

# 電信數據電路客戶流失資料探勘

## Churn of Telecom Data Circuit Using Data Mining Technique

葉燉烟<sup>1</sup> Duen-Yian Yeh  
環球技術學院資訊管理系

鄭景俗<sup>2</sup> Ching-Hsue Cheng  
國立雲林科技大學資訊管理系

傅雲龍<sup>2</sup> Yun-Long Fu  
國立雲林科技大學資訊管理系

<sup>1</sup> Department of Information Management, Transworld Institute of Technology and

<sup>2</sup> Department of Information Management, National Yunlin University of Science & Technology

(Received May 25, 2004; Final Version February 22, 2005)

**摘要：**民國九十年二月起交通部陸續核發三家(台灣固網、速博電信、東森寬頻電信)固定通信網路經營特許執照，加上原有之中華電信共計有四家均可經營第一類電信業務，包括出租各種速率之數據電路供應國內市場需求。因而整個電信市場瞬間形成多家激烈競爭的局面，業者紛紛推出各種優惠促銷方案來拉攏客戶。各家業者為維持市場佔有率及降低日趨嚴重的客戶流失問題，均不惜降低利潤來保有客戶。據統計，企業得到一位新客戶所花的成本是維繫一位老客戶的五至十倍代價。因此業者莫不想盡辦法來留住老客戶，同時也對客戶的消費行為進行探勘分析，企圖找出可能流失的客戶並及早謀新的行銷策略。本研究主要是應用叢集分析(Cluster Analysis)技術，以IBM Intelligent Miner及MATLAB為資料探勘 (Data Mining) 工具，分別將國內某電信業者的數據電路退租資料進行「類神經叢集化」(Neural Clustering)、「人口統計叢集化」(Demographic Clustering)、「模糊C均值叢集化」(Fuzzy C-Means Clustering) 探勘分析，進而選取較佳之叢集分析結果做合理推論，以探勘出可能流失的客戶及哪些產品在競爭市場上已漸居劣勢，必須速謀因應對策。經由探勘分析，本文發現：(1) Group2的16565位客戶為獲利性較低的高忠誠客戶。(2) Group3的31位客戶是具有高度流失傾向的大企業。(3) Group4的319位客戶為高收益且移轉性低的

高忠誠中小企業。在本研究預測出的31個高潛在流失客戶中，經過印證實際潛在流失客戶數後，發現有23個客戶確實已與其他同業進行新合約的協商，印證準確率達74%。這些經資料探勘的成果，將對電信固網業者之數據電路經營具有提昇競爭力的加分效果。

**關鍵詞：**顧客流失、資料探勘、叢集分析、類神經叢集化

**Abstract :** The Ministry of Transportation and Communications had issued three concession licenses of Type I telecommunications to Eastern Broadband Telecom, Sparq Telecom, and Taiwan Fixed Network Telecom respectively from February 2001. Counting the original Chung-Hwa Telecom, there are four telecom companies permitted to operate Type I telecommunications business, including renting out any rate of data circuit to meet the needs of the domestic market. As a result, fierce competition soon arises in the telecommunication market. In order to attract the clients, the providers come up with various favorable sale promotions, which make chaos in the telecommunication data circuit market. On the purpose of maintaining market share and reducing the downward problem of customer moving out, the providers choose to cut down profits to retain customers. According to statistics, however, the cost for a company to gain a new customer is as much as five to ten times more value than that of retaining an old customer. Therefore, the service providers spare no effort to retain old customers. Meanwhile, they try to analyze the attitudes of the customers with data mining technique, attempting to find out customers of the possible moving out and thus help companies to propose new marketing strategies. This study mainly applied the technique of Cluster Analysis, used IBM Intelligent Miner and MALAB as tools, and analyzed the customer's withdrawal data with Neural Clustering, Demographic Clustering, and Fuzzy C-Means Clustering, then selected a best one of churn data to infer the result reasonably. The purpose is to find out the possible moving out of customers and products that are going downward in the market, and thus launches strategies against the problems. These will give an effective help to the company decision maker, particularly obvious benefits and improvements on the marketing strategy planning and the performance. According to the results of data mining, this paper finds: (1) In Group2, 16565 customers are high loyal customers with low profitable. (2) There are 31 large enterprises with high shift tendency in Group3. (3) In Group4, 319 customers are high loyal firms with high profitable and low shift. These results will provide the conduct of data circuits of Telecom with the bonus effect of competitiveness improvement.

**Keywords :** Churn, Data Mining, Clustering Analysis, Neural Clustering

## 1. 前言

全球電信自由化潮流席捲世界各地，2001年新增的固定通信網路業者如雨後春筍般，使整個電信數據電路市場呈現一片混亂。一般電信數據電路主要客戶來自於各大中小企業，平均每家企業均租用有一路以上的數據電路，隨著各行業客戶經營業績的起伏，或競爭者產品新促銷方案或客戶搬家遷移等，都會有增租或退租數據電路的現象，造成相當高的客戶流動率。另一方面，根據統計指出，企業得到一位新客戶所花的成本是維繫一位老客戶五倍至十倍的代價（林德國，民90），因此企業經營者為維持市場佔有率及降低日趨嚴重的客戶流失問題，均不惜降低利潤來保有老客戶，而其中積極的作為即是對客戶的消費行為進行探勘分析，企圖找出可能流失的客戶及相關訊息，並且及早謀新的市場行銷策略供營運決策者迅即採取因應策略，以維持市場佔有率及採取差異化服務來留住客戶。

國際電信聯盟 (ITU: International Telecommunication Union) 公佈未來電信發展趨勢，已由早年的數據經由聲音方式傳送 (Data over Voice) 逐漸轉變為聲音經由數據方式傳送 (Voice over Data)，其意謂著未來電信業者的主要營收來源，將是由一般語音收入轉為以數據電路的營收佔主要大宗，這可由電信業者的數據電路營收逐年蓬勃成長得到印證，未來數據電路市場佔有率將是居於電信業者整體經營成敗具有舉足輕重的地位（毛治國，民91）。

