國立臺灣大學管理學院資訊管理學系

碩士論文

Department of Information Management

College of Management

National Taiwan University

Master Thesis

運用資料探勘輔助商品分類之需求預測方法

Demand Forecasting Using Data Mining Aided Product Classification

黃聖祐

Huang, Sheng-Yu

指導教授：陳靜枝 博士

Advisor: Chern, Ching-Chin, Ph.D.

中華民國九十八年六月

June, 2009

運用資料探勘輔助商品分類之需求預測方法

本論文係提交國立台灣大學

資訊管理學研究所作為完成碩士

學位所需條件之一部分

研究生：黃聖祐　撰

中華民國九十八年六月

國立臺灣大學碩士學位論文
口試委員會審定書

運用資料探勘輔助商品分類之需求預測方法

本論文係黃聖祐君（學號 R96725031）在國立臺灣大學資訊管理學系、所完成之碩士學位論文，於民國 98 年 6 月 25 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

陳青文　　葉飛卡

林我聰　　蔣明晃

所　　長：陳靜枝

謝詞

本論文之完成，最需要感謝的是恩師陳靜枝教授的悉心指導與督促。在大學時期於陳老師的必修課就能感受到老師治學嚴謹與教學認真。研究所時有幸能成為陳老師的研究生，在老師的循循善誘之下，總能突破思考問題的困境與盲點。老師不僅在論文寫作上給予指導，也不吝惜地分享許多珍貴的生活經驗，讓學生獲益良多。此外，也感謝口試委員蔣明晃老師、林我聰老師與蕭正平老師於口試期間給予學生的建議，使本論文更臻完備。

研究所兩年的時光將隨著本論文的完成而結束，感謝所上全體老師對課程精心的安排與規劃，讓學生在資管領域獲得更加專精的認識與訓練。在此也特別感謝工工所陳正剛老師，他所開設的線性代數與隨機過程兩門課為學生奠下扎實的數學基礎；商研所蔣明晃老師在供應鏈模型這門課更加激發學生對研究領域的挑戰。

感謝研究所全體同學豐富了我研究所的生活，有緣與大家一起努力是我此生的榮幸。感謝實驗室昌佑學長、錫濤學長的關心與鼓勵，並且在遭遇困難的時候不吝傾囊相授。感謝馨梅學妹、佳綺學妹、邦瑋學弟的陪伴與打氣，尤其在撰寫論文苦悶的時候，你們在實驗室的笑語舒緩了我們的壓力。特別感謝心惟同學的協助，經過彼此的切磋與激勵，使我能順利完成論文。感謝女友素瑜的包容與相伴，讓我經過在一起的這段時光不只是在學業上，也在待人處事與感情付出上更加成長。

最後，僅以此論文獻給我最親愛的家人，有你們的栽培與支持，才能成就今天的我，願此刻的喜悅與你們分享。

黃聖祐　謹識

于台大資訊管理研究所

中華民國九十八年六月

論文摘要

作者：黃聖祐　　中華民國九十八年六月
指導教授：陳靜枝　博士
論文題目：運用資料探勘輔助商品分類之需求預測方法

商品分類為所有與管理商品相關資訊活動的核心，每個公司都會為商品分類，用於銷售管理、採購管理、存貨管理等方面。需求預測是需求管理中最重要的功能，過去許多研究提出改進傳統時間序列趨勢預測方法的準確度，其中之一是合併不同商品的銷售記錄，降低資料的變異度，以合併後的資料進行預測再用適當的比率分配給各個商品。合併商品銷售記錄的依據便是商品分類。以往管理者以質性觀點所建立的商品分類架構並非完全適用於需求預測，貿然將銷售發展趨勢差異太大的商品歸為同一類，會導致類別商品的發展趨勢扭曲或模糊。本研究希望以量化觀點輔助調整商品分類架構使其適合需求預測所用。

針對此問題，本研究根據資料探勘分類方法中的距離基礎方法，定義商品之間銷售發展趨勢的相似程度，進而提出一兩階段最佳化目標模式：首先在固定分群數目下最小化群集內樣本間距離總平均；然後比較不同分群數目時，最大化群集間距離總平均。

本研究的最佳化目標函式並非線性，且解集合的型態近似於整數規劃，必須在每個整數點上搜尋，可行解區域大小隨著分類的商品數目呈現指數速度成長，因此無法利用有限資源求出最佳解。本研究提出一啟發式演算法，使上述問題在可接受的時間內找到一趨近最佳解的分類結果。

本研究啟發性演算法主要流程為：在前置作業中，以時間序列分析進行分類所需之資料轉換，然後根據前一步驟的分析結果建構分類階層，並以基因演算法為基礎搜尋最適分類結果。新的分類架構匯入一需求預測學習系統進行預測準確度評估。

最後，本研究實作出此分類架構建立系統，以兩個實際案例進行驗證本研究所提出之方法確實可行且具有效率。經過實驗之後發現，數百個商品若擁有長期的銷售歷史，且具有明顯的長期趨勢與季節性波動，經過本研究所提之方法分析可以有效判別商品之間的異同，並加以群集，提升預測準確度。本研究的適用商品不限產業，也足以應用於供應鏈管理其他功能。

關鍵詞：供應鏈管理、時間序列分析、基因演算法、商品分類、資料探勘、需求預測。

THESIS ABSTRACT

AUTHOR: Huang, Sheng-Yu JUNE 2009
ADVISOR: Dr. Chern, Ching-Chin
TITLE: Demand Forecasting Using Data Mining Aided Product Classification

Product classification is the core of every information activity related to product management. Almost all companies classify their products according to some attributes for different management purposes such as sales, procurement, and inventory control. Within these business functions, demand management is the leading pulling force while demand forecasting is the most critical function of demand management. Previous studies have suggested many ways to improve the accuracy of prediction using traditional time-series analysis with trend, and one of notable techniques is aggregating sales records of individual product. The purpose of sales aggregation is to reduce the data variation, which can then result in a better sales forecast. Product classification can be used as the scheme for deciding which items should be combined into one product class.

Most companies cluster or group their products based on qualitative features such as brand, color, package, etc, even though for different purposes. The sales trends might be distorted or become unremarkable if the products are carelessly clustered together. This study aims to cluster products by analyzing their quantitative characteristics, namely sales pattern, and make it more suitable for demand forecasting. This study defines the similarity of sales pattern among various products by adopting distance-based method in data mining, and furthermore develops a two-phase optimization model: starting with a given number of groups, minimizing the average distance within groups, then looking for maximization of average distance among separated groups through incrementing number of groups assigned.

Because of the non-linear nature of objective function, integer programming is a popular way to solve the problem. However, when the number of items to be classified increases, the size of feasible solution set grows exponentially as well and makes the problem insolvable due to the time and computing resource it requires. To conquer the difficulty, this study proposes a heuristic algorithm, called Data-Mining Aided Product Classification (DMAPC).

DMAPC first analyzes sales records using time-series analysis and transfers them into a number of indexes which can best describe their patterns. Then, DMAPC searches the optimal product grouping result using GA-based heuristic and the extracted indexes from first stage. A demand forecasting learning platform is used in the final stage. In order to show the effectiveness and efficiency, a prototype was constructed and tested to demonstrate the power of DMAPC using complexity and computational analysis.

Keywords: Data Mining, Demand Forecasting, Generic Algorithm, Product Classification, Supply Chain Management, Time Series Analysis.

