

第二章 文獻探討

本研究旨在運用資料探勘分類技術，建構台灣師大游泳會員流失區別模型。本章之主要目的在於探討與本研究相關之文獻，俾對研究主題能有更深入之瞭解，以作為本研究之理論基礎。全章共分五節：第一節、資料探勘；第二節、資料探勘分類功能運用的技術；第三節、顧客流失與資料探勘；第四節、相關文獻之探討；第五節、本章總結。

第一節 資料探勘

資料探勘(Data Mining, DM)，也有學者稱之為「資料採礦」、「資料挖礦」、「資料採掘」、「資料挖掘」或「資料發掘」。本節先對資料探勘的定義做一整理，再對知識發現、資料倉儲與資料探勘的關係；資料探勘的流程；資料探勘的技術與功能；資料探勘的應用逐一探討。

壹、資料探勘的定義

資料探勘是目前學術與業界炙手可熱的研究主題，許多學者就不同的角度為資料探勘下了諸多定義，以下列舉數位資料探勘研究領域常引用之定義，羅列於後：

Hall(1995)的定義：資料探勘是一種結合資料視覺化(data

visualization)、資料庫、統計方法以及機器學習等技術，以便從龐大資料量中，萃取法則或模式所寓含的知識。

Fayyad 等人(1996a)的定義：資料探勘是知識發現其中的一個步驟。資料探勘透過演算法，將資料作一分析與應用，以找出其特徵(pattern)與模式(model)的過程。

Berry 與 Linoff(1997)的定義：資料探勘是使用自動或半自動的方式對大量的資料作分析，以找出有意義的關係或法則。

陳文華（民 88）的定義：資料探勘是從一個龐大的資料庫中將正確、以前未發覺卻非常重要的資訊加以抽離出來，並利用這些資訊來做出重要決策的過程。

李昇暉(民 89)的定義：資料探勘是從大量的資料庫中找出相關的模式(Relevant Patterns)並自動地萃取出可預測的資訊。

彭文正(民 90)的定義：資料探勘是一門結合統計與人工智慧的新興科學，它幫助我們整理資訊、分析資訊、達到去蕪存精，化腐朽為神奇的目的。資料探勘是經由自動或半自動的方法探勘及分析大量的資料，以建立有效的模型及規則。

綜合上述學者對資料探勘的定義，將其歸納成以下概念：

一、資料探勘是知識發現其中的一個重要步驟，該步驟是從資料倉儲或其它資訊儲存體分析大量資料的過程。

- 二、資料探勘分析過程是自動或半自動的處理程序，該程序結合資料視覺化、機器學習、統計方法以及資料庫等多種技術。
- 三、資料探勘的目的是從資料庫中萃取出有效的、事前未知的，以及潛藏有用的資訊，並找出資料間重要的特徵與模型。
- 四、資料探勘的應用是將結果中可用資訊提供決策人員參考。

貳、知識發現、資料倉儲與資料探勘的關係

在資料探勘相關文獻上，知識發現 (Knowledge Discovery in Database, KDD)、資料倉儲 (Data Warehouse, DW) 與資料探勘等名詞常相提並論，觀念上也容易混淆，茲將三者的關係分述於下：

一、知識發現與資料探勘的關係

根據 Fayyad 等人 (1996b) 對知識發現的定義：知識發現是一個從資料中找出合理、新奇及潛在效益的一個非瑣碎 (Nontrivial) 流程，其目的在瞭解資料的特徵 (Patterns)。整個知識發現的程序如圖 2-1 所示，簡略說明於下：

(一) 選擇 (Selection)

決定知識發現目的，理解要應用的領域，熟悉相關知識，並且選擇整個知識發現程序中處理的相關資料，以建立目標資料集。

(二) 資料前置處理 (Pre-processing)

對重複、錯誤資料做處理，去除錯誤或不一致等雜訊資料。

(三)轉換(Transformation)

將資料的格式、單位、維度等轉換成資料探勘程序中所能處理的資料格式。

(四)資料探勘(Data Mining)

利用資料探勘的技術建立資料模式或找出資料的分類型態。

(五)解釋與評估(Interpretation/Evaluation)

測試與檢驗所挖掘出來的模式，並對此資料模式加以解釋與使用，除能瞭解所找出知識是否有符合原本的要求外，並可透過實際的應用，讓探勘出來的資料模式成為企業的重要知識。

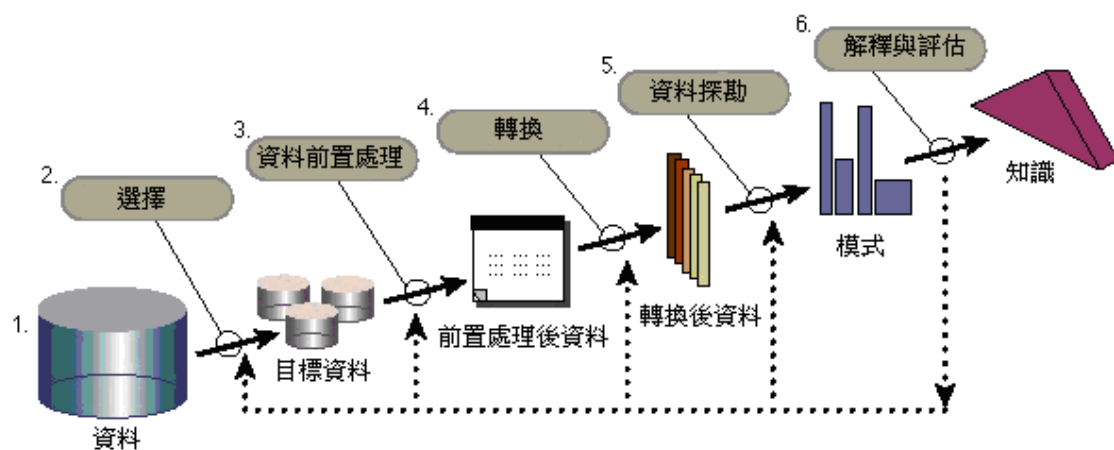


圖 2-1 知識發現之流程(The KDD Process)

資料來源：Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of Data. *Communications of the ACM*, 39(11), 27-34.

綜合而言，知識發現是一個循環的程序，一直重複上述步驟，最後才得到一些有用的知識。由此可知，知識發現的是一連串的程序，而資料探勘是其中的一個重要步驟。

二、資料倉儲與資料探勘的關係

依據資料倉儲之父 Inmon(1994)的定義：資料倉儲是一種整合性、主題導向(Subject-Oriented)、因時而易(Time-Variant)以及不常變動(Non-Volatiled)的資料庫，用來收集資訊以支援管理決策。謝邦昌(民 90)也指出：資料倉儲是蒐集來自其他系統的有用資料，存放在一整合的儲存區內。綜合言之，資料倉儲是一個經過整合、處理且容量特別大的資料庫，用以儲存決策支援系統(Design Support System)所需的資料，供決策支援或資料分析使用。

陳致平(民 84)針對資料倉儲的建立提出以下二個步驟：

(一)資料的萃取(Data Extraction)

此步驟是建立資料倉庫最基礎的工作。資料的萃取是將原始資料從原始的資訊系統中抽取出來，經過適當的介面處理就能將需要的資料完整導出，並存放到資料倉儲資料庫，以備資料探勘之用。

(二)資料分類及整理(Data Catalog & Summize)

此步驟是將萃取出的資料依各企業對資訊提供的模式及需求加以分類及整理。將原始資料(Raw Data)取得後，經過設計的分類及整

理，成為資訊的半成品或成品，以供資訊使用者利用。

由上述可知，資料倉儲本身所含的資料是正確的(不會有錯誤的資料參雜其中)、完整的，而且是經過整合的。因此，要進行資料探勘之前，資料倉儲應先行建立完成，如此資料探勘才能有效率的進行。謝邦昌(民 90)曾妙喻資料倉儲為礦坑，資料探勘就是深入礦坑採礦的工作。礦坑若沒有夠豐富完整的礦產，入坑採礦是很難挖掘出什麼有價值的寶藏。相對的，若沒有夠豐富完整的資料，也很難期待資料探勘能挖掘出什麼有意義的資訊。由此可知，資料倉儲與資料探勘密不可分，兩者的關係可敘述為：資料探勘是從巨大的資料倉儲找出有用資訊之一種過程與技術。