今日資訊科技的進步及全球化市場高度競爭環境衝擊下，各類訊息的收集及運用成功與否，已成為企業在產業競爭力及獲利率的重要關鍵（張寶誠等，民89）。資料探勘的應用主要是於龐大的資料庫中尋找出有價值的隱藏訊息，藉由統計及人工智慧的科學技術，將資料做深入分析以找出其中隱藏的知識，於探勘過程中可根據企業的問題構建不同的模型，並將探勘結果提供企業作營運決策之參考。目前已有許多企業應用資料探勘的成功案例，例如：行動電話公司可藉由資料探勘的技術，將其龐大的客戶通話資料做篩選分析，以預測出哪些可能是潛在的流失客戶，使企業營運者及早因應以避免客戶流失。另外是利用資料探勘技術分析出某項產品在市場上已逐漸失去競爭力，該項產品有被競爭者產品取代的危機，這些資料探勘所得訊息有效提醒企業經營者，促其即時採取因應對策扭轉被市場淘汰的危機。過去在資料庫應用是以資料查詢和簡單的統計分析功能來進行資料的解析，現今這些方法似乎不容易找到隱藏在資料庫中未知的訊息。因此如何透過特定的程序以及方法，幫助企業從大量的營運資料中萃取出以往未知的有用資訊，以提供企業決策者進行決策的參考依據，是目前資料探勘的主要目的。

得利於資訊科技的快速進步，市面上已發展出許多種成熟的資料探勘 (Data Mining) 工具 (Fayyad *et al.*, 1996; 邱志洲等，民91)，方便用來處理龐大如山的資料，其應用範圍甚廣，例如：晶圓圖分類 (簡禎富等，民91)，半導體製造資料特徵萃取 (簡禎富等，民92)，三維資料探勘 (Bohlen *et al.*, 2003)，加州大地震模式的探索 (Bolton *et al.*, 2004)，知識挖掘 (Schuster and Wolff, 2004) 及地圖改良 (Schroedl *et al.*, 2004)，並且獲致良好的成果。鑒於此，本研究的主要目的在

於針對電信業數據電路產品退租交易資料，以資料探勘的技術進行分析研究，企盼提供具有價值的商業訊息，進而能有良好應對策略以因應市場的變化。一般的資料探勘技術主要有六類：叢集分析 (Clustering Analysis)、分類分析 (Classification Analysis)、聯結法則分析 (Association Rule Analysis)、循序特徵分析 (Sequential Pattern Analysis)、時間序列相似性分析 (Time-series Similarity Analysis) 及鏈結分析 (Link Analysis) 等，各類技術依其特性之各異而適用於不同之領域；另一方面，在現實生活中有很多事物不是可以明確的用Yes、No或0、1來分類，所以引用模糊理論而發展出來的探勘技術亦不在少數，模糊聚類分析方法即是一例。

本研究主要是應用叢集分析技術，以IBM Intelligent Miner及MATLAB為資料探勘工具，分別將國內某電信業者的數據電路退租資料進行「類神經叢集化」(Neural Clustering)、「人口統計叢集化」(Demographic Clustering)、「模糊C均值叢集化」(Fuzzy C-Means Clustering) 探勘分析並作相互比較，以得知三者間之差異性，再依其差異性來選取最適宜及符合原始資料特性之分群方法，進而選取較佳之叢集分析結果做合理推論，以探勘出可能流失的客戶及哪些產品在競爭市場上已漸居劣勢而急須速謀因應對策。經由探勘分析，本文發現：(1) Group2的16565位客戶為獲利性較低的高忠誠客戶。(2) Group3的31位客戶是具有高度流失傾向的大企業。(3) Group4的319位客戶為高收益且移轉性低的高忠誠中小企業。這些成果將對電信業者之數據電路經營具有提昇競爭力的加分效果。

本文接續的架構如下：第二部份針對資料探勘的相關技術與RFM分析模式做簡單描述。第三部份為本研究主要之資料探勘分析程序，其中包含了探勘的步驟與資料前置處理的方式。第四部份為實際個案驗證，即針對數據電路產品的退租資料，進行客戶流失探勘分析以及結果解讀與驗證。最後則是結論與後續研究之建議。

## 2. 資料探勘相關技術

本章將針對資料探勘所需的相關技術做概述性介紹，其中包括探勘技術、模糊分群技術、資料倉儲及普遍應用於評量客戶忠誠度與客戶貢獻度的RFM分析模式。

### 2.1 資料探勘技術

資料探勘主要目的是從資料庫 (Database)、資料倉儲 (Data Warehouse) 或其他資訊儲存體中的大量資料，將有價值的隱藏知識發掘出來的過程 (謝邦昌，民90)。換言之，資料探勘著重的是資料庫的再分析，包括模式的建構或是資料樣式的決定，而其主要目的是用以發現資料庫擁有者先前關心卻未曾知悉的有價值資訊 (Hand, 1998)。準此，資料探勘可說是一種整合性的專業技術，其中的相關技術包含資料視覺化 (Data Visualization)、機器學習 (Machine Learning)、

統計技巧 (Statistical Techniques) 及演繹式資料庫 (Deductive Database) 等 (Fayyad *et al.*, 1996; 邱志洲等, 民91)。一般常用的資料探勘技術主要有下述六類：