目 錄

圖 目 錄

表 目 錄

第一章　緒論

第一節　研究動機

「分類(classification)」即是將一群觀察對象依照設定的架構進行指派，使得他們各自屬於某一個群集。這些群集的標籤代表研究者所關心的類別屬性(class attribute)，可以是進行指派前已經定義，或是研究者根據該群集所包含的樣本進行解釋再給予。而分類任務的用途在於新出現的觀察對象可以使用一樣的架構將之歸類，藉此協助決策；或是做為其他分析活動中的步驟，為後續研究提供增益，改善最後的結果。

分類活動持續應用於人類生活中各個領域。生命科學中區辨生物是否屬於任何已知的物種[21]；醫學中判斷患者的症狀屬於哪種病源引起，依此判斷施予療程[16, 24]；商業活動中，銀行根據申請人過去的信用記錄來決定是否核准其貸款申請，產品經理人判斷具有哪些特質的消費者會購買產品[9]；行銷計畫中，若能選擇正確的客戶寄發促銷廣告訊息則能提高回覆率並達成行銷計畫目標[5, 26]。

在商業應用裡，商品分類不僅與資料庫設計息息相關，也幾乎是所有使用與管理商品資訊活動的核心[14]。隨著使用者的觀點不同，會用不一樣的屬性去描述該項商品[28]：以行銷的觀點，商品屬性會包含銷售區域、出貨日期、售價、商品描述等等；以生產的觀點，商品屬性則會包括原物料號、生產時間、生產成本等

等。不一樣的商品屬性往往會導致相異的商品分類架構，因此分類架構有不一樣的用途與分析目標。一個好的商品分類架構能夠使分析得出的結論更具有意義，甚至能讓消費者更容易找到心目中最適合的商品並進行購買。

學者 Chopra 等等在[6]中表示，預測需求是所有供應鏈中策略與規劃決策的基礎。Sheikh 在[27]中所提出的生產資源規劃(Manufacturing Resource Planning, MRP II)架構中亦指出，需求管理(Demand Management)是所有規劃模組的起點，如圖 1-1 所示。不論是主生產排程(Master Production Scheduling, MPS)、物料需求規劃(Material Requirements Planning, MRP)，或是採購運輸排程的研究，都假設擁有良好的需求管理，也就是準確的銷售預測，然後才宣稱可以透過他們的研究與提出的模式方法達成整體供應鏈管理的目標。

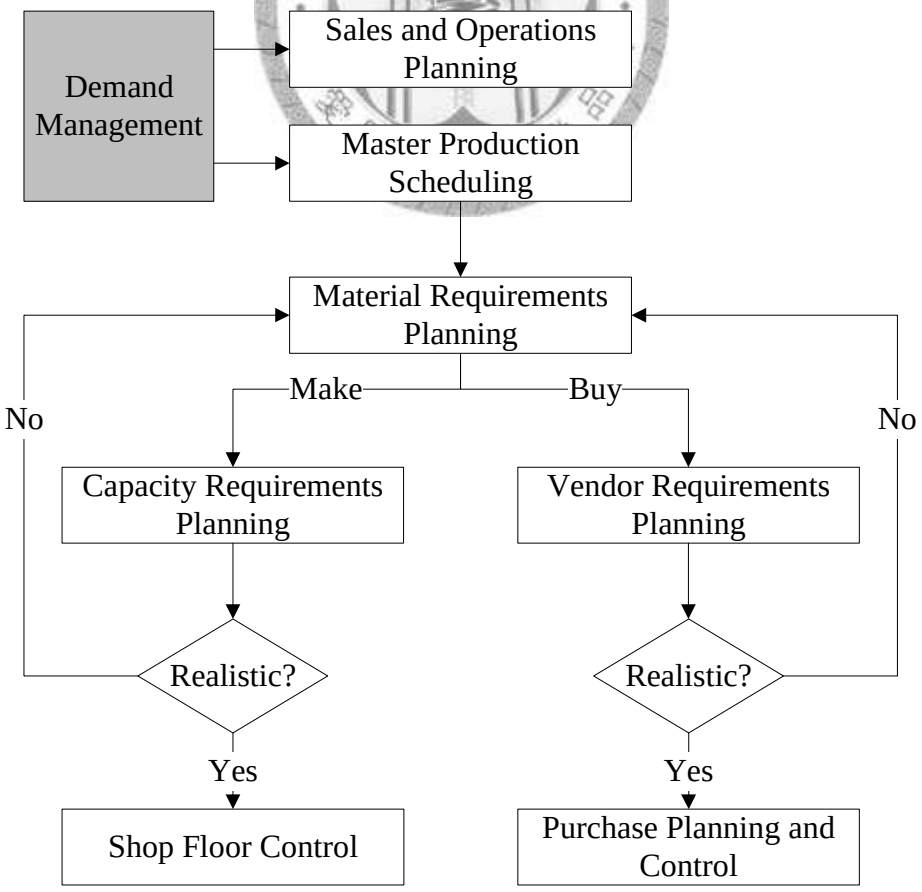

圖 1-1：MRP II Framework

(資料來源：[27]部分修改)

丁在[1]所整理的銷售預測改進方法其中之一為：將眾多商品的銷售記錄依商品分類架構加以合併，降低銷售資料的變異性，合併的銷售記錄經學習系統判斷後，使用選定的預測方法進行預測，最後將分類商品的銷售預測數量依適當比率往下分配給單項商品。這樣的銷售預測模式有一個關鍵的步驟就是「依商品分類架構整合銷售記錄」。任意的商品分類架構並非都適合銷售預測所用，他可能是廠商的專家或行銷業務人員依照自己的專業所設定的分類，僅僅適合財務分析或生產管理所需。貿然採用不適當的分類架構進行銷售記錄整合，若誤將銷售發展趨勢相反的商品置於同一商品分類，彼此互相抵消之後會造成該商品分類整合的銷售發展趨勢產生扭曲的現象，往後也將因此產生錯誤的預測，導致整體銷售預測模式的正確性與有效性低落。

如圖 1-2 所示，同類商品 A、B 的銷售記錄分別呈現成長與衰減的趨勢。若以商品各自的趨勢進行預測，會得出 100 單位的預測銷售量給商品 A，50 單位的預測銷售量給商品 B；若將兩商品的銷售記錄合併，則會得到一條水平的趨勢線，以此趨勢進行預測會得出維持 150 單位的預測銷售量，再依個別商品在觀察期間內總銷售量佔類別商品總銷售量的比例進行分配，商品 A、B 會各得到 75 單位的預測銷售量，反而低估了商品 A 的趨勢，高估了商品 B 的趨勢。

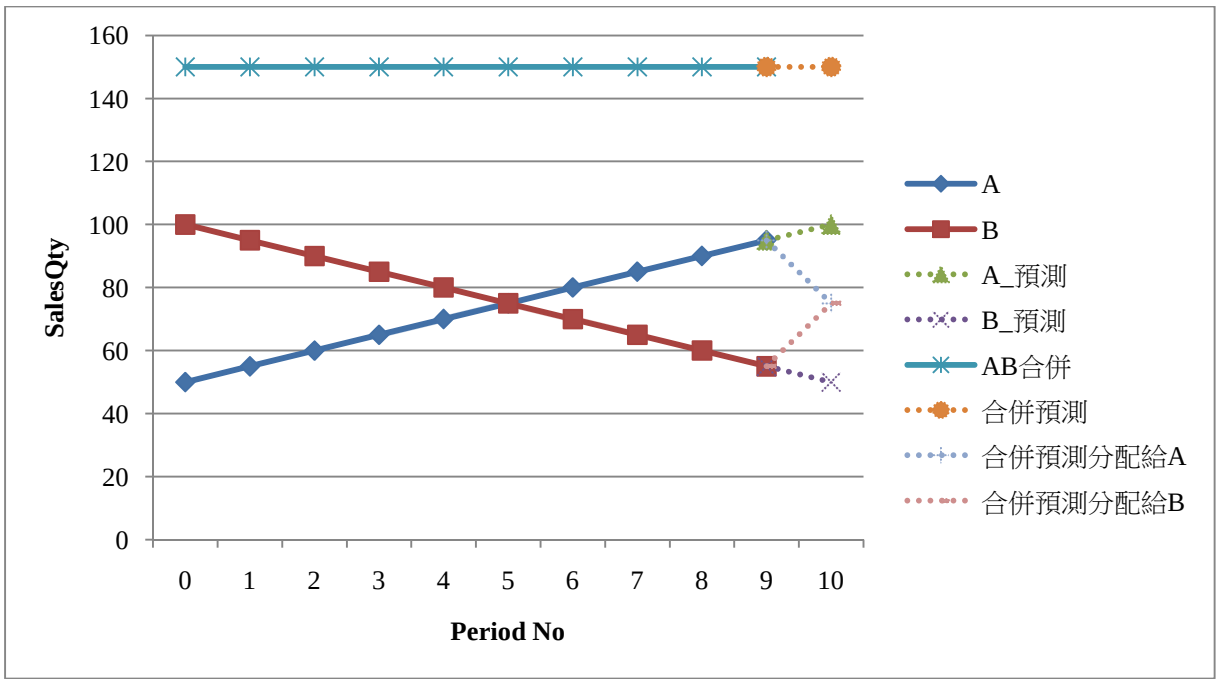

圖 1-2：合併銷售記錄之不良效果範例

(資料來源：本研究整理)

近年來由於電腦硬體運算與儲存科技已發展穩定且廣泛為企業組織所採用，資訊網路與資料庫技術的發展與佈署，更進一步擴大企業所能收集的資料廣度與深度，大幅加速資料的累積，包括每天的商業交易活動記錄、客戶與供應商資訊、產品資訊；儲存內容的格式也不限於資料庫記錄，也包括影像、聲音、3D 物件模型等等多媒體格式。