三、知識發現、資料倉儲與資料探勘的關係

綜合前述可知，知識發現的是一連串的程序，而資料探勘是其中的一個重要步驟，此步驟是從巨大的資料倉儲找出有用資訊之一種過程與技術，圖 2-2 清楚顯示三者之關係。

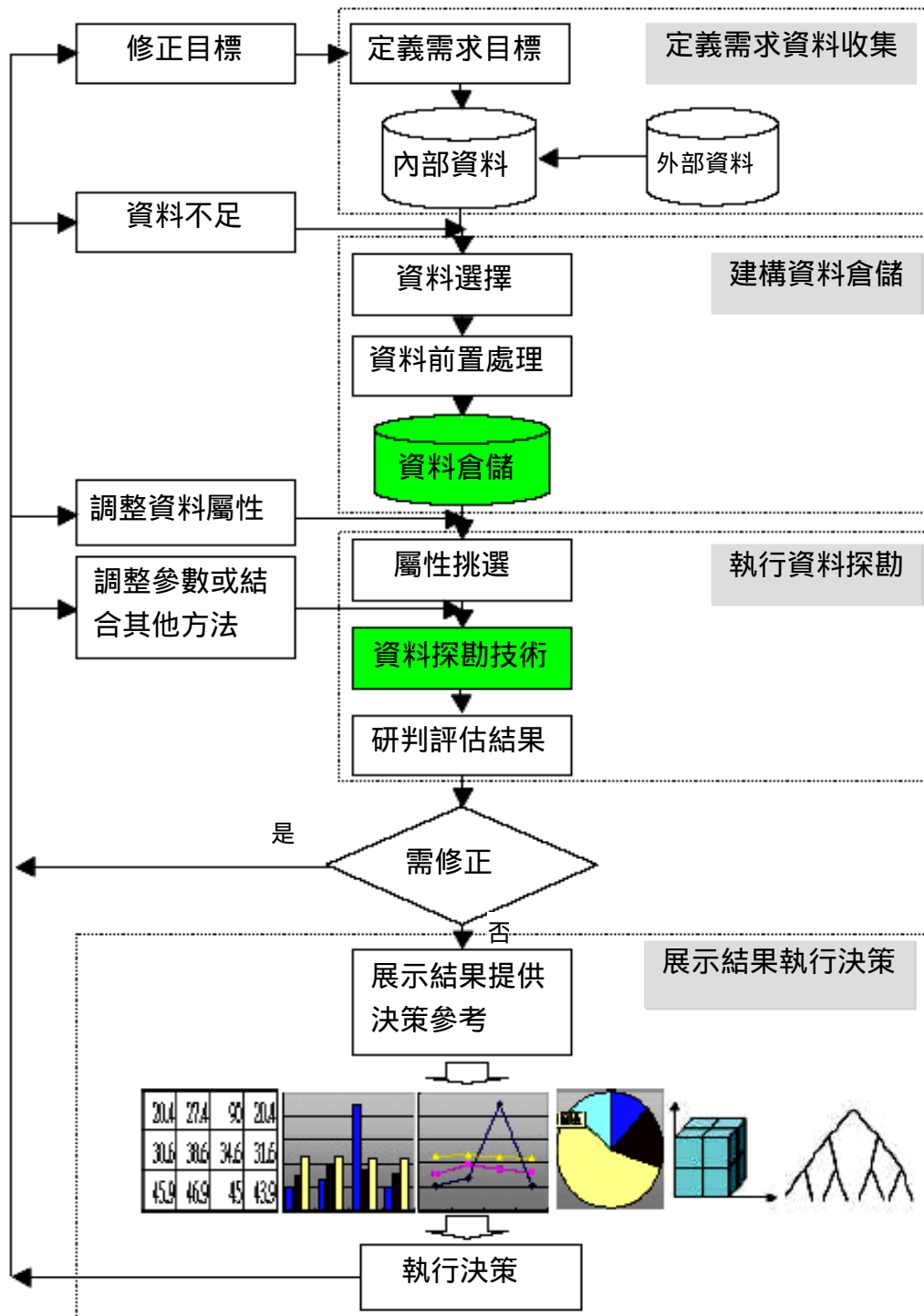


圖 2-2 整合資料倉儲及資料探勘的知識發現流程

資料來源：謝文雄(民 91):以資料探勘探討顧客消費之行為。雲林：雲林科技大學資訊管理學系碩士論文。

參、資料探勘的流程

最早應用資料探勘技術的 IBM(1998)公司，將資料探勘流程分為以下四個步驟，如圖 2-3 所示。

- 一、選取輸入資料：此步驟是指定要探勘和分析的資料，資料來源可能從一個或多個資料庫中，取得表格、概略表或記錄文字檔。
- 二、轉換資料：此步驟是將收集的資料作整理、清除重複或無效的資料記錄，並且確保資料的完整。
- 三、資料探勘：此步驟是依探勘的任務類型，使用相關的理論技術從轉換後的資料中發掘存在的特徵及資訊。
- 四、解釋結果：此步驟是以文字及圖形顯示結果，並進一步解釋結果，提供決策人員參考。

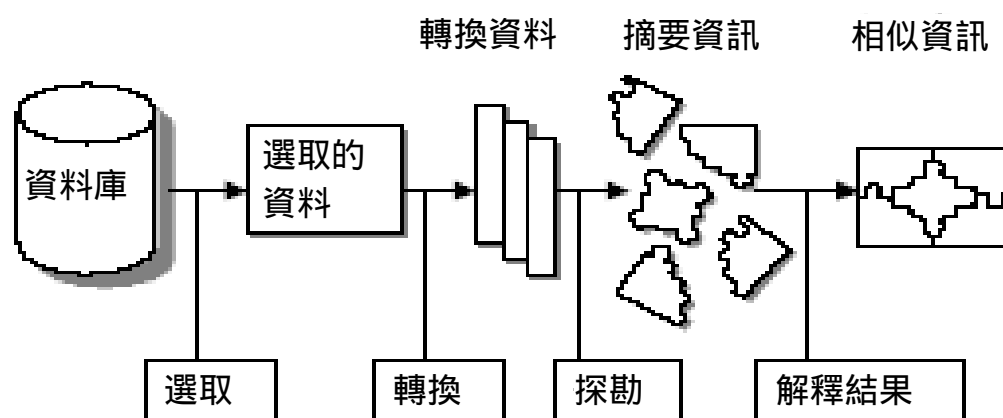


圖 2-3 IBM 的資料探勘流程

資料來源：黃彥文(民 90)。資料探勘之應用 - 會員消費特徵之發掘
屏東：屏東科技大學資訊管理系碩士論文。

肆、資料探勘的技術與功能

一、資料探勘的技術

依據學者專家的看法，資料探勘是結合統計方法、機器學習、資訊科學、資料庫、資料視覺化以及其他相關領域的理論，發展出各種分析技術(Hall, 1995)。換言之，資料探勘不是一種新技術，而是結合多項專業領域的技術研究。就技術面而言，資料探勘的技術可分為傳統技術與改良技術，簡述於下：

(一)傳統技術

傳統技術以統計分析為代表，舉凡統計學內所含之敘述統計、機率論、迴歸分析、類別資料分析等皆屬之，尤其以多變量分析中用來分類的鑑別分析(Discriminant Analysis)、用來綜簡變數的因素分析(Factor Analysis)及用來區隔群體的分群分析(Cluster Analysis)等，在資料探勘過程中特別常用(謝昌邦，民 90；駱至中、林錦昌，民 91)。