- (1) 叢集分析：是將所有的物件或資料依其本身的特性分成若干叢集的過程，每個叢集內的物件都具有高度的相似性，而不同叢集內的物件具有高度的不相似性 (Estivill-Castro and Murray, 1997)，叢集分析的目的主要是將叢集與叢集間的差異找出來。
- (2) 分類分析：從已知的類別物件集合中，依據資料的屬性來建立分類模式(如：決策樹)來描述物件屬性與類別的關係 (Han and Kamber, 2000)。然後將未經分類的資料或新的資料依此模式進行預測，這些用來尋找特徵的已分類資料，可能是歷史資料或是其他相關的資料。
- (3) 聯結法則分析：藉由顧客之購物明細，來分析客戶的購物特性，其可得知顧客購物時可能會同時購買哪些商品，並歸納成爲數種規則 (Agrawal and Srikant, 1994)，進而主動推薦符合顧客興趣的商品，如此不會造成傳統行銷中的盲目推銷，而導致行銷成本增加。
- (4) 循序特徵分析：此模式乃針對資料的順序關係進行探勘，例如在商業上從一連串有次序性的交易中，將客戶常依某次序購買的項目找出來，進而歸納出客戶長期購買的行爲 (Agrawal and Srikant, 1995)。
- (5) 時間序列相似性分析：以時間序列曲線來表示一連串的實數數據，這些數據均各別表示該時間點所對應到的數值，這些數值可用來表示銷售量、費率、通話量等 (Agrawal *et al.*, 1995)。
- (6) 鏈結分析：主要針對含有鏈結性的資料，例如：電話通話資料，其記錄有發話者及受話者電話號碼、雙方通話起訖時間等，將這些資訊以節點 (Node) 形式表示 (如電話號碼)，或以鏈結 (Link) 形式表示 (如通話時間)，再依據此鏈結圖形來找尋某些特性資訊，例如形成通話的電話號碼群，這些鏈結資訊極適合用於行銷活動 (Berry and Linoff, 1997)。

各類技術依其特性之各異而適用於不同之領域，例如：鏈結分析被美國電信業者利用於找出傳真電話號碼及供通話及傳真共用之電話號碼，並且據此電話號碼作行銷服務；時間序列相似性分析讓企業可預測明年度的獲利率，或者預測股價未來可能發展趨勢；循序特徵分析除可供企業了解客戶長期的購買行爲，有效的調整進貨種類及降低商品的庫存之外，另一方面也可針對客戶已購買的商品種類依特徵分析，推薦其後續極可能續購的商品，這將使推薦的商品交易成功率大爲提昇；聯結法則分析幫助企業藉由客戶的購物行爲特性分析，方便調整商品的擺設位置，更有效的調整商品庫存量，或強化設計商品促銷方案，最有名的例子是購物籃分析MBA (Market Basket Analysis)；而分類分析可例用於對不同族群客戶給予不同的產品及服務，如可將客戶信用狀況區分爲：高度風險、中度風險及低度風險，或是將客戶區分爲：高貢獻度族群、低貢獻度族群、高忠誠度族群、低忠誠度族群等。

叢集分析與分類分析不同的是，叢集化沒有依事先明確定義的類別來進行分類，資料是根據自身的相近性來叢集在一起的，因此，叢集化可說是分類的前置作業。1990年美國太空總署曾使用叢集分析技術，將數位影像的訊號找出金星表面上的火山位置 (張德民, 民88)。在商業上

的使用方面從客戶的交易資料中，找出共同特徵的客戶，從各種叢集的共同特徵中發現出哪些是潛在的流失客戶，以供企業營運單位及時作出因應策略。依電信業者統計指出使用電信數據電路之對象，大部分為政府機構、學校及各類企業組織等，其中影響客戶流失主要原因為：產品價格因素、業者產品在市場之評價及定位，每逢各電信業者推出各種產品優惠促銷活動時，將直接影響到以價格為考量之客戶忠誠度，因此客戶退租行為之研究極適合以叢集分析法來進行，此即為本研究採用叢集分析技術之主因。

## 2.2 模糊分群技術

模糊集合理論是由 Zadeh (1965) 教授所提出，其將無明顯界限的集合命名為模糊集合 (Fuzzy set)。其擴充傳統集合的概念，使人類思維中很難捉摸的模糊觀念加以具體化與數學化，其特點在解決複雜及非線性的問題。模糊集合是以歸屬函數 (Membership Function) 來表示元素與集合間之歸屬程度 (Grade of Membership)，歸屬函數是在 $[0,1]$ 區間上連續取值，其歸屬度值之大小表示整個集合的程度。歸屬函數是模糊理論的基本概念，透過歸屬函數我們才能對模糊集合進行量化，進一步處理模糊訊息。

此外，規則庫中存放的是模糊推論引擎進行推論時所需要之推論規則，推論規則的型式則是仿效人類的思考推理方式，每個推論規則幾乎都採用 IF-THEN 的形式，依輸入與輸出形式不同大致可分三種：Mamdani 規則庫、Takagi-Sugeno 規則庫、Yasunobu、Miyamoto 和 Ihara 規則庫。另外，解模糊化 (Defuzzifier) 的功能與模糊化相反，它主要在於將推論引擎所得到的模糊輸出值轉換成適當的明確值，以利於實際操作。通常採用的解模糊化方法有面積中心法 (Center of Area Method, COA) 及最大平均值法 (Mean of Maximum Method, MOM)。若一組資料將每個元素均嚴格劃分其屬於某一群集或不屬於某個群集，通常稱之為硬式分群 (Hard Clustering)，如此分群方式顯得太沒有彈性。另一種是模糊分群方法，亦即某一元素屬於某群集之程度為一個 0 至 1 之間的數值，其為模糊集合之歸屬度，常見的方法有「模糊 C 均值」(Fuzzy C-Means) 簡稱為「FCM」來作分群 (Bezdek, 1981)。