此外，隨著網路的興起，電子化的出版品數量開始快速增加，例如線上新聞、學術研究論文、電子書、電子郵件訊息、企業內部文件、網頁內容、部落格(blog)文章，與線上討論區的意見。這些數不盡的電子化資料構成龐大的文字資料庫，相較於商業資料庫中的記錄，被視為半結構化資料(semistructured data)，例如每份文件都有標題、作者等等必須記載的資訊，但是摘要與內容這類文字區塊則屬於非結構化，不同文件所使用的描述方式沒有固定的格式。

資料探勘(data mining)意指從大量且未經處理的資料中發掘珍貴的知識，作為一種有效的分析工具與自動化流程，解決「資

料充足卻資訊貧乏(data rich but information poor)」的處境[9]。而文字探勘(text mining)的興起，協助研究者處理結構化程度較低的文字資料，從龐雜的文件海中篩選出相關者，再進行字詞索引、關鍵字出現頻率等等分析步驟，然後與資料探勘方法連結，完成從文件資料中發掘知識的任務。

使用資料探勘技術來完成資料分類的任務，學者們已經提出許多方法[15]，包括以下幾種類別：邏輯基礎演算法(logic based algorithms)、類神經網路基礎方法(perceptron-based techniques)、統計學習方法(statistical learning algorithms)、實例基礎學習法(instance-based learning)、支援向量機(Support Vector Machines)。這些不同類別的方法為了要處理多樣化的資料類型與不同的研究目的所需，各自發展出迥異的演算法基礎。此外，基因演算法(genetic algorithm)、粗略集合(rough set)、模糊邏輯(fuzzy logic)等等方法的加入，在不同方面上改善只使用一種方法的分類結果。

商品除了有基本資訊可用於資料分類之外，來自零售商的銷售量記錄對銷售量預測的目標而言更為重要。這些隨時間不斷累積且具有時間序列(time series)特性的資料是以往任何知識發掘(knowledge discovery)方法所沒有考慮過的。不論是資料探勘或文字探勘的方法，都是收集大量屬於不同個體的樣本，例如不同的客戶、不同的生物、不同消費者的購物記錄、不同文件，而不是同一個觀察對象在記錄中不斷出現，以表現其發展趨勢。

因此，要發展一套適合需求管理的商品分類架構，除了將各商品連續性的銷售表現納入考慮，以求聚集在一起的商品不會扭曲模糊該分類的銷售發展趨勢，同時也要考慮商品之間的關係，例如是否高度相關、相依，還是互補商品，甚至是新商品取代生

命週期已經到退出市場的舊商品。

本研究欲針對上述問題引入資料探勘的自動化方法與理論，提出一套新的方法模型，協助管理者發展適合銷售量預測與需求管理的商品分類架構，達成提高商品銷售量預測準確度的目標。

第二節　研究目的

本研究試圖針對商業資料庫中的商品資訊進行分析，以期建構出一套分類架構，將商品分類。本研究特別針對商品銷售記錄，包括商品基本資料、商品描述、門市營業行事曆、門市營業據點資料、門市商品每日銷售量、特殊事件記錄等等。銷售量會因為市場的機制與消費者的反應所產生的購買行為而在記錄上產生上升或下降的趨勢表現；長期發展趨勢則可能表現出季節性的成長與下跌或是更長期的循環表現；也可能受特殊事件影響，例如節慶、促銷活動、社會新聞事件，而產生短期異常的發展趨勢導致短期缺貨的問題。商品與商品之間的關係是互補、互斥、還是替代的關係都會影響該商品在分類架構中的歸屬。

已知商品銷售預測改進模式中，將單項商品的銷售記錄整合成分類商品銷售記錄可以降低資料的變異性，取得顯著的發展趨勢，而不適當的商品分類架構會造成發展趨勢迥異的單項商品被歸類為同一分類，使得分類商品的銷售發展趨勢受到扭曲或抵銷的效應。因此，從時間序列資料中辨認各項商品銷售模式的相似相異程度並加以區分是本研究的重點。

同時，本研究亦不能完全捨棄廠商所提供的商品分類標籤與各項屬性與描述，從中可以分析得出商品之間的關係，並導出具有解讀意義的分類規則，一方面，如果未來有新商品的加入可以快速歸類，另一方面也較容易為使用者所理解接受。

過去並未有研究針對調整商品分類架構使其適合供應鏈成員進行多樣商品的整體需求管理與銷售預測，僅有學者提出商品分類與消費者線上選購產品的關係[14]，或是研究銷售預測卻只針對少量或單一商品。本研究將發展一套方法，適用於前述的複雜商品銷售發展趨勢與資料型態，以供零售商進行準確的銷售預測，進而達成存貨成本下降，甚至將預測資料分享給上游製造商、配銷商，提升整體供應鏈的效率。

第三節　研究範圍

本研究的最終應用目標屬於需求管理中的銷售預測，商品分類架構的建立屬於銷售預測模式中最先開始的步驟。因此本研究針對供應鏈末端的零售商，以一般商品流通業為主要研究對象，目標商品必須具有充分的歷史銷售資料。

本研究不探討流行性商品或易受促銷活動影響的商品，亦即不分析流行期間的起訖與量化效果幅度；促銷活動與預期之外新聞事件爆發對商品銷售量的影響視為不規則變動，在分析的過程中以其他方法去除。

本研究假設所收集之歷史銷售資料具有良好的品質，即不存在錯誤或不完整的資料。零售商擁有足夠的資訊科技建設與能力，能記錄每日的商品銷售記錄，並且透過網路連線將各分店記錄匯集至一整合資料庫，呈現單一商品在最小預測時距中全區銷售量總和的觀點。

第四節　研究架構

本研究之架構與流程，如圖 1-2 所示，說明如下：

一、文獻探討：

蒐集整理和資料探勘、商品分類、需求預測等相關研究之定義與方法論文獻。

二、問題定義：

依據相關文獻對此類型產品分類問題之定義，界定本研究之研究範圍、定義和相關之假設與限制條件。

三、模式建立：

根據本研究問題之定義和假設限制，以數學方式建構出此問題模式。

四、演算法建構：

根據相關文獻之研究心得，建構出解決此問題之演算法。

五、模式與演算法驗證：

驗證模式與演算法的正確性和適用性，並做適當的修正。

六、模式分析：

實際導入台灣產業以驗證本研究模式之可行性，並驗證其成果。

七、結論與建議：

歸納總結研究結果，討論研究中仍有的限制和未來可以研究的方向與建議。

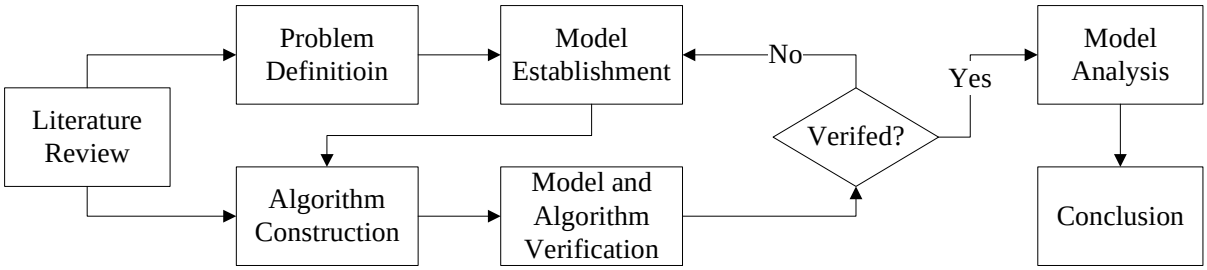

圖 1-3：研究架構

(資料來源：本研究整理)

第二章　　文獻探討

第一節　資料分類之定義

根據 Han 等學者在[9]中的定義，分類是一種資料分析的任務，透過分類演算法建立分類模型或分類子(classifier)，用來預測研究者對於觀察目標所關心的類別標籤(categorical labels)，例如一件貸款申請是「安全」還是「有風險」，一個消費者會「買」還是「不買」一件特定的商品，一個前來求診的患者應該接受「處方 A」、「處方 B」還是「處方 C」；這些類別可以用離散的數值來表示，其數值順序與大小並無比較上的意義。「預測(prediction)」有時會與分類在字詞上混用，但是在此將之定義為專用於預測連續型數值函式(continuous-valued function)，例如預測商品銷售量。兩者藉由預測目標的資料類型加以區分。