(二)改良技術

在改良技術方面，應用較普遍的有類神經網路(ANNs)、多元適應性雲形迴歸分析(MARS)、模糊邏輯(fuzzy logic)、決策樹理論(Decision Trees)、購物籃分析(Market Basket Analysis)、規則歸納法(Rules Induction)、基因演算法(genetic algorithms)、連結

分析(Link Analysis)、記憶基礎理解(Memory-Based Reasoning, MBR)、整合二種以上技術的分析方法...等(呂奇傑, 民 90; 謝昌邦, 民 90; Berry & Linoff, 1997)。

二、資料探勘的功能

運用上述各式理論技術, 資料探勘因不同的功能或任務, 建立各種的模式。而關於不同功能所建立的各種模式, Fayyad(1996)、Berry 與 Linoff(1997)以及 Jason(1998)提出不同的看法, 整理於表 2-1。

表 2-1 資料探勘的功能

學者/ 年代	功能(任務、模式)			
Fayyad (1996)	預測 (Prediction)	關連 (Association)	群集 (Clustering)	序列特徵 (Sequential)
Berry 與 Linoff (1997)	類別	監督式資料探勘 (Directed Data Mining)		非監督式資料探勘 (Undirected Data Mining)
	項目	分類 (Classification)		關聯 (Association)
		推估 (Estimation)		群集 (Clustering)
		預測 (Prediction)		描述 (Description)
Jason (1998)	類別	分類區隔類	推算預測類	序列規則類
	項目	分類 (Classification)	迴歸 (Regression)	關聯 (Association)
		分群 (Clustering)	時間序列 (Time-series)	循序發現 (Sequence Discovery)

資料來源：本研究整理

其中 Berry 與 Linoff(1997)依資料探勘不同的功能, 將其分為

監督式資料探勘和非監督式資料探勘。監督式資料探勘的目的是利用現有的資料建立模型，藉此來描述特定變數；而非監督式資料挖掘中，並沒有指定目標變數，其目的是找出所有變數中存在的關係。

Jason (1998)所提出的六種功能，較能涵蓋其他學者的看法，將其摘要並整理於表 2-2。

表 2-2 資料探勘功能摘要表

類別	功能項目	摘 要
分類 區隔類	分類 Classification	<ul style="list-style-type: none"> ◇根據一些變數的數值做計算，再依照結果作分類。 ◇用一些根據歷史經驗已經分類好的資料來研究它們的特徵，然後再根據這些特徵對其他未經分類或是新的資料做預測。
	分群 Clustering	<ul style="list-style-type: none"> ◇將資料分群，其目的在於將群間的差異找出來，同時也將群內成員的相似性找出來。 ◇與分類不同的是，分析前並不知道會以何種方式或根據來分類，所以必須要配合專業領域知識來解讀這些分群的意義。
推算 預測類	迴歸 Regression	◇使用一系列的現有數值來預測一個連續數值的可能值。可利用 Logistic Regression 來預測類別變數。
	時間序列 Time-series	◇用現有的數值來預測未來的數值。與 Regression 不同的是，Time-Series 所分析的數值都與時間有關。
序列 規則類	連結 Association	◇找出在某一事件或是資料中會同時出現的東西。如果 A 是某一事件的一種選擇，則 B 也出現在該事件中的機率有多少。例如：如果顧客買了火腿和柳橙汁，那麼這個顧客同時也會買牛奶的機率是 85%。
	循序發現 Sequence Discovery	◇Sequence Discovery 與 Association 不同的是，Sequence Discovery 事件的相關是以時間因素來作區隔。例如：如果 A 股票在某一天上漲 12%，而且當天股市加權指數下降，則 B 股票在兩天之內上漲的機率是 68%。

資料來源：本研究整理

三、資料探勘分類功能運用的技術

Berry 與 Linoff(1997)的研究顯示，資料探勘的各種技術，並非對所有的功能類型皆能適用，必須依據分析資料的特性以及所要達成的目的來決定。換言之，不同的資料探勘功能必需選擇適合的探勘技術。

依據資料探勘功能之分類規則，顧客流失區別模型的建構屬於分類類型。由文獻整理所得，適合於資料探勘分類功能運用的技術有：鑑別分析、迴歸分析、類神經網路、多元適應性雲形迴歸、記憶基礎理解、連結分析、基因演算法、整合型的分析技術...等(呂奇傑，民 90；許俊源，民 90；謝昌邦，民 90；陳麒文，民 91；葉怡成，民 91；許哲瑋，民 91；Berry & Linoff，1997)。

由上述的分類功能所運用的技術中，本研究選取傳統技術的鑑別分析，改良技術中多元適應性雲形迴歸、類神經網路以及整合多元適應性雲形迴歸、類神經網路兩項技術的分析方法。此四種資料探勘分類功能所運用的技術將在下一節及第三章第四節中進一步說明。

伍、資料探勘的應用

資料探勘技術應用的領域十分廣泛，其應用領域擴及行銷、財務、銀行、製造廠、通訊、保險、體育...等產業，以下就數位學者對資料探勘的應用說明做一整理。

Frawley(1991)的研究顯示，資料探勘技術主要的三種應用方式都在行銷領域方面，分別是：

一、顧客特性分析(Customer Profiling Analysis)

透過此分析能找出顧客的一些共同的特徵，藉此預測哪些人可能成為顧客，以幫助行銷人員找到正確的行銷對象。

二、目標市場(Targeted Marketing Analysis)

從現有顧客資料中找出特徵，利用這些特徵在潛在顧客資料庫裡篩選可能成為顧客的名單，做為行銷人員推銷的對象。行銷人員依此名單寄發廣告資料，以降低成本，提高行銷的成功率。

三、購物籃分析(Market-Basket Analysis)

此分析主要是幫助零售業了解顧客的消費行為，例如哪些產品顧客會一起購買；或是顧客在買了某一樣產品之後，在多久之內會買另一樣產品...等。零售業者使用此種分析可以更有效的控制進貨量、庫存量、決定貨品的擺設或是評估促銷活動的成效。

李昇暉(民89)認為，企業應用資料探勘技術在以下六個方向：(一)獲取新客戶；(二)維繫客戶；(三)放棄客戶；(四)購物籃分析；(五)需求預測與目標行銷；(六)交叉銷售與主動銷售。

吳佩珊(民91)將資料探勘在各行業的應用整理成表2-3。

表 2-3 資料探勘在各行各業的應用

信用卡公司	信用卡公司可使用資料挖掘來從事有關購買授權決定、分析持卡人的購買行為、偵測詐欺行為等，成功的案例有美國聯邦銀行及花旗銀行。
零售商	可利用資料挖掘來瞭解顧客購買行為及顧客偏好的資訊，例如：透過購物籃分析或是應用其他技術來偵測收銀員詐騙的行為，成功的案例有 Wal-Mart。
金融服務機構	使用資料挖掘來分析大量的財務資料，建立交易及風險模式來發展投資策略。利用資料挖掘以瞭解顧客貸款活動、調整金融商品以符合顧客需求、尋找新的顧客、及加強顧客服務。
銀行	利用資料挖掘以識別顧客的貸款活動、調整金融商品以符合顧客需求，並用來尋找新的顧客及加強顧客服務。
電話銷售及直銷	電話銷售及直銷公司因使用資料挖掘已節省許多金錢並且能夠精確的取得目標顧客。直銷公司可依顧客過去的購買資料及地理資料來配置及郵寄他們的產品目錄；對於電話銷售公司而言，不但能減少通話數，而且可以增加成功通話的比率。
航空業	對於航空業者而言，瞭解顧客需求已經變得極為重要，主要是藉由資料挖掘獲取顧客相關資訊，來制定因應策略。
製造業	資料挖掘已廣泛的被使用在製造工業的控制及排程技術生產程序，可用來偵測潛在的品質問題，以降低產品不良率的發生。
電信公司	利用資訊挖掘，電信公司可以提供更符合顧客需求的新服務。而電信巨人像 AT&T 和 GTE 正應用這些快速偵測不尋常的行為技術來防止竊打。
保險公司	資訊挖掘能幫助保險業者從大型資料庫中取得有價值的資訊進行決策，讓保險業者更瞭解顧客、並有效偵測保險詐欺。
政府單位	美國聯邦政府利用資料挖掘的技術找出犯罪的關連性及可能性。另外，財政部也利用此方法找出洗錢及詐欺的案件。