## 2.3 資料倉儲

一般電信公司日常運作均需藉助各種資訊系統，例如：帳務系統、客戶服務資訊系統、數據電路服務資訊系統、行動電話服務資訊系統、障礙服務資訊系統...等，目前業者所面臨之挑戰是如何整合各系統及利用這般龐大的資料庫。這些資訊系統中的資料往往不具有一致性，不但可能部分有重複性，而且是分散在各個不同系統的資料庫內或大型主機中，資料倉儲 (Data Warehouse) 的功能正好是用來解決此一問題的最有效方案 (Han and Kamber, 2000)。

根據最近一項對於全球廿八個電信業者所做的報告顯示，資料倉儲提供電信公司的重要應用包括：客戶型態的分類、目標市場的行銷、產品收益分析以及促銷活動的管理等，其中最主要

的是客戶型態的分類。適當的客戶型態描繪可讓電信公司發掘、拓展並掌握具有價值的客戶，也能幫助電信公司查覺潛在的市場機會（中華電信，2000）。

## 2.4 RFM 分析模式

普遍應用於評量客戶忠誠度與客戶貢獻度方面，其中 R (Recency) 值表示客戶最近一次購買的時間有多久，是一種時間量度，由最後一次購買起算至現在之時間，如最近購買日期離現在越遠，則表示此客戶的購買行為可能改變。F (Frequency) 值表示客戶在最近一段時間內購買的次數，如於一星期或一個月或一個季的區間內，其表示客戶與業者雙方互動的程度，因此 F 值越高表示該客戶與業者互動性很高。M (Monetary) 值表示客戶在最近一段時間內購買的總金額，M 值越高表示客戶貢獻度越高 (Bult and Wansbeek, 1995)。

一般而言，顧客關係管理 (CRM) 分析著重在對於客戶貢獻的分析，相對的，RFM 分析則在強調以客戶的行為來區隔分群，依每個分群特性決定如何作因應對策，因此 RFM 分析已普遍被業界廣為採用的一種評估方法。另外，Sung and Sang (1998) 文中亦提到使用 RFM 行銷策略，用來探勘旅館免稅店之顧客消費行為。

## 3. 資料探勘分析程序

本章主要是說明電信數據電路客戶流失資料的探勘分析程序，共分為資料探勘步驟、資料萃取、資料轉換、數據電路退租RFM五等分法及分群演算法等五個部份。

### 3.1 資料探勘步驟

本研究整個資料探勘的步驟敘述如下，並如圖1所示：

- (1) 將數據電路資訊系統內之資料庫，於每個月初定期匯出上個月數據電路退租資料做為探勘原始資料 (Raw Data) 檔。
- (2) 以MS SQL之Data Transformation Service Wizard (DTS) 程式將資料匯入MS SQL Server，以方便SQL語言進行欄位資料的選取及轉換。
- (3) 以SQL語法在SQL Server Query Analyzer中將原始資料檔進行前置處理，其程序首先是資料萃取然後進行資料轉換及數據電路退租R、F、M五等分法等程序。
- (4) 待前置處理完之資料整合後匯出成平坦檔案 (Flat File)，以符合Intelligent Miner探勘所需之格式。
- (5) 將欲分析之平坦檔案存入Intelligent Miner目錄以進行分群演算法探勘分析。
- (6) 將探勘分析結果以電信數據電路領域知識進行判讀，並轉成易讀之報表以供企業營運決策者進行客戶挽回之策略。

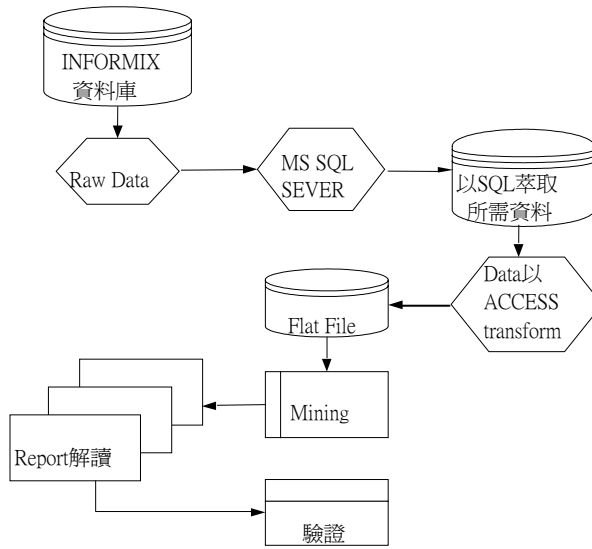


圖1 資料探勘步驟

(7) 高度流失傾向客戶可能持續退租，依預測其可能很快轉移至其它固網業者，因此列印出名單與下一個月之退租客戶名單互相比對，以估算出分群之正確性。

### 3.2 資料萃取

本文以中區某電信業有關客戶對數據電路產品退租交易資料進行實際探勘研究分析，在數據電路退租資料庫中，原始資料共有12個欄位的交易紀錄。第1欄位表示高價值客戶以『\*』表示，其定義為平均每個月電信費約十萬元以上貢獻度之客戶；第2欄表示數據電路甲端裝設所屬地區，即以市內電話之區域號碼來分辨；第3欄表示數據電路型態；第4欄表示數據電路編號；第5欄表示數據電路客戶編號；第6欄表示數據電路客戶名稱；第7欄表示數據電路速率；第8欄表示數據電路甲端裝設地址；第9欄表示數據電路乙端裝設地址；第10欄表示數據電路退租日期；第11欄表示數據電路起租日期，而第12欄表示數據電路所附掛之加值業務。有關資料探勘的資料前置處理說明如下：

- (1) 選擇資料：原始資料中有許多欄位資料，並非全部為本研究所需資料，因此需以SQL語法篩選出。
- (2) 刪除及修正資料：篩選出資料中有許多是重複與錯誤的部分，必需進行整理與更正，以符合探勘資料之正確性。
- (3) 轉換資料：篩選出的資料中其資料值或資料型態可能不符合探勘所設定之格式需求，因此必需進行資料轉換的工作。