分類包含兩個主要的階段[9]，如圖 2-1 所示：

一、學習(learning or training)：這個步驟顧名思義是指分類模型將透過分類演算法從一群事先準備好的資料中學習而得，分類模型通常會以分類規則來表示，而分類規則是一群「若則(IF-THEN)規則」的集合。這群資料被稱為訓練集(training set)。每一筆樣本(tuple)以一個 n 維的屬性向量(attribute vector)表示，例如 $X = (x_1, x_2, \cdots, x_n)$。同時有另一個資料庫屬性被指定為類別標籤屬性(class label attribute)，擁有離散且無序的數值代表研究者所關

心的類別。因為每一個訓練集樣本的類別標籤屬性值都是已知，因此這個步驟也被稱為「受監督的學習(supervised learning)」。

二、分類：已經建構好的分類模型在這個步驟會經由另一群稱為測試集(test set)的樣本進行準確度檢驗。測試集的樣本通常是隨機從一般的觀察群體中抽取，並且不同於訓練集中的樣本，以避免得到過分樂觀的準確度結果。經過測驗通過研究者所設定的準確度門檻的分類模型，終於能用於預測類別標籤屬性值未知的新樣本；反之，若未達到要求的準確度，則必須返回前一個步驟重新進行學習，採用新的分類演算法或不同的訓練集樣本。

有別於上述的定義，叢集分析(clustering analysis)也是一種分類的方法[9]，但是在學習階段所使用的訓練集樣本並不具有已知的類別標籤，憑著分類演算法中的相似度計算將樣本聚集成不同的群集，被稱為「未監督的學習(unsupervised learning)」。經過叢集分析所得到的分群結果將不是以準確度作為最終的評量標準，而是以可解釋性與作為其他延伸分析的基礎是否帶來良好的效果來判斷其方法的優劣。

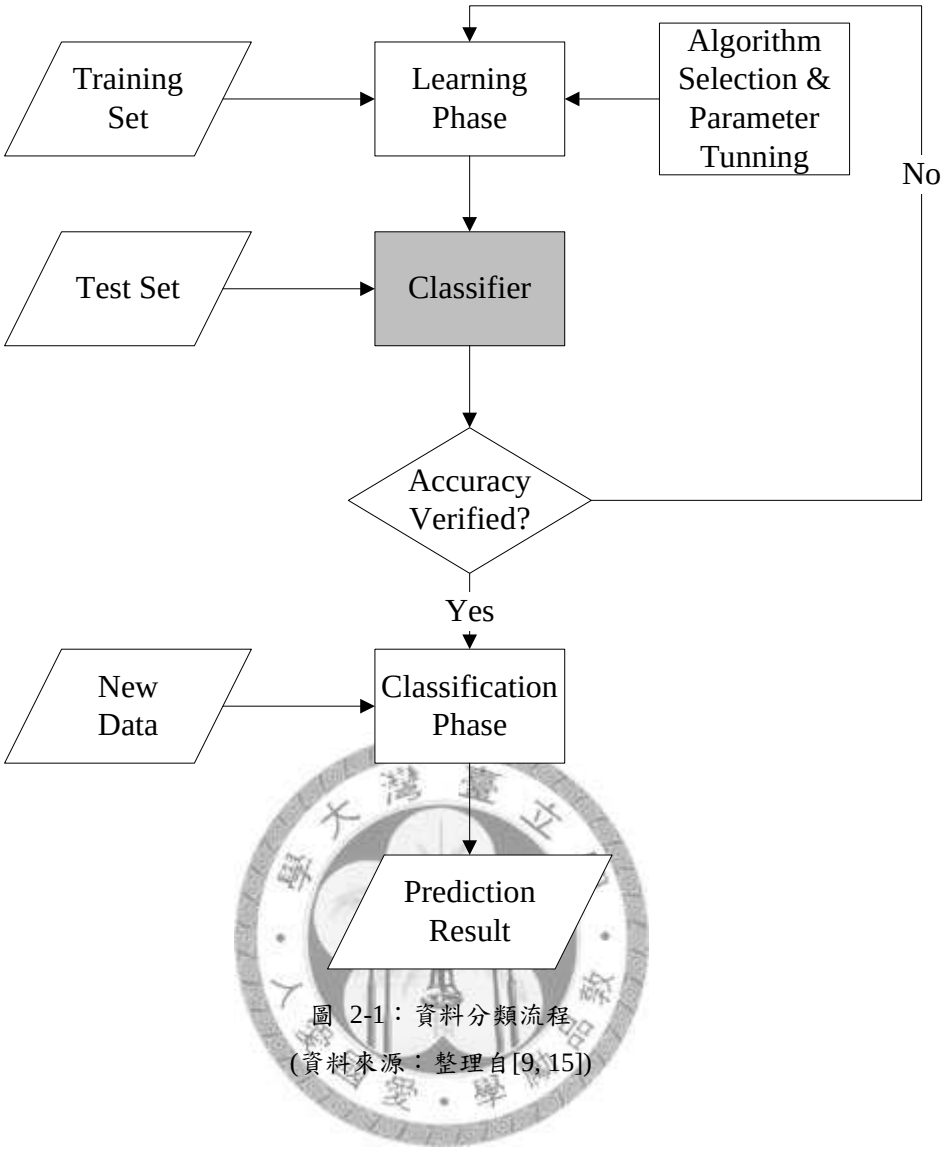

圖 2-1：資料分類流程

(資料來源：整理自[9, 15])

第二節　資料分類之議題

2-2-1 分類之資料準備

在正式將所收集到的樣本引入資料分類分析流程之前，樣本資料可以經過下列前處理(preprocessing)的步驟，以提升分析流程與結果的品質[9]。

一、資料清理(data cleaning)：

這個步驟的目的在於去除或減輕雜訊(noise)對資料分析的影響。樣本資料在收集的過程中，可能因為使用者或資訊系統流程的疏失，導致某些屬性欄位數值有缺漏或

是不合理。這些品質不佳的樣本稱之為雜訊，在資料前處理階段可以將之去除，但是如果樣本數量稀少或取得成本較高，較不適用；或將缺漏的屬性欄位填入最有可能的數值，但是此做法必須謹慎控制避免對分析結果造成扭曲的影響。多數的分類演算法都有掌控雜訊的機制，但是經過清理的資料仍然可以減低學習階段的困擾。

二、相關性分析(relevance analysis)：

儲存於資料庫中的資料在平時交易作業的時候經常會記錄相當多樣的屬性，但是真正在分析時會發現許多屬性欄位彼此之間存在高度的相關性。這些與分類模型相關卻無法提供新資訊的多餘欄位(redundant column)，或是根本與分類模型無關者(irrelevant column)[8, 19]必須加以處理。特徵選擇(feature selection)不同於特徵抽取(feature extraction)，前者僅是選出最佳的屬性子集合，後者則牽涉到將屬性轉換並重新組合成新的屬性。有許多學者針對樣本的特徵選擇提出許多研究，希望能增進預測效能、提出更快更有成本效益的分類模型，並且對產生資料的流程提出更好的解釋。

特徵選擇方法有兩大方向[8, 32]：一是獨立於主要的歸納分析演算法之外，用過濾器(filter)的方式去除不相關的特徵；另一個則是在進行歸納演算法的同時，以包裝(wrapper)的方式評估所選擇的特徵子集合，根據適度(fitness)的比較進行調整。Yuan 等學者在[32]提出同時採用兩種方法並結合基因演算法的兩階段特徵選擇方法。Hanczar 等學者在[10]針對基因序列分析中，樣本數很少，但是基因數龐大的特殊情況，提出從相似樣本的群集中

選出代表該群集的原型(prototype)作為資料分類學習階段的訓練集。Swiniarski 等學者在[29]將粗略集合(rough set)與主成分分析(principle components analysis, PCA)引入特徵選擇的方法協助分類模型的建構。Liu 等學者則在[19]中將現存的特徵選擇演算法依照搜尋策略、評量條件、資料探勘任務分做一個三維的分類架構，並建構了一個統一的平台協助對特徵選擇演算法細節無知的研究者選擇適當的特徵選擇方法。

三、資料轉換(data transformation)：

同樣屬於數值資料的欄位，為了避免數量級的差距過大造成屬性之間相對的重要性受到扭曲，透過標準化(normalization)將資料轉換至固定間距內[9]，例如將原始數值範圍數萬至數百萬的年收入屬性數值轉換為 0.0 至 1.0 之間。這種資料轉換方法通常適合應用於衡量樣本點之間的距離。另一個方向是普遍化(generalization)，將原本連續的數值資料轉換為離散的範圍類別，例如將年收入依照設定的界線轉換為低中高三種類別。如此一來因為原始的資料受到壓縮，也就降低了分類模型在學習階段的輸入輸出動作，提升學習速度。