資料來源：吳佩珊(民 91)：資料倉儲技術於電子商務環境下顧客關係管理之研究。高雄：義守大學資訊工程學系碩士論文。

由上述文獻可知，資料探勘大部份應用在商業金融、異常偵測等領域，且大部份是運用在顧客關係管理上。資料探勘在體育運動領域的應用並不多見，而以學校游泳會員為研究對象者更是付之闕如。學校雖為非營利單位，但游泳池對外開放的營運，其經營管理的觀念就應該仿效企業，須導入顧客關係管理的觀念，善用資料探勘技術。就資料探勘的應用而言，只要是擁有具分析價值與需求的資料倉儲或資料庫，皆可利用資料探勘相關技術進行分析。現今學校資訊教育發達，各種資料多已資訊化，游泳池之游泳會員資料亦然，運用資料探勘技術分析游泳會員資料，建構游泳會員流失的區別模型，除可使學校在游泳池的經營管理更為完善，亦可做為體育運動領域相關研究之參考。

第二節 資料探勘分類功能所運用的技術

本研究屬於資料探勘的分類功能，運用的探勘技術為鑑別分析、類神經網路、多元適應性雲形迴歸及整合多元雲形適應迴歸與類神經網路兩項技術的分析方法。本節將對鑑別分析、類神經網路、多元適應性雲形迴歸做一探討。

壹、鑑別分析(Discriminant Analysis)

鑑別分析是由 Fisher 於 1938 年所提出來一種劃分群體的技術，

其主要的目的在於計算一組「預測變項」(自變項)的線性組合，此線性組合即為區別函數，再依區別函數對依變項(間斷變項)加以分類，並檢查其再分組的正確率(吳明隆，民 89)。以下就鑑別分析的概念、基本假設與原則及用途與應用做一整理說明。

一、鑑別分析的概念

林傑斌、劉明德(民 91)認為：鑑別分析是在已知的分類情況下，增添新樣本時，可以利用此法選定一種判別標準，以判定如何將新樣本放置於何種族群之中。簡言之，鑑別分析是將新增的樣本，判定至已知分類群體中的分析技術。用一簡單例子說明：已知有兩群客戶，一群為忠誠度高的客戶，另一群為忠誠度低下的客戶。現在有一位新客戶，要判定新客戶可能成為何種客戶的問題即為鑑別分析。Fisher 提出的線性鑑別函數，是解決上述問題的分析技術之一，Fisher 依花卉的各種特徵(如花冠和花萼的色澤長寬等)，利用線性組合方法，將這些多變數資料轉換成單變數資料，再化成單變數的線性組合變數值來鑑別事物的差異，而線性鑑別函數便是該研究中主要的分析工具。

二、鑑別分析的基本假設及原則

吳明隆(民 89)指出，在進行鑑別分析時須符合以下基本假設：

(一)樣本須來自一個多變量常態分配的母群，其共變數矩陣需相等。

(二)自變項(預測變項)必須是連續變項(等距/比率變項)，所屬母體

是一個常態分配母群，而依變項則屬間斷變項。如果預測變項是非連續變項，與進行迴歸分析一樣，應先轉化為虛擬變項。

Tacq(1997)提出使用鑑別分析，在分析階段應把握以下原則：

(一)事前組別分類標準要儘可能有可靠性。

(二)初始分析時自變項之數目不能太少，亦即研究者要從許多不同的特性中搜集統計資料，統計分析的平均數與加權總和會決定自變項的鑑別能力。加權總和即是預測變項間的線性組合，也就是所謂鑑別函數(Discriminant Function)。

(三)如果個別變項與加權總和有顯著的鑑別能力，則可以有效的將觀察值歸類為組別中的一組，此歸類的正確率愈高愈好。

(四)挑選具重要特性又有鑑別力的變項，以最少變數達到高區別力的目標。

二、鑑別分析的用途與應用

鑑別分析的功用和迴歸分析一樣，主要用於解釋與預測兩方面。

張紹勳等人(民 89)將鑑別分析的用途整理如下：

(一)歸類：根據觀察值的特性，將其歸納到某一組別中。

(二)處理分群的問題：鑑別分析可讓使用者將多個預測變數加以線性組合，以進行已知組別的分群處理，並了解線性組合的鑑別能力。

(三)預測：根據鑑別函數所得到的鑑別分數，可預測某一觀察值可能屬於那一個組別或其日後的表現。

(四)鑑別分析可用來決定某一變數在鑑別類別時，其相對重要性以及影響力大小。

(五)鑑別分析可用於不同類別的人員甄選。

由此可知，鑑別分析可應用範圍非常廣泛，舉凡醫學、商業、化學、生物、教育、生理、行銷研究、以及考古學等都有其應用之處。鑑別分析經普遍地應用於各種不同的領域之後，已是一個廣為人知的統計技術，相較於其他相關的方法而言，鑑別分析是最常被使用於分類問題的統計方法之一(呂奇傑，民 90)。例如：Desai,Crook 與 Overstreet (1996)運用鑑別分析在信用卡及銀行領域中，建立信用分數的鑑別模型；Kim 等人(2000)使用鑑別分析對韓國的房地產市場進行市場區隔分析，並預測消費者的購買行為。

三、鑑別分析的優缺點

鑑別分析能廣泛的運用，主要是具有以下的優點：(一)結果容易瞭解；(二)建構好的模式容易再使用；(三)能夠整合預測變數。其主要的缺點有：(一)違反應有假設時的鑑別結果不佳；(二)不易說明每個變數的相對重要性及區隔的演算法；(三)很難使用在時間序列資料的問題上(呂奇傑，民 90)。

貳、類神經網路(Artificial Neural Networks, ANNs)

1943 年心理學家 McCulloch 與數學家 Pitts 兩位學者依據人類腦神經組織特性提出了神經細胞的數學理論基礎模式(MP 模式)，此模式開啟了類神經網路的研究領域，此領域主要的研究是運用電腦來模擬人類的腦神經細胞網路。以下就類神經網路的概念、分類及應用做一整理說明。

一、類神經網路的架構

類神經網路是一種以電腦來模擬人類腦神經細胞網路的科學，也就是人腦的抽象電腦模式，它和人腦相似的地方就是由許多的人工神經元(Artificial Neurons)相互的連結所組成，換言之，類神經網路是一種使用大量簡單的相連人工神經元來模仿生物神經網路的能力的一種計算系統。人類神經網路的架構可分為神經細胞、神經結、神經網路，類神經網路的架構也模擬此架構，分為人工神經細胞、層、網路等三個層次，簡述於下。