- (4) 資料合併：篩選出的資料可能分散在數個資料表中，因此必需進行資料合併的動作。
- (5) 切割資料：篩選出的資料中可能集中在一個資料表內，因此必需進行資料切割的動作，以得到所需要的資料。
- (6) 統計資料：為減輕資料探勘程序之負荷，並提升探勘資料處理速度，須預先將資料做必要之加總(SUM)、平均(AVERAGE)、取最大值(MAX)、取最小值(MIN)等運算。

### 3.3 資料轉換

在SQL SERVER中，經上述程序萃取初步資料後，將這些資料進一步以SQL語言轉換成我們所定義的格式，以方便Intelligent Miner進行資料探勘分析，格式定義如表1所示。當資料完成轉換後，將以相同的數據資料進行「類神經叢集」、「人口統計叢集」、「模糊C均值叢集」等三種叢集方法分析並比較其特性。

### 3.4 數據電路退租 RFM 五等分法

RFM五等分法 (Bult and Wansbeek, 1995) 中，每個級距值的設定直接影響整個分析結果之正確率，其值的訂定屬於研究探勘模型的重要一環，該值必須定期依大環境的變動適時檢討，以維持整個探勘模型的正確率 (Sung and Sang, 1998)。以下之RFM五等分法之各級距值，乃以中區某電信業之客戶數據電路產品退租交易資料，依據該業者多年的經驗法則而設定，其中考量因素包含每等分內之戶數多寡、後續服務及行銷成本等。

#### • 距最近電路退租時間(R)

(1) 前1個月給5分；(2) 前2個月給4分；(3) 前3個月給3分；(4) 前4個月給2分；(5) 前5個月以上給1分。

#### • 分析期間內之退租電路總數(F)

(1) 總數大於30路給5分；(2) 總數大於20路至29路間給4分；(3) 總數大於10路至19路間給3分；(4) 總數大於4路至9路間給2分；(5) 總數大於1路至3路間給1分。

#### • 分析期間內之退租電路總值(M)

依市內數據電路月租費表，將其分成五等份，退租電路總價值越高給分越高。

(1) 電路速率45Mbps以上給5分；(2) 電路速率於768Kbps至2046Kbps間給4分；(3) 電路速率於128Kbps至512Kbps間給3分；(4) 電路速率64Kbps給2分；(5) 電路速率為14.4Kbps、56Kbps等音頻級給1分。

R、F、M三個構面均具有同等權值，表示其每個構面均同等重要，於表2為退租客戶特徵分群表中顯示資料探勘後得到的八個叢集，分別歸納對照到四種個別特質的分群中，每個分群內高R值叢集之潛在客戶流失預測正確率比低R值叢集高。表中Group3所示的客戶是本研究主要鎖定的目標，此群客戶對業者而言貢獻度最高，即其獲利性最高，但此群客戶極易移轉至競爭業者，

表1 R、F、M定義因子

因 子	定 義
R	以目前為基準距上次數據電路退租之期間。
F	於觀察期間內之數據電路退租總數。
M	於觀察期間內之數據電路退租總價值。

只要其它競爭業者略施利誘此群客戶極易流失，因此必須速謀因應對策留住此群客戶以保住高額營收。Group1的客戶是屬於獲利性較低的群組，但其亦有高度的流失傾向，照顧好老客戶將使所花的營運成本較低，因此必須要有積極的挽回客戶的措施。Group2的客戶是屬於獲利性較低的群組，但是此群客戶具有極高度的忠誠性，不易轉移至競爭者，日後必須以各種行銷優惠策略配合將可提升為穩定的獲利客戶。Group4的客戶是屬於獲利性高及移轉性低的群組，此群是業者最佳的客戶，宜善加經營以長久保有此類客戶，其將為公司帶來穩固的營收。

### 3.5 分群演算法

包括「類神經叢集化」、「人口統計叢集化」兩種探勘分析，及以相同之平坦檔案於MATLAB中進行「FCM」叢集分析，再從中選擇最適合本資料特性之分析方法。

## 4. 個案驗證

本研究分析原始資料來自國內一家大型電信固網公司，資料期間為民國九十一年四月起至民國九十一年十月止為期六個月，探勘資料含蓋區域為台灣中部六縣市，即北自苗栗縣南迄雲林縣。在探勘結果解讀方面主要是以Intelligent Miner的人口統計分群結果，說明每個分群之客戶特

表2 退租客戶特徵分群表

客戶分群	Group1	Group2	Group3	Group4
問題特徵	獲利性低 移轉性高	獲利性低 移轉性低	獲利性高 移轉性高	獲利性高 移轉性低
R / F / M	↑↓ ↓↑	↑↓ ↓↓	↑↑ ↓↑	↓↓ ↑↑
說明	本群客戶屬獲利性低群組也有流失傾向，需迅速採取應對策略。	本群組客戶雖屬獲利性低，但屬於忠誠客戶宜加速開拓商機以提昇獲利。	本群組雖屬貢獻度高之客戶，但極易移轉至其它競爭者，需迅速採取應對策略以留住客戶。	本群組客戶屬貢獻度較高之客戶及其移轉性低，宜善加經營長久保有此類客戶。

性及找出最可能流失之客戶，以提供營運單位迅速進行該目標客戶群的挽回對策，並檢討市場行銷策略以鞏固客戶防止客戶流失，這樣將可保持既有之市場佔有率及達成公司之年度營收目標。於驗證程序中是將研究結果找到之可能流失客戶進行驗證比對工作，以確定本模式為一種可靠的客戶流失探勘模型，以供日後快速有效分析作為未來營運作業的重要參考環節。