本研究假設所收集到的資料不存在屬性值缺漏或可能因為記錄失誤產生品質不佳的狀況，因此可略過資料清理的步驟，將所有資料引入分類模型中。銷售歷史記錄資料的特性為，描述單一資料列的屬性很少，資料筆數相對龐大，所以屬性篩選能提供的助益不顯著。本研究中最重要的資料前處理步驟為，將原始資料轉換為適用於下一章將提出的分類模型。

2-2-2 資料分類之評量

不同的分類方法可以用下列條件來評比[9]：

一、準確性(accuracy)：

代表一個經過指定分類演算法學習而得的分類模型正確預測一個未見過的樣本的類別標籤的能力。準確性其實是透過測試而得的估計值，為了提升估計值的可信度，可以使用多個測試集交叉驗證(cross validation)，又因為是估計值的關係，也可以引入統計的方法給予信心區間(confidence interval)的描述[7]。

二、速度(speed)：

指該分類模型的建構與使用上的運算複雜度。任何演算法都必須考慮速度，若是一個分類模型需要超過可接受的時間來運算，即使準確度很高，也會因為喪失時效性而變得不適用。

三、強健性(robustness)：

指該分類方法在雜訊樣本的干擾下，正確預測類別標籤的能力。雖然在資料前處理的時候可以將雜訊刪除或修正，但是那畢竟是手動去更改所收集到的樣本資料，若是能從演算法的角度證明雜訊不會影響預測正確性，更加有說服力。

四、擴張性(scalability)：

指面對大量的資料是否能有效率地建構分類模型。現在的商業資料庫中所累積的資料數，至少都以千筆為單位，甚至數萬數十萬，發展有效處理如此巨量的資料甚至是在運算階段會超過電腦主記憶體容量的分類演算法[18]，是目前學者們研究的重點。

五、解讀性(interpretability)：

指經由分類演算法所建構的分類模型是否容易為研究者所理解，描述模型的規則是否具有解讀的意義，而不是變成一個只能使用而無法提供研究者新的知識的「黑盒子」。解讀性是一個相當主觀的評量條件，有時候研究者只在乎分類結果是否提供後續應用分析良好的基礎。

以上條件經常會發現彼此衝突的情況，例如準確度很高，分類模型卻難以解讀；模型的建構速度很快，分類準確度卻偏低。隨著研究者所期待的分類模型的功能與應用目的不同，上述的評量標準也就不需要每一項都達到最佳化。本研究所在意的重點是速度與擴張性，因為本研究所需要分析的資料量相當大，而且預想中，引入更多的資料能得出更適用於銷售預測的分類結果。

第三節　資料分類之方法

2-3-1 統計基礎方法

貝氏分類模型(Bayesian classifiers)採用統計中的貝氏理論，計算一個樣本在已知的屬性表現下屬於各類別標籤值的條件機率，當一個樣本屬於某一個類別標籤值的機率大於屬於所有其他類別標籤值時，貝氏分類模型就將該樣本歸類為那個類別[9]。同時，為了降低計算複雜度，天真貝氏分類模型(naïve Bayesian classifier)假設樣本屬性之間互相獨立，但是現實上屬性是有相關性存在，此時可以貝氏信念網路來表示。

當樣本的屬性符合貝氏分類模型的假設，而且相關的機率資料可以取得的情況下，貝氏分類模型與其他分類模型相比擁有最小的分類錯誤率以及最短的分類模型建構時間與分類時間[9]。但

是因為樣本屬性之間經常具有相關性，在為這些假設鬆綁的同時，貝氏分類模型的整體準確度會下降，若為了確保假設而進行屬性轉換則會使得分類方法複雜化失去快速分類的優點。

本研究所針對的商品銷售記錄資料有高度的自我相似性，同時也是貝氏分類模型所無法處理的連續型數值，所以統計基礎的貝氏分類方法並不適用於本研究。

2-3-2 規則基礎方法

分類規則通常是用於描述分類模型，決策樹(decision tree)是一個適合產生分類規則的方法[15]。一棵決策樹是一個類似流程圖的樹狀結構[9]，有根節點(root node)、內部結點(internal node)，與葉節點(leaf node)，節點與節點之間以分支連結，而且每個子節點只有一個父節點。每個內部的節點代表對指定屬性的測試，每一條分支代表測試的結果，葉節點則代表一個類別標籤的值。一個樣本進行分類的過程即是從根節點開始，依循每個節點測試結果選擇對應的分支不斷往下，直到抵達葉節點取得類別標籤的值。

決策樹的建構關鍵在於每個內部節點應該擺放的屬性測試，選擇的標準是根據每個屬性對分類所能提供的資訊增量(information gain) [9]。決策樹的建構不需要專業領域知識與參數設定，建構與分類過程相當簡單快速，產生的規則也容易被研究者所解讀吸收，而且大致上擁有不錯的準確度，因此獲得廣泛的使用，或當作新分類方法的評比標準。但是決策樹的建構需要品質良好的資料，否則會因為雜訊的影響產生許多小分支，必須經過修剪(tree pruning)才能增進準確度。Carbalho 等學者在[4]提出結合基因演算法的混合決策樹分類方法，有效處理涵蓋樣本數較

少的小分支，抽取出研究者關心的分類規則。

分類規則也可以直接從訓練集學習而得[15]。換句話說，規則是由廣泛至精確的方式搜尋建立，每學習一條新的規則，就將這條規則所涵蓋的樣本移除，然後用剩下的樣本繼續學習。

這些若則關係的邏輯判斷式可以被賦予順序，每個樣本根據符合的規則來決定應該歸屬於哪個類別。這樣的做法必須考慮當一個樣本同時符合多條規則，而這些規則可能導出相反類別標籤結果的衝突情況[9]。

聯合性分類模型(associative classification)是另一種規則基礎的分類模型[9]，將常見的聯合性規則分析應用於資料分類任務上，將每一個樣本屬性與值的配對(attribute-value pair)視為一個項目(item)，在訓練集中尋找頻繁項目集合(frequent itemset)並檢驗其信心(confidence)與支援(support)，藉此建構分類模型。

規則基礎方法必須輸入離散的屬性值，如果該屬性為連續性的數值，也必須先間隔離散化。本研究所使用的樣本資料包含大量的連續型數值，並且擁有時間序列的特性，同一件商品會因為時間不同而有不一樣的屬性表現；另外，本研究所需要的結果並非完整的分類規則，只需要經過分類的商品群集能有效提升後續的銷售預測模式的正確性，因此規則基礎的分類方法不適合單獨使用於本研究。

2-3-3 類神經網路

源自於心理學家與神經生物學家為了在計算模型上模擬大腦神經元所開發出來，類神經網路可以視為一組互相連結的單元(unit)，他們之間的連結有輸入輸出的區別還有不同的權重[9, 17]。類神經網路的學習即是在往前輸入(feed-forward)測試集樣本資料

與往後傳播(backpropagation)錯誤評量的過程中反覆調整每條連結的權重，直到錯誤率降低到研究者所設定的標準以下。

類神經網路長久以來被批評所建構出來的分類模型難以解讀[9]，而且學習時間較長，選擇此方法進行分類模型建構時必須考慮結果的時效性是否重要；但是在另一方面，類神經網路對於雜訊資料的容忍度很高，即使不清楚樣本屬性之間的相關性也可以使用，再加上分類演算法有平行運算的特性，可以加以發揮縮短運算所需的時間。

類神經網路分類模型與規則基礎方法相反，輸入與輸出都需要數值類型的屬性值，如果要處理非連續數值的屬性，也必須加以編碼才能引入分類模型。本研究所要分析的資料型態雖然適用，但是仍然需要加以調整，加速分類模型的建立。

2-3-4 距離基礎方法

將每個量化的屬性視為一個維度，樣本可以被看作分佈在這個樣本空間的資料點[15]，樣本與樣本之間的「距離」也就可以經由研究者定義得出。尤其在叢集分析中，沒有事先標記的樣本經過分群演算法之後，距離相近的「相似」樣本會被歸為同一群。

分割方法(partition methods)[9]在起始時將所有樣本視為一個群體，然後隨機分配成研究者所設定的群集個數，k。樣本經過反覆的重新安排，終於不再於群集之間移動時代表分群結果已達穩定，演算法終止。群集個數 k 值如果沒有專業領域的知識輔助判斷，則需要經由窮舉搜尋(exhaustive search)最適當的值。分割方法適用於建立球體分佈(spherical-shaped)的樣本群集以及中小規模的樣本數。