(一)人工神經細胞

即為人工神經元或處理單元，是組成類神經網路的基本單位。其基本的運作方式是摹仿生物的神經細胞(圖 2-4)，其模式如圖 2-5。

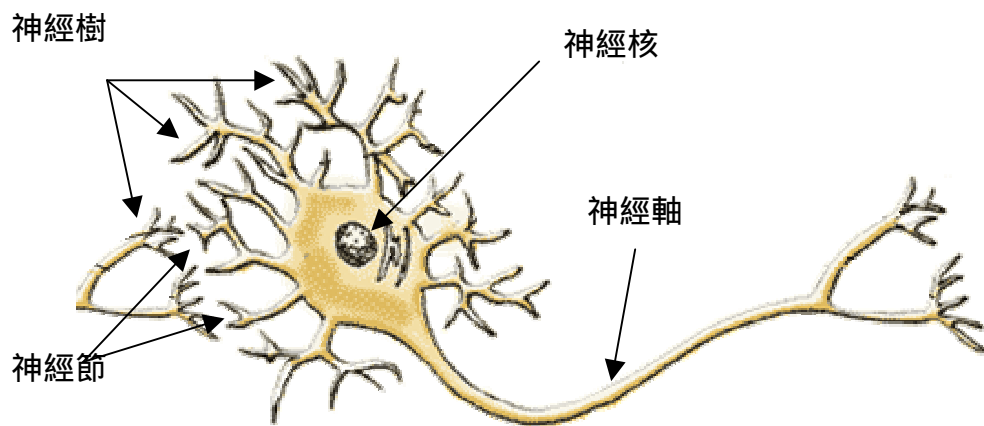


圖 2-4 生物的神經細胞模型

資料來源：<http://www.bud.org.tw/Hu/essay28.htm>

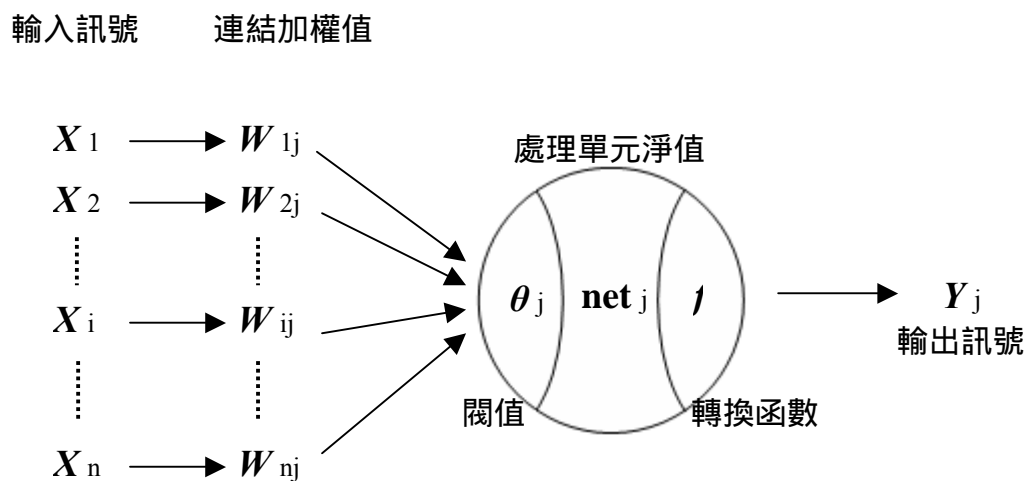


圖 2-5 人工神經細胞模型

資料來源：葉怡成(民 91)：類神經網路模式應用與實作。台北：儒林圖書有限公司。

其中

$X_1 \sim X_n$ 為模仿生物神經細胞模型的輸入訊號。

W_{ij} 為模仿生物神經細胞模型的神經節強度, 又稱連結加權值

θ_j 為模仿生物神經細胞模型的閾值(threshold value)。

net_j 為處理單元淨值。

f 為模仿生物神經細胞模型的轉換函數(transfer function) ;

Y_j 為模仿生物神經細胞模型的輸出輸出訊號。

(二)層

由許多處理單元所組成, 一般分為輸入層(Input Layer)、隱藏層(Hidden Layer)與輸出層(Output Layer)。輸入層處理單元用以輸入外在環境的訊息; 輸出層處理單元用以輸出訊號給外在環境; 隱藏層處理輸入資料間的交互作用。

(三)網路

由若干層集合構成網路架構。網路架構根據處理單元連接的方式可以分為前向式(forward)及回饋式(feedback)。前向式架構的神經元分層排列, 形成輸入層、隱藏層、輸出層, 每一層只接受前一層的輸出作輸入; 而回饋式架構是指從輸出層回饋到輸入層, 或神經元不分層排列而只有一層, 或者是層內各處理單元間有連結者。

二、類神經網路的分類

類神經網路發展至今已有許多模式被提出，Jain, Mao 與 Mohiuddin(1996)依學習策略，將各種模式分為監督式學習、非監督式學習、與聯想式學習等三類網路架構。而葉怡成(民 91)將之分為：監督式(Supervised)、非監督式(Unsupervised)、聯想式(Associate)與最適化應用網路(Optimization Application Network)等四類。在眾多的網路模式中，屬於監督式學習式的倒傳遞類神經網路(Back-Propagation Network, BPN)具學習精度高、回想速度快、理論簡明等特性，在應用上也最為廣泛。根據 Vellido, Lisboa 與 Vaughan (1999)的研究指出，在 1992 到 1998 年間，商業領域中約有 78%的高比例使用類神經網路做為研究方法。因此，本研究以倒傳遞類神經網路做為分析技術之一，此部分將於第三章說明。

三、類神經網路的應用

Lipmann(1987)指出，若問題存在了許多的不確定性，且輸入及輸出間存在複雜的非線性關係，可利用類神經網路解決該問題。又因其具有嚴謹的數學基礎、巨量平行的處理能力、容錯能力、高聯想力、與能過濾雜訊等特性，因此，應用越趨廣泛，在市場區隔、股價指數預測等研究上都有類神經網路的應用。目前類神經網路研究已跨越各領域，到了整合的階段，包括資訊科學、控制、認知科學、神經生理學、心理學、生化、數學、物理、工程等學域。茲將類神經網

路常用模式與其主要應用領域整理於表 2-4。

表 2-4 類神經網路用模式與其主要應用領域

分類	模 式	主要應用領域
監督式	感知機	字母識別
	倒傳遞類神經網路	樣本識別、分類問題、函數合成、適應控制、雜訊過濾、資料壓縮、專家系統、預測
	機率神網路	樣本識別、分類問題
	學習向量化網路	樣本識別、分類問題
	反傳遞網路	樣本識別、函數合成、資料壓縮、分類問題
非監督式	自組織映射圖網路	拓撲映射、聚類問題
	自適應共振理論網路	樣本辨識、聚類問題
聯想式	霍普菲德網路	聯想記憶問題、雜訊過濾、資料擷取
	雙向聯想記憶網路	資料擷取、雜訊過濾
最適化	霍普菲德-坦克網路	組合最適化問題
	退火神經網路	組合最適化問題

資料來源：葉怡成(民 91)：類神經網路模式應用與實作。台北：儒林圖書有限公司。

四、類神經網路的優缺點

類神經網路的主要優點有：(一)其轉換函數為一連續函數，可應用範圍很廣；(二)學習精度很高，可處理高度非線性的函數合成問題；(三)容錯率高、回想速度快。其主要缺點有：(一)學習過程為黑箱作業，無法篩選重要影響變數；(二)網路架構與網路動態的決定缺乏系統化的方法；(三)學習速度過慢、執行時間太長。(葉怡成，民 91；陳麒文，民 91)

參、多元適應性雲形迴歸 (Multivariate Adaptive Regression Splines, MARS)

MARS 是由史丹福大學統計與物理學家 Friedman(1991)所提出，是一種多變量無母數迴歸處理的新方法，目的是解決多元資料問題。以下就 MARS 的概念、功用、應用及優缺點做一整理說明。

一、MARS 的概念

根據 Friedman(1991)的說明，MARS 運用數段基本方程式(Basis Function, BF)加總組合出一個較具彈性的預測模型。其運作原理是根據原始資料本身變數間之交互關係，評估其損適性(Loss of Fit, LOF)，獲得最適合之變數組合。以此組合將資料分成數個區間，並利用各單獨區間的基本方程式做加法運算，構成各區間內的資料行為。各單獨區間內的模型雖然是線性函數，但其斜率皆不相同，使得整體模型變為非線性模型。