#### 4.1 選擇最適叢集分析方法

- (1) **類神經叢集化**：探勘功能是使用「Kohonen 特性對映」類神經網路，「Kohonen 特性對映」是利用一種稱為自行組織的程序，將相似的輸入資料分成同一組，可以指定叢集數及處理階段數，這些參數負責控制處理時間，以及將資料記錄指定給叢集時，所使用的顆粒程度。類神經叢集化的主要工作是尋找每一個群聚的中心，這個中心也稱之為叢集原型，就每一個輸入資料中的記錄而言，「類神經叢集化」探勘功能主要計算最接近群聚原型的記錄得分。在執行類神經探礦功能時，使用最佳化時間選項，如同類神經分類、類神經叢集、或類神經預測，因為在訓練資料上，類神經功能比其它探礦功能需要更多的處理階段，使用最佳化時間選項會明顯地減少處理的時間。輸入本研究之退租客戶RFM純文字檔內容進行類神經叢集化分析得到全部叢集明細，其分為八個叢集，如表3所示，其中最小叢集數為292屬叢集0，最大叢集數為4,710屬叢集2。全部探勘資料之平均信賴度值皆大於0.58。
- (2) **人口統計叢集化**：之探勘功能是使用另一種通常會讓模型收斂更快的演算法，因此這種方式會導致叢集化模型有點特殊，此探勘功能是將指定記錄叢集的相關資訊儲存到暫存檔中，當重新執行該功能時會使用到該暫存檔。找出叢集的主要原則是將具有相似特性的記錄分成同一組，Intelligent Miner通常會搜尋整個探勘基地，以發現出同樣最常出現的特性，將相關的記錄分成同一組。在此探勘功能的結果中，包含偵測到的叢集數以及構成每個叢集的特性，其結果都會顯示這些特性在整個叢集內的分佈情況。對超大型資料庫而言其可進行快速和自然的叢集化作業，它會自動判斷要產生的叢集數目。人口統計叢集化是一種反覆的處理過程，通常需要逐步調整其設定值，以取得最好的叢集結果，在叢集化模式中，該功能會試著形成最理想的叢集，並以指定數目的最大量來建立叢集。此叢集明細共分為八個叢集，如表3所示，其中最小叢集數為5屬叢集5，最大叢集數為15,797屬叢集0。全部探勘資料之平均信賴度值大於0.77。
- (3) **模糊C均值叢集化 (FCM)**：其可方便使用MATLAB工具之模糊系統函數來進行分析以建立叢集，模糊叢集方法應用在數據的組織與分類或目標的識別與分類，或是數據的壓縮與還原及模型建構的應用等方面，利用模糊叢集技術可將特徵數據聚類，使數據聚集在叢集中心附近。輸入本研究相同之客戶RFM純文字檔資料進行FCM分析，運算所得結果，再以Data Transformation Service Wizard程式將其匯入MS SQL SERVER中，使用Query Analyzer進一步

計算各個叢集數量，其結果共分爲八個叢集，如表3所示，其中最小叢集數爲199屬叢集7，最大叢集數爲4,710屬叢集3。全部探勘資料之平均信賴度值大於0.62。

綜合上述三種叢集理論分析比較結果，以人口統計叢集化分析具有較高之平均信賴度值，以及每個分群結果較能完全映照出本研究個案資料的分佈特性，故本研究以採用「人口統計叢集化分析演算法」較佳。

## 4.2 結果分析與驗證

以「人口統計叢集化分析演算法」之結果爲後續研究對象，就叢集0爲例，圖2表示叢集0的分佈特性，本叢集佔母體百分比爲93.39%，其中M值爲1的佔該群100%；F值爲1的亦佔該群100%；而R值爲1的佔該群30%，2的佔該群14%，5的佔該群20%，因此叢集0的RFM歸屬值爲低/低/低，依退租電路特徵分群表可對照出本叢集屬於Group 2，意即本群組客戶的特性是：帶來的獲利性雖低但其移轉性低，是屬於高忠誠度的客戶，業者可配合開拓新商機以提昇獲利率。觀其退租行爲大都屬於日常之業務調整或搬家異動或因價格因素改租同一固網業者之其他業務。其客戶種類顯示一般客戶佔99%，企業客戶佔1%，退租電路型態以DL居多約佔96%，退租電路速率平均以64Kbps居多約佔97%以及其它低速電路。退租地區以台中地區居首如地區代碼42 (北台中21%) 及43 (南台中19%) 共計佔40%。

相同的步驟重複對叢集1至7實施後，歸納結果如表4所示。由叢集分析總表中可看出其共分爲八個叢集，依退租電路特徵分群表歸類爲Group2的有二個叢集，其爲叢集0及叢集1計有16565位客戶。歸類爲Group3的有二個叢集，其爲叢集5及叢集6計有31位客戶。歸類爲Group4的有四個叢集，其叢集爲2、3、4、7計有319位客戶。顯然的，採用「人口統計叢集化分析演算法」之結果做爲歸類之依據，Group 1並不存在。

表3 三種叢集演算法分析比較特性

	類神經叢集	人口統計叢集	模糊C均值叢集
Cluster0	292	15,797	420
Cluster1	345	768	2,736
Cluster2	4,710	262	2,243
Cluster3	377	27	4,710
Cluster4	2,347	19	499
Cluster5	3,193	5	2,915
Cluster6	2,915	26	3,194
Cluster7	2,736	11	199
平均信賴度	> 0.58	> 0.77	> 0.62