與分割方法相反，階層式方法(hierarchical methods) [9]在起

始時將每個樣本各自視為一個群集，然後不斷地將最靠近的兩個群集合併在一起，直到所有的樣本都合併為一個群集。合併過程會以樹狀圖(dendrogram)記錄，研究者依照專業領域的知識或窮舉搜尋的方式決定最佳的群集個數。

與前述的分割方法一樣，這兩個方法有一個關鍵在於選擇計算樣本與群集之間距離(linkage)的方式，使用不同的計算方式有時候會導致不一樣的分群結果。

為了消除以樣本之間距離為相似度基礎的方法只能發現球體分佈群集的缺點，在密度基礎方法(density-based methods) [9]中，一個群集會持續增長擴張範圍當週遭的樣本分佈密度高於指定的門檻值，換句話說，在該群集裡的每一個樣本在指定半徑內都會包含一定數目以上的樣本。這樣的方法不僅可以發現分佈呈任意形狀的群集，也可以有效過濾雜訊。

支援向量機(support vector machine, SVM)是一種新興的分類方法[9]，適用於區分在樣本空間內呈線性或非線性分佈的資料，並且以優異的分類準確度吸引研究者的注意與應用。主要的做法是將原始樣本資料經過非線性轉換到一個高層次(維度數目較少)的空間，再尋找一個能區隔訓練集樣本的最佳線性超平面(hyperplane)。

被稱為懶惰學習者(lazy learner) [9]的分類方法並不積極於學習階段建構分類模型，僅將訓練集的資料儲存，直到測試集的資料引入或新樣本的出現，才啟動分類的機制。例如 *k*-nearest-neighbor classifier 會計算新樣本與已經儲存於樣本空間中的訓練集樣本的距離，而新樣本的類別標籤值就與最近的 *k* 個鄰居一致。這樣的做法導致進行分類的運算複雜度與時間相當久，必須依靠平行運算的技術加以改善。

本研究試圖將商品置入未知的分類架構中，這樣的分析任務與叢集分析相仿，因此將採用距離基礎分類方法，將指定期間銷售歷史記錄轉換之後表現相似的商品群集為同一分類。但是本研究所收集的樣本還包含其他可能有助於分類模型建構的資訊，所以需要定義其他離散型屬性值的距離關係，再加以引入距離基礎分類方法。

2-3-5 其他方法

基因演算法(genetic algorithm) [4, 9, 32]的引入，將研究者所期望的可行解編排成如同生物基因的形式，同時考慮多組可行解，然後模擬生物界交配(crossover)與突變(mutation)的方式造成基因序列變換產生子代(新的可行解)，並以適者生存的法則，根據研究者所定義的適度(fitness)作為演算法終止的條件。基因演算法的分類規則搜尋法可能顯得難以控制，不像其他分法是由一個起始解開始漸漸往最佳解的方向搜尋，但是也正因為這種跳動的特性，將創造跳脫區域最佳解(local optimal)而得到全域最佳解(global optimum)的機會，或是避開雜訊樣本的干擾，並且多點搜尋的策略將有效縮短複雜問題的解題時間。適度函式的設計與子代數目的設定是使用基因演算法的重要參數，影響最後所得到的解的品質與求解速度。本研究所必須分析的樣本無論在數量上與屬性維度上都可稱為複雜問題，因此適合引入基因演算法提升分類模型建構時的速度與強健性。

使用粗略集合(rough set)與模糊集合(fuzzy set)的概念能增進分類模型處理不精確的規則，或是不完整的樣本資料的能力[9, 17, 22]。某種程度上提升了分類模型的強健性，但是同時也增加了衝突發生的情況。例如使用模糊集合將原本間隔沒有重疊的屬性數

值判斷條件鬆綁，符合間隔重疊區域的樣本將同時符合多個分類規則，因此有可能導出衝突的分類結果。本研究最後期望的分類結果是每個商品只屬於一個分類，因此不允許模糊集合所產生的衝突結果。

第四節　分類方法與預測

許多學者運用上述的資料探勘的分類方法協助各領域的研究，包括醫學研究中區別不同類型的 DNA 表現[10, 16, 32]、網際網路中網頁內容自動分類[31]、資訊安全中偵測入侵的網路行為[18]，或是商業分析中的資料庫行銷[20]。但是大部份的研究所做的分類僅僅是將觀察對象指派給各個預先定義的類別，然後進行解讀，並沒有延伸應用到數值預測上。

近年來，有學者 Cardoso 等人曾經針對報紙的銷售量預測與各銷售點的供補貨建議[3]，但是僅針對單一商品。本研究所要面對的是流通業多達萬種的商品，無法以單一商品的觀點進行銷售量預測，必須先將商品進行分類，否則將導致無法負擔的計算與時間資源成本。

第五節　預測成果評估方法

應用於需求預測的研究都希望提升預測準確度。因應不同的情境與資料特性，預測準確度有數種量化的評估方式，以下列四種最為常見[2, 13]：

一、平均方差(mean squared error, MSE)：

預測值與已知歷史記錄差之平方平均值。平方的做法雖然避免了高估與低估效應互相抵消的問題，但是也使得

大誤差更為顯著。

二、平均絕對差(mean absolute deviation, MAD)：

預測值與已知歷史記錄差之絕對平均值。計算方法簡單，但是會受觀察對象數量級不同的影響，例如預測數千到數萬之間的誤差與預測數十到數百的誤差，可能前者所得出的 MAD 較大但是並不代表預測準確度較低。

三、平均絕對百分比差(mean absolute percent error, MAPE)：

已知歷史記錄與預測值差之比值絕對平均值。採用比值的作法避免了數量級不同的影響，將誤差絕對值轉換為與歷史記錄的百分比。但是所有使用比值的計算公式都必須考慮分母不得為零的限制，因此採用 MAPE 時也必須注意。

四、最大絕對差(largest absolute deviation, LAD)：

預測值與已知歷史資料記錄之差最大絕對值。與前述三種方法相比，計算方式最為簡單。但是只留下最大絕對值的作法在多數的情境下捨棄了太多資訊，容易導致偏差的分析結論，或受原始資料誤差影響。

依照學者 Kahn[12]的研究顯示，四種比較標準中以 MAPE 最廣受一般企業採用，且皆是使用數量進行計算。本研究將採用平均絕對百分比差(MAPE)來驗證商品分類架構對於銷售量預測準確度的改善效果，其目的在於去除不同商品銷售量數量級之間的差異，並用實際銷售值為分母的公式，以避免過度高估銷售的狀況。

本研究將運用資料探勘的分類方法找出最符合目標的商品分類架構，並且應用於商品銷售量預測分析。資料的前處理步驟中，資料轉換是本研究的重點，使得原始銷售歷史記錄可以適用於下一章將提出的模型。另外，透過上面的整理比較，本研究將

結合距離基礎分類方法，並搭配基因演算法提升最佳解的搜尋效率，最後在各個需要評比的階段使用 MAPE 做為量化的標準。

第三章　問題描述與最小距離群集模型

第一節　問題描述

本研究將分析商品基本資訊與銷售歷史記錄，進而建構一商品分類架構。依此分類架構，每個商品會被歸屬於一個類別，然後同類別的商品銷售記錄會在後續的預測分析中整合形成單一的類別銷售量進行預測，最後再將類別預測銷售量分配給單項商品做為最終的銷售量預測。

此分類架構所預期產生的商品分類是：銷售表現相近的商品會屬於同一個類別。如此將避免進行銷售預測時，在整合分類商品銷售歷史記錄時，誤將銷售表現迥異的商品置於同一分類，造成該分類商品銷售表現的扭曲或產生互相抵銷的效果，然後因此產生錯誤的預測，導致整體銷售預測模式的正確性與有效性低落。

在商品資訊方面，可分為預先定義的資訊，即上市時就已經確定，由供應商所提供的資訊，包括建議售價、銷售地區、廠商定義的商品管理分類架構；與隨著該商品持續出現在門市通路上不斷累積的銷售量歷史記錄，形成一串連續性的時間序列數值資料。

本章之第一節將詳細介紹建構商品分類架構所需的商品銷售資訊，第二節為本研究之假設條件，第三節則提出最小距離群集模式。

3-1-1 銷售歷史記錄

隨時間不斷累積的銷售記錄，在資料庫裡呈現出與一般資料探勘所面對的資料不同的型態，如表 3-1 與表 3-2 對照所示，研究者所關心的對象在資料庫記錄裡通常會擁有不重覆的代碼；但是在銷售歷史紀錄中，若只看商品代碼會發現同一件商品的記錄重覆出現在資料庫中，必須同時參照日期這項屬性才會使得每一筆記錄符合資料庫管理的唯一性(uniqueness)要求。