二、MARS 的功用

依據陳麒文(民 91)引述 Salford 系統白皮書(<http://www.Salford-systems.com/whitepaper.html#mars>)中的說明，MARS 主要的功用如下：

- (一)自動變數搜尋(Automatic Variable Search)：使用更有效率的演算法檢查全部的變數，並辨認出所有可能的變數。

(二)自動變數轉換(Automatic Variable Transformation)：每一個被選擇進入模型的變數會接受非線性反應的重覆檢查，非線性函數會經由分段迴歸程序中被精確地找出。

(三)自動限制交互作用搜尋(Automatic Limited Interaction Searches)：MARS 會從分析者所允許的交互作用中反覆地搜尋與遞迴分隔系統(Recursive Partitioning Schemes)。不同的是，它可被限制去禁止使用某些型式的交互作用，因此它可允許一些主要作用的變數進入，同時也允許其它變數以交互作用的形式進入，但僅限其他變數的特定子集。

(四)變數分類(Variable Nesting)：只有在符合特殊條件時，某些變數在模型中才是被認為有意義的。

(五)內建的測試訓練(Built-in Testing Regimens)：分析者可以選擇保留一個任意的資料子集來測試，或者使用 V 層交叉驗證的方式來對最後模型選擇的參數做除錯的工作。

三、MARS 的應用

MARS 是一個逐漸成熟的新興方法，且又相當具有可塑性，雖然發展的時間仍短，但許多軟體逐漸支援，相信在未來的應用上有不可限量的成長，目前較被廣泛運用的領域，大部份是資料探勘方面(許俊源，民 90)。茲將相關研究的應用領域整理於表 2-5。

表 2-5 MARS 相關研究應用領域整理表

作者/年代	應用領域
De Veaux, Gordon, Comiso & Bacherer (1993)	地球科學
Friedman and Roosen (1995)	醫學
Bose (1996)	資料探勘技術比較
Nguyen-Cong, Dang & Rode (1996)	遺傳學之基因工程
Griffin, Fisher, Friedman & Ryan (1997)	礦石區分
De Gooijer, Ray and Krager (1998)	匯率預測
陳慧滢(民 83)	進口車、國產車需求預測
許峻源(民 90)	信用卡客戶申請分析
陳麒文(民 91)	健康俱樂部會員流失分析
顏毓靜(民 91)	股票型基金之績效評估
黃明輝(民 91)	債券型基金之績效評估
唐筱菁(民 91)	企業財務危機預測
葉素美(民 91)	匯率預測
陳怡萍(民 91)	IC 產業經營績效
黃正鳳(民 91)	醫學
黃立維(民 92)	資料探勘技術比較
陳怡妃(民 92)	健康俱樂部會員保留分析
陳靜怡(民 92)	失業率預測

資料來源：本研究整理

四、MARS 的優缺點

MARS 主要的優點有：(一)沒有假設限制、具備建構非線性模式的優越性；(二)可篩選重要的影響變數。其主要的缺點是：解釋方程式的數量沒有一定的準則(陳麒文，民 91)。

第三節 顧客流失與資料探勘

由先前資料探勘中的文獻回顧可以發現，資料探勘大部份應用在企業的顧客關係管理上。諸多學者指出顧客流失是顧客關係管理內的一環(夏載，民 90；陳麒文，民 91；Brown, 2000)。由此可知，顧客流失與資料探勘之關係密不可分。本節第一部分探討顧客流失之定義、因素及重要性，以闡明顧客流失之概念；第二部分說明資料探勘在顧客流失管理上的運用，包括執行的步驟及所能達成的目標。分述於下。

壹、顧客流失

一、顧客流失的定義

由相關文獻中發現，對於顧客流失的定義並不明確，過去學者常將顧客保留、顧客流失、顧客忠誠度、顧客轉換與重複購買行為做為研究的構面，探討其相互關係。茲將數位學者之看法羅列於下：

Keaveney (1995) 認為轉換意願是指顧客意圖停止消費目前的品牌或從原先使用的品牌，轉換至其他品牌的一種心理傾向。

Bolton (1998) 以持續期間(duration)為重要變數，瞭解顧客滿意度與顧客保留的關係，結果顯示滿意度的水準高低更能解釋顧客流失行為。

Strouse(1999)認為顧客流失(customer churn)與顧客保留(customer retention)是一體兩面，顧客保留最大化即是顧客流失最小化，降低顧客流失就是要降低其轉換意願與行為。

Gerpott, Rams 與 Schindler (2001)認為顧客保留是有關維持老顧客與供應商間商務關係的建立。

王正雄(民 92)認為顧客的滿意度會影響顧客的再購行為，再購率的高低顯示出顧客的忠誠度，而忠誠度直接影響顧客留在企業內的意願。

綜合上述，顧客流失、顧客轉換與顧客保留三者的內涵相同，受顧客的滿意度與忠誠度的影響，而其衡量標準為顧客的再購率。因此，顧客不再重複購買，或終止原先使用之服務即是顧客流失。

二、顧客流失的因素

瞭解顧客流失的因素，能讓企業瞭解更多過去流失客戶的輪廓，有助於企業鎖定即將可能流失的現有客戶，從而建立顧客流失預警模式，使企業能夠在情況未持續惡化之前及早得到預警，讓企業能針對即將可能流失的客戶做挽留的行動。

顧客流失的因素眾多，Berry 與 Linoff (吳旭志，賴淑貞譯，民 90) 將其分為兩種類型：自願的(Voluntary)離開與非志願(Involuntary)的離開。自願的離開包括：顧客搬離服務區域、顧客

過世、顧客沒能力負擔開支、顧客被競爭者引誘走、顧客換工作、顧客退休、顧客生重病等因素；非自願的離開包括：顧客已經好幾個月沒付帳款、顧客付不出帳款等因素。

佐藤知恭(蕭宏誠譯，民 91)指出，顧客離去的因素以企業對顧客的漠視居第一位佔 68%；對產品不滿意居第二位佔 14%；競爭企業的影響居第三位佔 9%。

柳靜慧(民 92)指出，顧客流失率提高的因素有：

- (一)生活型態轉變導致需求的改變，例如結婚、死亡...等。
- (二)因為經濟景氣的波動，造成顧客消費能力的改變。
- (三)企業無法滿足顧客的新需求或需求的轉變。
- (四)顧客發現更好的選擇與解決方案。
- (五)顧客對服務感到不滿意。
- (六)粗魯無禮的第一線服務人員。
- (七)企業所提供的產品與服務和顧客的預期有差距。
- (八)消費流程讓顧客覺得不方便。
- (九)企業帶給顧客不良的消費經驗。
- (十)遭遇競爭者毫無理性的殺價競爭。

綜合上述可發現，顧客流失的因素大多是企業和顧客間互動不良所產生，然而這些因素並非顯而易見的，但結果卻會造成新客戶滿意

度欠佳，舊客戶忠誠動搖，嚴重者可能轉而他去，以至於企業也因此而蒙受損失。在學校游泳池的經營管理上亦是如此，游泳會員停止消費的因素常是隱而不見的，然而游泳會員的流失，除造成收益上的實質損失外，如流失的因素未能及時發現修正，更可能造成學校形象負面的影響。而資料探勘正是瞭解顧客流失問題的有效工具，運用資料探勘技術建構分析模型，找出顧客流失的可能因素，針對發掘的因素儘早採取因應措施，達到將顧客流失率降至最低的目的。

三、降低顧客流失的重要性

顧客的流失對企業存活影響深遠。以獲利來說，降低顧客流失比起經營規模、市場佔有率、單位成本等競爭優勢更具有效果。由以下研究顯示，降低顧客的流失亦即保留舊顧客遠比開發、吸收新顧客更顯其重要。