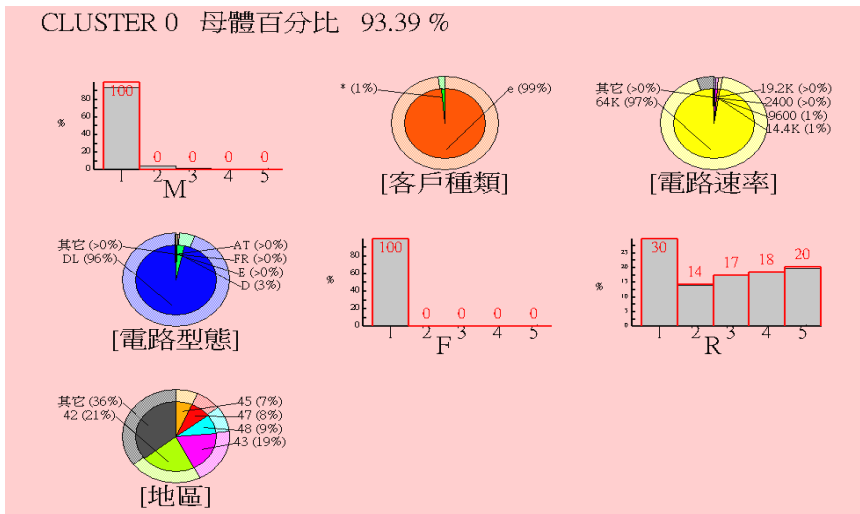


圖2 叢集0的分佈特性

表4 叢集分析總表

Cluster	客戶比例	客戶數	R平均值	F平均值	M平均值	Group
0	93.39%	15,797	2.85	1	1	2
1	4.54%	768	2.75	1	2	2
2	1.55%	262	2.61	1	3	4
3	0.16%	27	3.85	1.67	4	4
4	0.11%	19	2.42	1	4	4
5	0.03%	5	1.40	3.6	4.2	3
6	0.15%	26	3.79	3.35	5	3
7	0.07%	11	2.18	2	3.27	4

表4所展示的意涵說明如下：(1) Group2的16565位客戶目前雖帶來的獲利性較低，但這些是屬於忠誠度較高的客戶，未來假以時日進行機會行銷，規劃一些適合的產品優惠組合方案，亦能開拓一片天地，給業者帶來穩定的收益。(2) Group3的31位客戶是具有高度流失傾向的客戶，經探勘得知大部份均為企業大客戶，這也是各競爭業者必爭的對象，這群客戶給業者帶來高收益，但其常以價格為導向，忠誠度較低。只要競爭業者給予較高優惠即會流失。針對本群客戶應提供合適的差異化服務，除了提供網路技術方面的整合 (Total Solution) 服務外亦應做產品價格的優惠包裝來守住客戶，畢竟業者80%的營收均來自此群客戶。(3) Group4的319位客戶能給業者帶來

高收益，其移轉性又低，這是所有企業經營者夢寐以求的，此群客戶應善加經營以長久保有此類高價值客戶，因此所有規劃優惠促銷活動均應包含此群的客戶不可遺漏。

由先前的分析發現Group3為高度流失傾向的客戶群，其大部分也都是屬於企業級大客戶，整個驗證程序首先以叢集5及叢集6分群資料合併成report\_g6的檔名，將檔案內31個Id與次月之數據電路退租資料檔endrent91\_11進行碰檔，將使report\_g6檔內之31個Id與次月電路退租資料檔endrent91\_11之數據進行互相比對，凡兩個檔案內有相同Id之數據電路退租資料將挑出全部存入新檔名為：lost91\_11，結果共查詢出31家總計有201筆。接著由SQL語法敘述分別歸納出潛在流失目標客戶群資料，以提交負責之服務人員進行客戶挽回行動及分析流失的可能原因。在服務人員進行客戶關懷訪談並了解客戶現況，同時發現有部分客戶正與其它競爭者進行洽談中，有部分客戶已與競爭者業者簽妥年度合約，也有部分客戶已同時租用數路其它競爭業者數據電路。表5呈現此次潛在流失客戶的印證結果，其中高潛在客戶流失預測數分別為叢集5有5個客戶，叢集6有26個客戶，經過印證實際潛在流失客戶數為叢集5有3個客戶，叢集6有20個客戶，準確率為74%。

於電信數據電路的經營模式中，業者主要以行銷為導向，配合網路技術為後盾，針對高流失傾向的客戶群施以推薦相關的加值服務內容，或善用優惠合約進行捆綁以留住客戶，經過適度的努力折衝及配合行銷策略的運用整體上客戶的挽回效果令人滿意，本模式所探勘之潛在流失客戶經統計其準確度達74%，這將使留住老客戶增裕營收的策略上提供有利的契機。本研究之探勘資料經由人口統計叢集分析演算法得到的叢集數或叢集客戶，數量上比較符合本資料的特性，因此探勘之潛在流失客戶其準確度可達到較高之比例，方便後續進行客戶挽回行動時可接近彈無虛發，收到較佳的客戶挽回效果。

## 5. 結論與建議

### 5.1 結論

本研究以國內一家大型電信固網公司為研究對象，資料期間為民國九十一年四月起至民國九十一年十月止為期六個月，而區域含蓋了台灣中部六縣市，即北自苗栗縣南迄雲林縣。資料業經「類神經叢集化」、「人口統計叢集化」及「模糊C均值叢集化」三種方法探勘分析比較後，以「人口統計叢集化」分析具有較高之平均信賴度值，並且每個分群結果，亦較能完全映照出本研究個案資料的分佈特性，因此以「人口統計叢集化」之結果為對象。在實務印證上，本研究預測出高潛在客戶流失數在叢集5內有5個客戶，叢集6內有26個客戶，合計31個客戶，經過實際印證後，發現叢集5有3個客戶，叢集6有20個客戶，合計23個客戶為潛在流失客戶，準確率達74%；另外，分析結果亦能有效提醒專責人員適時察覺，這對留住老客戶提供一個很重要的契機，例如：由客戶流失的案例中可發現，通常客戶佔大部分比例不曾詳加了解與電路提供者所簽訂的合約內容，僅片面以價格高低為主要考慮因素，而逕行與業者簽約，當使用者發現到每個月之通信費並