表 3-1：非時間序列資料表(以客戶資料為例)

CustomerID	Gender	Age	Occupation	Annual Income	…
00001	Female	24	Student	120,000	…
00002	Male	35	Sales	1,500,000	…
00003	Male	44	Manager	3,000,000	…
⋮	⋮	⋮	⋮	⋮	⋮

(資料來源：本研究整理)

表 3-2：時間序列資料表(以銷售歷史記錄為例)

ProductID	Date	SalesQty
00001	2008/12/01	99
00001	2008/12/02	220
00001	2008/12/03	266
00002	2008/12/01	145
00002	2008/12/02	383
⋮	⋮	⋮

(資料來源：本研究整理)

因此，本研究必須先將同一商品但時間不同的銷售記錄整合成單一資料列，使得不同時間的銷售量成為該商品的一項屬性，如表 3-3 所示。若將銷售量與時間作為兩個維度，則可以畫出該商品的銷售量隨時間變化的發展趨勢，如圖 3-1 所示。

表 3-3：時間序列資料表，整合單一商品記錄

ProductID	…	SalesQty	SalesQty	SalesQty	SalesQty	…

		@2008/12/01	@2008/12/02	@2008/12/03	@2008/12/04	
00001	…	99	220	266	192	…
00002	…	145	383	251	133	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮

(資料來源：本研究整理)

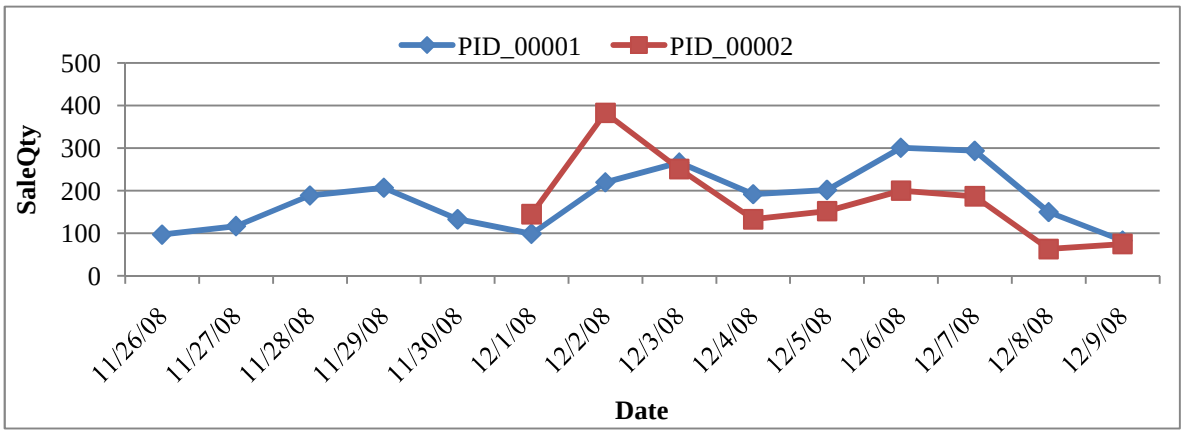

圖 3-4：時間序列趨勢圖

(資料來源：本研究整理)

由上圖可發現另一個問題是，並非所有商品都在同一時間上市，也就造成每個日期不是所有商品都有銷售記錄。以資料探勘的角度來看，上市時間較短的商品就變成屬性值有缺漏，品質不良的資料，無法直接引入分類模型中進行分類。在第二章所提到的資料前處理方法中，有一種方法是將品質不良的資料直接捨棄，但是本研究的目標是要讓所有商品在最後都屬於某一個商品分類，所以不能將任何一個上市時間較短的商品銷售記錄捨棄；另一種做法是手動填補「最有可能」的屬性值，可以是當日其他商品的銷售記錄的平均值，或是出現次數最多的值，但是這種做法等同於盲目將所有商品的銷售表現整合成單一數值，必然會對接受此數值的商品產生扭曲銷售表現趨勢的效果。當然，也不可能只取所有商品都有銷售記錄的期間，若不是完全沒有交集，也會是交集期間很短，無法代表任何一件商品的銷售量表現趨勢。

為了去除上市時間先後不同所造成的歷史記錄資料量不一的影響，本研究採取時間序列趨勢分析方法，將每項商品的銷售量

歷史記錄轉換為下列四種指標：

一、長期趨勢指標(T, trend index)

代表一時間序列資料在整段觀察期間內所展現出持續增加或持續減少的趨勢，如圖 3-2 所示。將銷售量當作觀察值 y，可以寫作一個時間 t 的線性函數(linear function)：

$$y_t = \beta_0 + \beta_1 t + \varepsilon$$

其中的β_0與β_1各自為未知的參數，分別代表趨勢線的截距與斜率。若$\beta_1 > 0$，則表示此一時間序列資料呈現持續增加的趨勢；$\beta_1 < 0$，則表示持續減少的趨勢；$\beta_1 = 0$，則表示沒有長期趨勢。$|\beta_1|$表示長期趨勢的線性遞增(或遞減)幅度，絕對值愈大表示遞增(或遞減)得愈快。ε代表實際歷史記錄與長期趨勢線預測值之間的誤差。但是，長期趨勢中，y 與時間 t 的關係並不一定為一次線性，若發現可能有非線性的趨勢表現，則可寫作

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon$$

則為二次多項式(quadratic function)的關係，其中，β_2代表此一呈拋物線趨勢的方向與幅度：$\beta_2 > 0$表示增加，$\beta_2 < 0$表示減少；$|\beta_2|$表示增加(或減少)的幅度，絕對值愈大表示趨勢愈陡。大多數的研究顯示，商業應用中的時間序列資料最適合用線性函式估計，或是二次多項式[Keller, 2005]，因此本研究也同樣採用這兩種函式。這些未知的參數可以使用迴歸分析(regression analysis)估計得知，最後將依據已知的銷售歷史記錄與迴歸模型所計算得出的估計值的誤差，也就是ε的比較決定採取一次線性或是二次多項式的模型最能代表此一時間序列發展趨勢。

二、循環性指標(C, cyclical index)

長期趨勢並不一定是持續增加或減少，有時候可以發現反覆發生的漲跌趨勢，但是這樣的循環週期長達一年以上，而且週期長度並不固定。循環性趨勢通常是由大環境的因素所造成，例如總體經濟景氣的循環，是以在本研究所收集的銷售歷史記錄中幾乎難以發現，故不予考慮。

三、季節性指標(S, seasonal index)

與循環性指標同樣用來表示反覆發生的漲跌趨勢，但是不同的部分在於，季節性的循環週期可以在一年以內發現，而且每一季的長度是固定的，如圖 3-2 所示。例如每年因為聖誕節與新年所帶起的購物潮，反映在銷售量記錄上成為一個高峰[13]。本研究對時間序列分析模型採用乘法模式(Multiplicative Model)，所以若令商品銷售量為因變數 y，前面所提出的關係式則變成

$$y_t = (\beta_0 + \beta_1 t) \times S_t + \varepsilon$$

季節性指標代表，每一季實際的銷售記錄與趨勢分析的預測值的比值。分析時所採用的「季」不一定是三個月，也可以是一個月，相對應的，採用不同長度的單季期間會使得季節性指標的個數不同，例如採用三個月為一季，則一年有四個季節性指標；採用一個月為一季，則一年有十二個季節性指標。與趨勢分析一樣，最適合的季節性指標可以透過比較不同單季長度設定下所產生的誤差來判斷。

四、不規則性指標(I, irregular index)

商品在市面上販售時，受無法預期的特殊事件影響，例如促銷活動，新聞事件等等，使得銷售量在短期內表現

出非長期趨勢與季節性因素所能解釋的大幅成長或衰減。特殊事件因為發生的時機無法預期，所以效應也就難以掌控，即使是由廠商主動推出的促銷活動，也無法保證促銷效果在設定的期間內與施行的區域上都對商品銷售量產生一樣程度的影響。