Kotler(1994)的研究顯示，吸引或開發一位新客戶所花的成本要比留住一位舊客戶大 5 倍之多，且保留舊顧客獲利率為開發一位新客戶之 16 倍。

Keaveney(1995)的研究顯示，企業一但無法留住顧客，將因顧客的轉換行為而造成企業成本負擔加重，不僅失去原有顧客的收益，且需花費更多的成本去尋求新的顧客以取代原有的顧客。

Zeithaml, Berry 與 Parasuraman (1996)的研究整理出顧客流失

對公司所造成的各種財務損失，其中包括為了爭取新顧客所需要的費用支出，包含廣告、促銷、銷售成本等等，而且新爭取的顧客並不會馬上對企業的營收有所助益，通常需要等到第三、四年後，業者才能回收以前所投資的成本。

Bolton(1998)指出，若能有效延緩顧客流失的速度，預估企業將可增加 15%的長期收益。

Ganesh, Arnold 與 Reynolds(2000)的研究中也顯示，高流失率會對企業的利潤與競爭力產生極大的負面影響，若能降低顧客流失率，公司的營收將會大幅增加。

李昇敦(民 89)根據研究調查指出，一般人認為受到不好的服務時，96%的人會默默地離去，其中 91%的人日後絕不會再光顧，而企業卻只能聽到 4%不滿顧客的抱怨；調查中也顯示，一位不滿的顧客會把他的抱怨轉告八到十個人，其中的 20%甚至還會超過二十人。

由上述可知，降低顧客流失率，對提高利潤及節省成本開支確實有實質上的助益。有鑑於此，學校運用資料探勘降低游泳會員的流失率，除能增加長期的利潤以挹注校務基金外，也能幫助學校建立良好的口碑與正面的形象，進而提升學校的競爭力。

貳、資料探勘在顧客流失管理上的運用

對企業而言，顧客快速流失是企業獲利可預期衰退的警訊，而資

料探勘技術是解決此顧客流失問題的重要方法之一。企業如何將資料探勘運用在顧客流失管理上，以下就其執行的步驟及所能達成的目標，分述於下。

Berry 與 Linoff(1997)指出，企業運用資料探勘從事顧客流失管理可從下列四個步驟著手：

一、定義企業問題，確認機會點

當企業發現具價值的客戶流失造成了企業損失，此時「降低流失率」變成為有價值的機會點。機會點可以從企業各階層中去挖掘及觀察，特別是一些在執行以及管理上所面臨的困境；另外透過與企業重要人物的訪談，也是確認機會點可行方法。

二、運用資料探勘技術，將資料轉換為足以行動的資訊。

顧客流失管理的重要任務，是如何找出最有可能流失的客戶，其中一個方法便是檢視以往流失的記錄，並從中歸納出可能流失的原因；另一個方法是經由問卷的方式來找出流失相關因素。透過了解流失原因，可幫助確認哪些資料對探勘分析能產生價值。透過資料探勘技術，結合相關資料以及企業機會點，產生能夠執行的行動方案。

三、根據資料探勘結果，採取因應行動。

將探勘結果轉化為行動，可根據企業機會點以及資料探勘分析結果，擬訂幾種可行措施，並選擇最佳方案付之實施。從事資料探勘最

終目標，是將資料探勘結果整合於企業客戶關係管理(CRM)機制，透過 CRM 解決方案，減少執行活動所需的時間，使企業反應變得更靈敏也更主動積極。

四、評估模型的效能

評估提供持續改善探勘結果的回饋機制。評估不僅只是回覆率、成本、平均數或者是標準差而已，它涵蓋了所有企業價值。資料探勘是一個循環工作，藉由預期的結果與真實狀況相互比較，同時也能激發下一個工作循環的機會點。每一個資料探勘的成果無論成功與否都將會是未來進行後續工作時的珍貴資產。

依據上述步驟，企業運用資料探勘發展顧客流失模型能達到下列目標（ARC 遠擎管理顧問公司，民 90）：

一、短期目標：

提供潛在流失的顧客名單，支援行銷或顧客維繫活動。

二、中期目標：

建構有效流失模型，建立企業流失管理應用(Churn Management Application, CMA)自動化程序。CMA 主要作業包括建構及管理模型、提供資料分析環境、輸入資料的整理與轉換、產生流失資訊。

三、長期目標：

整合流失模型於企業顧客關係管理，使顧客流失管理成為顧客關

係管理系統的一環。

綜合上述，企業依據資料探勘的執行步驟，發展顧客的流失模式，瞭解客戶的流失問題，並訂定顧客流失管理的目標。如此，有助於加強企業與客戶之間的聯絡、溝通及互動，進而提高客戶忠誠度，降低流失率。以此觀念，應用在學校的游泳會員流失管理上，建構游泳會員流失模式，了解游泳會員的流失特徵。如此，學校游泳池更能有效的管理其資源，處理游泳會員的意見，明白游泳會員的需要，建立學校與游泳會員間共同的價值與尊重，以一種可信賴的關係，幫助學校厚植穩固的客戶群。

第四節 相關文獻之探討

關於資料探勘應用於顧客流失的相關研究並不多，以下就相關的研究整理羅列於下：

蔡永恆(民 89)使用 IBM 所開發的 Intelligent Miner 資料探勘工具軟體，分析國內一家銀行的 ATM 自動櫃員機交易記錄，並以記錄中顧客所使用過的金融服務進行顧客分群，進而分析各群顧客的特性。研究的結果有助於銀行業找出不同顧客群的消費特性及合適的金融商品，以增加顧客滿意度，達到保留顧客的目的。

宮政達(民 90)以國內一家加油站現有資料，針對顧客價值做保

留分析，其中包含流失率分析、異常交易分析與佔有率分析等。在流失分析方面，利用決策樹表示流失狀況，所使用的特徵值有會員層級、消費頻率、及消費週期，結果預測流失之準確度達 84.8%。

邱義堂(民 90)運用 C4.5 決策樹歸納技術並配合多專家決策 (Multi-Expert Strategy)分類方法，對國內一家行動電話業者客戶的大量通聯記錄進行分析，探索客戶退租前通話行為變化，從中尋找流失徵兆。實證結果，利用其流失模型，選取 10.03%的全體客戶，平均可預測出 50.64%的「會流失客戶」；選取 29.00%的全體客戶，平均可預測出 68.62%的「會流失客戶」。

吳坤泉(民 91)運用羅吉斯迴歸(logistic regression)分析國內一家行動電話業者客戶的人口統計資料及使用資料為，探索客戶流失前後通話行為變化，從中尋找流失徵兆，並進一步對流失用戶作問卷電話訪談。研究主要結論有：一、每月平均通話費、平均每通話費與使用量對流失與否關係顯著；二、性別與租用期間對用戶流失與否的影響並不顯著；三、申裝原因與用戶流失與否關係顯著。

蔡明憲(民 91)運用流失模式學習器及策略模組建構器的混合式的架構，對國內一家電信業者之客戶歷史紀錄資料進行分析。結果顯示，已建構完成的流失模式學習器大約有百分之八十五的正確性，已建構的策略模組目前並無適當資料加以評估，

陳麒文(民 91)利用資料探勘分類技術中的鑑別分析、羅吉斯迴歸、類神經網路、多元適應性雲形迴歸等方法，建立中興健身俱樂部之顧客流失分析模式，並由顧客流失分析模式來瞭解流失顧客之重要特徵。研究結果：一、由資料結構中瞭解中興健身俱樂部顧客的組成結構；二、多元適應性雲形迴歸的整體分類績效最佳達 86.52%；三、最佳流失分析模式所找出的流失顧客特徵為：年齡介於 30 35 歲、會齡為一年到二年、月繳 2,500 元的月費且以現金支付費用的會員。

王秀育(民 92)運用不同的資料探勘技術，建構無線電信業者客戶流失預測模式，並經由效能比較，提出一個較佳的預測模型。結果顯示：一、利用客戶的基本資料、合約/服務狀況、通話明細、及客服相關資料等所建置的模式，能有效地達到準確的預測。二、資料探勘技術的選擇，決策樹或類神經網路，均可達到不錯的預測準確度。