表5 潛在流失客戶預測統計表

Group3 目標客戶群	流失預測數	印證實際數	準確度
Cluster 5	5	3	-
Cluster 6	26	20	-
Total	31	23	74%

無實質的降低，或者數據電路品質與預期的有所落差時，客戶通常會主動接洽想返回，但由於已受該合約捆綁處理起來將倍加困難。如此，可歸納出會產生客戶流失的原因如下數點：

- (1) 競爭者低價促銷。
- (2) ADSL廣泛應用致使其他產品競爭力會漸趨於弱勢。
- (3) 數據電路品質不穩定。
- (4) 無法提供整合服務能力。

## 5.2 建議

本研究探勘分析結果有效的篩選出高流失傾向的客戶資料，加上經營者的成功策略，能有效挽回這些高流失傾向的客戶措施，使企業整體的營收目標邁向達成之路，雖然目前獲得預期成果，於此亦提出數點建議以供後續研究之參考。

- (1) 詳細比較選擇各種分析技術的適用性，有助於達成更理想的目標效果，例如本研究以相同的分析資料分別進行「類神經叢集化」、「人口統計叢集化」、「模糊C均值叢集化」等三種叢集方法進行分析比較其各分群特點，以找出最適用於本研究目標的方法。
- (2) 重複選擇測試各項參數以找出最合適的結果，例如：使用Intelligent Miner時選擇最大處理階段數及精確度改良值或相似性臨界值等不同的參數選擇，皆會有不同的分析結果，其所表現不同的結果可由領域知識來判別其適用性並加以選擇。
- (3) 於客戶流失的探勘議題中亦有多種探勘模式，個案宜視資料性質選擇不同之演算法，並將結果進行適當的驗證程序。研究模型的適用程度亦會隨時空環境的不同而會產生不同的結果，因此模型需要不斷的修正或重新建構以配合整體的環境變化，為確保模型品質宜定期進行評估與重新學習。

## 參考文獻

- 毛治國，中華電信，第三十六卷第一期，民國91年。  
林德國，口碑行銷，台北：遠流出版社，民國90年。

- 邱志洲、李天行、周宇超、呂奇傑，「整合鑑別分析與類神經網路在資料探勘上之應用」，工業工程學刊，第十九卷第二期，民國91年，9-22頁。
- 張德民，資料探勘：從搜尋金星火山到偵查考試作弊，資訊傳真周刊，民國88年3月。
- 張寶誠、楊景棠、袁建中、陳俊傑，「台灣通訊工業發展趨勢預測與研究」，管理與系統，第七卷第二期，民國89年，227-248頁。
- 謝邦昌，資料採礦入門及運用，台北：資商訊息顧問股份有限公司，民國90年。
- 簡禎富、李培瑞、彭誠湧，「半導體製造資料特徵萃取與資料挖礦之研究」，資訊管理學報，第十卷第一期，民國92年，63-84頁。
- 簡禎富、林鼎浩、劉巧雯、彭誠湧、徐紹鐘、黃佳琪，「建構晶圓圖分類之資料挖礦」，工業工程學刊，第十九卷第二期，民國91年，23-38頁。
- Agrawal, R. and Srikant, R., 1994, "Fast Algorithms for Mining Association Rules," In *Proceedings of International Conference on Very Large Data Bases*, Santiago, Chile, 1994, pp.487-499.
- Agrawal, R. and Srikant, R., "Mining Sequential Patterns: Generalizations and Performance Improvements," Research Report RJ 9994, IBM Alma den Research Center, San Jose, California, 1995.
- Agrawal, R., Lin, K., Sawhney, H.S., and Shim, K., "Fast Similarity Search in the Presence of Noise, Scaling and Translation in Time-Series Data bases," In *Proceedings of the 21<sup>st</sup> International Conference on Very Large Data Bases*, Zurich, Switzerland, 1995, Sep.
- Berry, J.A. and Linoff, G., *Data Mining for techniques: For Marketing Sale and Customer Support*, Canada: John Wiley & Sons Inc., 1997.
- Bezdek, J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York, 1981.
- Bohlen, M., Bukauskas, L., Eriksen, P.S., Lauritzen, S.L., Mazeika, A., Musaeus, P., and Mylov, P., "3D visual data mining- goals and experiences," *Computational Statistics & Data Analysis*, Vol.43, 2003, pp.445-469.
- Bolton, R.J., Hand, D.J., and Crowder, M., "Significance tests for unsupervised pattern discovery in large continuous multivariate data sets," *Computational Statistics & Data Analysis*, Vol.46, 2004, pp.57-79.
- Bult, J.R. and Wansbeek, T., "Optimal selection for direct mail," *Marketing Science*, Vol.14, No.4, 1995, pp.378-381.
- Estivill-Castro, V. and Murray, A.T., 1997, "Spatial Clustering for Data Mining with Generic Algorithms," Technical Report FIT-TR-97-10, Faculty of information Management, Queensland University of Technology, 1997.



- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., "The KDD process for extracting useful knowledge from volumes of data," *Communications of the ACM*, Vol.39, 1996, pp.27-34.
- Hand, D.J., "Strategy, methods, and solving the right problem," *Computational Statistics*, Vol.13, 1998, pp.5-14.
- Han, J., and Kamber, M., *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2000.
- Schroedl, S., Wagstaff, K., Rogers, S., Langley, P., and Wilson, C., "Mining GPS traces for map refinement," *Data Mining & Knowledge Discovery*, Vol.9, 2004, pp.59-87.
- Schuster, A. and Wolff, R., "Communication-efficient distributed mining of association rules," *Data Mining & Knowledge Discovery*, Vol.8, 2004, pp.171-196.
- Sung, H.H. and Sang, C.P., *Application of data mining tools to hotel data mart on the intranet for database marketing*, Elsevier Science Ltd., 1998, pp.1-31.
- Zadeh, L.A., "Fuzzy sets," *Information and Control*, Vol.8, 1965, pp.338-353.