量約為臺鐵局全部客運量的 87.7%<sup>[36]</sup>，相信已具有不錯的代表性。未來臺鐵局在實際運用時，建議可以再加入各線的資料加以分析。
4. 臺鐵局的車票預售資料庫隨著時間的進行會有資料的更新，並可能因此導致的不同集群結果與決策樹規則，不過，這也是因為會員的價值隨著後續訂票、購買的情形在全部的會員中會相對地改變所致，本研究所提供的是一套制定會員忠誠計畫規則的決策架構與方法，臺鐵局未來仍須隨著資料的更新持續地維護更新以及監控會員的價值變化狀況，以提供適合的服務暨行銷活動。

## 參考文獻

1. 林繼國、陳佩棻，**臺灣地區軌道運輸系統整合規劃**，交通部運輸研究所，臺北，民國 91 年。
2. 臺灣鐵路管理局，「臺鐵統計資訊」，<http://www.railway.gov.tw/tw/CP.aspx?sn=7460&n=6886/>，民國 102 年。
3. 臺灣鐵路管理局，「臺鐵會員服務系統」，<https://members.railway.gov.tw/>，民國 102 年。
4. Sharp, B. and Sharp, A., "Loyalty Programs and Their Impact on Repeat-Purchase Loyalty Patterns", *International Journal of Research in Marketing*, Vol. 14, 1997, pp. 473-486.
5. Stauss, B., Schmidt, M., and Schoeler, A., "Customer Frustration in Loyalty Programs", *International Journal of Service Industry Management*, Vol. 16, No. 3, 2005, pp. 229-252.
6. Berman, B., "Developing an Effective Customer Loyalty Program", *California Management Review*, Vol. 49, No. 1, 2006, pp. 123-148.
7. Kadar, M. and Kotanko, B., "Designing Loyalty Programs to Enhance Value Growth", *Mercer on Transport and Travel*, Vol. 8, No. 2, 2001, pp.28-33.
8. Nunes, J. C. and Drèze, X., "Your Loyalty Program is Betraying You", *Harvard Business Review*, Vol. 84, No. 4, 2006, pp. 124-131.
9. Dowling, G. R., "Customer Relationship Management : In B2C Markets, Often Less Is More", *California Management Review*, Vol. 44, Spring, 2002, pp. 87-104.

10. Ferguson, R. and Hlavinka, K., “The COLLOQUY Loyalty Marketing Census : Sizing Up the US Loyalty Marketing Industry”, *Journal of Consumer Marketing*, Vol. 24, No. 5, 2007, pp. 313-321.
11. Oakes, G., “What Big Data Means to Marketers, and How to Handle It”, [http : //www.marketingmagazine.co.uk](http://www.marketingmagazine.co.uk), 2012.
12. Dachis, J., “Big Data is the Future of Marketing”, [http : //www.businessinsider.com](http://www.businessinsider.com) , 2012.
13. Porter, S., “Big Data Changes the Marketing World Forever”, [http : //www.businessreviewaustralia.com](http://www.businessreviewaustralia.com), 2012.
14. Reinartz, W. and Kumar, V., “The Mismanagement of Customer Loyalty”, *Harvard Business Review*, Vol. 80, No. 7, 2002, pp. 86-94.
15. Liu, Y., “The Long Term Impact of Loyalty Programs on Consumer Purchase Behavior and Loyalty”, *Journal of Marketing*, Vol. 71, 2007, pp. 19-35.
16. Dorotic, M., Bijmolt, T. H. A., and Verhoef, P. C., “Loyalty Programmes : Current Knowledge and Research Directions”, *International Journal of Management Reviews*, Vol. 14, No. 3, 2012, pp. 217-237.
17. Oracle, “Ensuring Customer Loyalty : Designing Next-Generation Loyalty Program”, *Oracle White Paper*, Oracle Corporation, CA , USA, 2005.
18. Marcus, C., “A Practical yet Meaningful Approach to Customer Segmentation”, *Journal of Customer Marketing*, Vol. 15, No. 5, 1999, pp. 494-504.
19. Mooi, E. and Sarstedt, M., *A Concise Guide to Market Research*, Springer, New York, 2011.
20. Jain, A. K., “Data Clustering : 50 Years Beyond K-means”, *Pattern Recognition Letters*, Vol. 31, No. 8, 2010, pp. 651-666.
21. Mangiameli, P., Chen, S. K., and West, D., “A Comparison of SOM Neural Network and Hierarchical Clustering Methods”, *European Journal of Operational Research*, Vol. 93, No. 2, 1996, pp. 402-417.
22. Steinley, D., “K-means Clustering : A Half-Century Synthesis”, *British Journal of Mathematical and Statistical Psychology*, Vol. 59, No. 1, 2006, pp. 1-34.
23. Kohonen, T., “The Self-Organizing Map”, *Proceedings of the IEEE*, Vol. 78, No. 9, 1990, pp. 1464-1480.
24. Kuo, R., An, Y., Wang, H., and Chung, W., “Integration of Self-Organizing Feature Maps Neural Network and Genetic K-means Algorithm for Market Segmentation”, *Expert Systems with Applications*, Vol. 30, No. 2, 2006, pp. 313-324.
25. Van Laerhoven, K., “Combining the Self-Organizing Map and K-means Clustering for On-line Classification of Sensor Data”, *Artificial Neural Networks*, B. Yegnanarayana (Ed.), 2001, pp. 464-469.
26. 林祥生、劉益豪，「應用資料採礦探討國際線航空旅客之線上購票行為」，*運輸計劃季刊*，第37卷，第2期，民國97年，頁197-235。
27. Aggelis, V. and Christodoulakis, D., “Customer Clustering Using RFM Analysis”, *Proceedings of*

- the 9th WSEAS International Conference on Computers, Athens, Greece, 2005.
28. Vesanto, J. and Alhoniemi, E., “Clustering of the Self-Organizing Map”, *IEEE Transactions on Neural Networks*, Vol. 11, No. 3, 2000, pp. 586-600.
  29. Hughes, A. M., *Strategic Database Marketing : The Masterplan for Starting and Managing a Profitable, Customer-based Marketing Program*, Probus Publishing, Chicago, 1994.
  30. McCarty, J. A. and Hastak, M., “Segmentation Approaches in Data-Mining : A Comparison of RFM, CHAID, and Logistic Regression”, *Journal of Business Research*, Vol. 60, No. 6, 2007, pp. 656-662.
  31. 廖述賢、溫志皓，**資料探礦與商業智慧**，雙葉書廊，臺北，民國 98 年。
  32. Lawrence, R. L. and Wright, A., “Rule-Based Classification Systems Using Classification and Regression Tree (CART) Analysis”, *Photogrammetric Engineering and Remote Sensing*, Vol. 67, No. 10, 2001, pp. 1137-1142.
  33. Sharma, S., *Applied Multivariate Techniques*, John Wiley & Sons, Inc., New York, 1996.
  34. Halkidi, M., Batistakis, Y., and Vazirgiannis, M., “On Clustering Validation Techniques”, *Journal of Intelligent Information Systems*, Vol. 17, No. 2, 2001, pp. 107-145.
  35. Tyler, T., “Loyalty Programs : Making Sure You Invest in Only the Right Customers”. [www.genroe.com](http://www.genroe.com), 2010.
  36. 臺灣鐵路管理局，「各站客貨運起迄量」，<https://www.railway.gov.tw/>，民國 102 年。