王正雄(民 92)利用遺傳規劃法，分析直銷化妝品市場顧客的基本資料及銷售歷史資料，並將顧客終身價值納入了評估指標，由顧客終身價值的高與低，判斷顧客流失機會的高與低。研究結果成功地建構了所有顧客流失率高的模型與高淨值顧客流失率高的模型，這兩個模型的產生，可以成為企業在擬定行銷策略時的一項參考。

陳怡妃(民 92)利用鑑別分析、類神經網路及整合類神經網路與多元適應性雲形迴歸，建構國內某健康休閒俱樂部的顧客獲利性保留

模式。結果顯示，整合模式整體鑑別正確率最高、誤置成本最低且模式運算時間最短，為最佳顧客鑑別模式。

柳靜慧(民 92)運用分群方法中的 K 平均值法(K-means)與自我組織映射圖(Self-organizing Maps, SOM)，分析電信顧客的基本資料與合約資料。結論如下：K-means 在分群效果上是有其一定的分群能力，而 SOM 的分群結果比 K-means 的分群效果更好，經過測量增益值驗證，SOM 在流失客戶的預警功能上特別顯著。

綜觀上述文獻可發現，資料探勘運用在顧客流失的相關研究，大都出現在最近四年內，是一新興的研究議題。從文獻中不難發現，研究對象多以電信業為主，體育運動領域只見二篇相關著作，顯示目前運動相關產業對於資料探勘的應用程度還不高。茲將國內運用資料探勘在顧客流失分析上之相關研究，整理如表 2-6。

表 2-6 應用資料探勘於顧客流失相關研究摘要表

作者/ 年代	論文題目	研究 行業別	資料探勘技術/ 工具軟體
蔡永恆 民 89	應用資料挖掘技術研究 銀行顧客消費行為	銀行業	●IBM 所開發的 Intelligent Miner
宮政達 民 90	顧客關係管理系統設計 與實作 - 保留分析	加油站	●決策樹
邱義堂 民 90	通信資料庫之資料探勘 客戶流失預測之研究	電信業	●C4.5 決策樹 ●多專家決策分類法

(續下頁)

(續上頁)

吳坤泉 民 91	行動電話顧客流失行為探討	電信業	●羅吉斯迴歸
蔡明憲 民 91	以混合式資料探勘技術強化客戶保留之工作	電信業	●流失模式學習器及策略模組建構器的混合式架構
陳麒文 民 91	健康休閒俱樂部顧客流失分析模式之研究	健康 休閒 俱樂部	●鑑別分析 ●羅吉斯迴歸 ●類神經網路 ●多元適應性雲形迴歸
王秀育 民 92	資料探勘於電信業客戶流失管理之應用	電信業	●決策樹 ●類神經網路
王正雄 民 92	應用遺傳規劃法在顧客流失模式之研究 - 以直銷化妝品業為例	直銷 化妝品	●遺傳規劃法
陳怡妃 民 92	資料探勘顧客保留分類模式之建構 - 以健康休閒俱樂部為例	健康 休閒 俱樂部	●鑑別分析 ●類神經網路 ●多元適應性雲形迴歸 ●整合類神經網路與多元適應性雲形迴歸
柳靜慧 民 92	顧客流失預警模型之研究 - 以電信產業為例	電信業	●K 平均值法 ●自我組織映射圖

資料來源：本研究整理

第五節 本章總結

綜合前述內容，總結如下：

- 一、資料探勘的定義：為知識發現其中的一個重要步驟，主要是從資料倉儲中，利用多種技術結合成自動或半自動的處理程序，將有效的、事前未知的以及潛藏有用的資訊抽離出來，並找出資料間重要的特徵與模型，再以這些資訊做為重要決策的參考。其流程

為：(一)選取輸入資料；(二)轉換資料；(三)資料探勘；(四)解釋結果。

二、資料探勘的技術：可分為傳統技術與改良技術，傳統技術以統計分析為代表；改良技術方面，應用較普遍的有類神經網路、多元雲形適應迴歸、模糊邏輯、決策樹理論、規則歸納法、購物籃分析、基因演算法、整合二種以上技術的分析方法...等。

三、資料探勘的功能類型：分類、分群、迴歸、時間序列、關聯、循序發現。本研究將游泳會員區分為已流失和未流失兩類，並以此建立分類模型，因此，本研究屬於分類功能類型。

四、資料探勘分類功能運用的技術有：鑑別分析、迴歸分析、類神經網路、多元適應性雲形迴歸、基因演算法、整合型的分析技術...等。本研究選取傳統技術的鑑別分析，改良技術中多元雲形適應迴歸、類神經網路以及整合多元雲形適應迴歸與類神經網路的分析方法。

五、鑑別分析的優點有：(一)結果容易瞭解；(二)建構好的模式容易再使用；(三)能夠整合預測變數。其主要的缺點有：(一)違反應有假設時的鑑別結果不佳；(二)不易說明每個變數的相對重要性及區隔的演算法；(三)很難使用在時間序列資料的問題上。

六、類神經網路的主要優點有：(一)其轉換函數為一連續函數，可應

用範圍很廣；(二)學習精度很高，可處理高度非線性的函數合成問題；(三)容錯率高、回想速度快。其主要缺點有：(一)學習過程為黑箱作業，無法篩選重要影響變數；(二)網路架構與網路動態的決定缺乏系統化的方法；(三)學習速度過慢 執行時間太長。

七、MARS 主要的優點有：(一)沒有假設限制、具備建構非線性模式的優越性；(二)可篩選重要的影響變數。其主要的缺點是：解釋方程式的數量沒有無一定的準則。

八、現今資料探勘大部份應用在企業的顧客關係管理上，其他方面的應用，只要是產業擁有具分析價值與需求的資料庫，皆可利用資料探勘相關技術進行分析。現今學校各種資料多已資訊化，游泳池之游泳會員資料亦然，運用資料探勘技術分析游泳會員資料，建構游泳會員流失的區別模型，除可使學校在游泳池的經營管理更為完善，亦可做為體育運動領域相關研究之參考。

九、顧客流失的定義：顧客流失、顧客轉換與顧客保留三者的內涵相同，受顧客的滿意度與忠誠度的影響，而其衡量標準為顧客的再購率。因此，顧客不再重複購買，或終止原先使用之服務。

十、游泳會員流失的因素：游泳會員停止消費的因素常是隱而不見的，然而游泳會員的流失，除造成收益上的實質損失外，如流失的因素未能及時發現修正，更可能造成學校形象負面的影響。而

資料探勘正是瞭解顧客流失問題的有效工具，運用資料探勘技術建構分析模型，找出顧客流失的可能因素，針對發掘的因素儘早採取因應措施，達到將顧客流失率降至最低的目的。

十一、降低游泳會員流失率的重要性：降低顧客流失率，對提高利潤及節省成本開支確實有實質上的助益。因此，學校運用資料探勘降低游泳會員的流失率，除能增挹注校務基金外，也能幫助學校建立良好的口碑與正面的形象，提升學校的競爭力。

十二、運用資料探勘在顧客流失管理的執行步驟：(一)定義企業問題，確認機會點；(二)運用資料探勘技術，將資料轉換為足以行動的資訊；(三)根據資料探勘結果，採取因應行動；(四)評估模型的效能。

十三、運用資料探勘在顧客流失管理所能達成的目標：(一)提供潛在流失的顧客名單，支援行銷或顧客維繫活動；(二)建構有效流失模型，建立企業流失管理應用自動化程序(三)整合顧客流失模型成為企業顧客關係管理系統中的一環。

十四、學校依據資料探勘的執行步驟，發展泳的流失模式，瞭解游泳會員的流失問題，並訂定游泳會員流失管理的目標。如此，有助於加強學校與游泳會員之間的聯絡、溝通及互動，進而提高游泳會員忠誠度，降低流失率。