綜合以上說明，本研究採用之商品銷售記錄表現特徵將包括長期趨勢指標與季節性指標來代表描述。

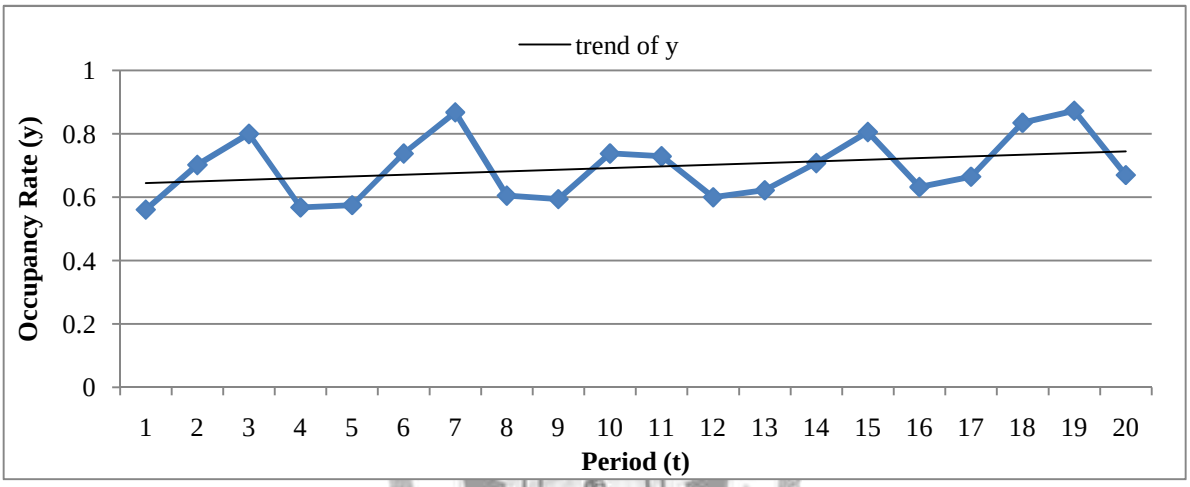

圖 3-5：時間序列資料長期趨勢與季節性

(資料來源：[13])

第二節　假設條件

根據前一節針對商品銷售量歷史記錄的資料轉換定義，可進一步歸納本研究之假設如下：

一、預測時距(prediction time bucket)以週為單位

一般零售商在訂補貨時並非根據每日銷售狀況調整次日的訂補貨量，而且從訂貨到補貨實際上需要前置時間。因此，本研究假設使用者採取固定時距檢查並規畫下一期的訂補貨量，而銷售量預測的最小時距設為週，所有日記錄都將先整合為週記錄再進行分析，預測未來的銷售量最短也是預測一週需求。

二、不存在可分析抽取的不規則性指標

因為特殊事件發生所造成的不規則性影響，其影響期間、影響效果皆難以掌控；此外，本研究所針對的長銷型商品較少有特殊事件發生，所以本研究假設不規則指標不存在，所有可能的變動都透過季節性分析加以去除。

第三節　最小距離群集模型

本節將以數學模型的方式，針對前兩節的問題描述與假設條件，來建構一最小距離群集模型。以下描述建構最小距離群集模型之設定，包括下標、參數、決策變數、限制式以及目標函式。

3-3-1 最小距離群集模型建構流程

本研究建構最小距離群集模型流程，先使用智慧型學習平台選擇最佳的參數計算，將銷售量歷史記錄轉換為時間序列指標。然後將這些指標所形成的特徵向量引入本研究所設計的最小距離群集模型，完成商品分類。

3-3-2 參數部分

P　：表示待分類商品所構成之集合。

$N(P)$　：表示待分類商品數目。

m　：表示欲區分的群集數目。

M_{min}　：表示群集數目之下限。$M_{min} \geq 1$。

M_{max}　：表示群集數目之上限。$M_{max} \leq N(P)$。

G　：表示所有可行的分類結果構成之集合。

G^m　：表示指定分為 m 群所有可行分類結果構成之集合。

G_C^m　：表示前述分類結果中，屬於群集編號 C 的商品所構成

之集合，$C = 1 \sim m$。

$N(G_C^m)$ ：表示前述分類結果中，屬於群集編號 C 的商品數目。

C_i ：表示商品 i 所屬於的類別群集編號，$i \in P$，$1 \leq C_i \leq m$。

B_i ：表示長期趨勢分析中，商品 i 所採用描述長期趨勢的係數個數。根據每項商品所選擇的最適關係式，各自有不一樣的係數個數：若選擇線性，則 $B_i = 2$ $(\beta_0$、$\beta_1)$；若選擇二次式，則 $B_i = 3$ $(\beta_0$、β_1、$\beta_2)$。

K_i ：表示季節性分析中，商品 i 所採用一年的季數。

若單季為三個月，則 $K_i = 4$；若單季為一個月，則 $K_i = 12$。

S_{iq} ：表示商品 i 第 q 期的季節性指標，$i = 1 \sim N$，$q = 1 \sim K_i$。

T_{ib} ：表示商品 i 第 b 個長期性指標，$i = 1 \sim N$，$b = 1 \sim B_i$。

3-3-3 決策變數

TS_i ：表示描述商品 i 銷售發展趨勢的特徵向量。

若使用某個參數設定時，有不需要的屬性，則該屬性值指定為 0。例如使用簡單線性迴歸與四期季節性描述時，T_{i3}，也就是時間序列模型中係數 β_2，與季節性指標 $S_{i5} \sim S_{i12}$ 皆指定為 0。

同時，因為模型中用來描述趨勢線截距的參數 β_0 並不影響後續預測的準確度，因此從特徵向量中移除。又為了避免描述趨勢線斜率的係數數量級過大的影響，導致季節性指標的差異被掩蓋，因此趨勢線斜率的係數必須經過標準化，將係數除以所有待分類商品的係數平均值 $(\bar{\beta}_1$、$\bar{\beta}_2)$，形成一平均值為 1 的長期趨勢指標。

$$TS_i = (T_{i2}, T_{i3}, S_{i1}, S_{i2}, S_{i3}, S_{i4}, \dots, S_{i12})$$

d_{ij} ：表示商品 i 與商品 j 在特徵向量空間內的距離。

根據各商品所選擇的最適長期趨勢關係式與季節性分析的季節數不同，使用不同參數描述發展趨勢的商品即屬於不同類別。因此，將原始銷售量歷史記錄轉換為時間序列分析指標之後，這些商品至少可依長期趨勢分析分作遞增與遞減；依季節性分析分作無季節性、四期季節性、十二期季節性三個類別。

不同參數設定下所計算得出的指標之間並不存在合理的轉換關係，例如在長期趨勢分析中，採用一次線性關係與二次函式所得出的β_1所代表的意義完全不同(在一次線性中代表遞增或遞減趨勢，在二次函式中則代表二次曲線最高最低點水平位移的幅度)，或是在季節性分析中透過資料縮減或添補的方法使得四期季節性與十二期季節性得以互相比較。

所以，唯有歸為同一類別的商品可以計算之間的距離，而商品 i 與商品 j 兩者之間的距離

$$d_{ij} = \left|TS_i - TS_j\right|^2$$

藉由採取平方和的方法去除商品之間指標值差異正負號加總時的抵銷效應，同時凸顯較大的 d_{ij} 對整體模型的影響應該愈大，應該優先考慮將彼此之間距離較遠的商品分為不同類別。

3-3-4 限制式

(1)群集數限制

$$M_{min} \leq m \leq M_{max}$$

若群集數目太少，會不足以將發展趨勢相異的商品區分為

不同類別，來避免類別商品發展趨勢扭曲的問題，因此需要選擇 M_{min} 做為下限；若群集數目太多，會使得每個商品自成一類，無法達成合併商品記錄來提升預測準確度的效果。M_{min} 與 M_{max} 的選擇方法留待演算法中討論，經過不同的情境實驗分析之後提出建議值。

3-3-5 兩階段目標函式

(1)最小化同群集內的樣本間距離總平均

$$Min\ aTDG = \frac{2}{\sum_{c=1}^{m} N(G_c^m)(N(G_c^m)-1)} \sum_{i=1}^{N(P)-1} \sum_{j=i+1}^{N(P)} d_{ij}, \qquad \forall g \in G^m$$

$aTDG$ 代表在分類結果 g 下，所有商品與屬於同一群集的其他商品彼此之間的距離平均，屬於不同群集的商品之間的距離則不加以計算。在解讀意義上，希望被歸為同一群的商品彼此愈相似愈好。

(2)最大化不同群集之間的距離總平均

$$Max\ AGD = \frac{2}{m(m-1)} \sum_{k=1}^{m-1} \sum_{j=k+1}^{m} d_{\bar{k}\bar{j}}, \qquad \text{for } M_{min} \le m \le M_{max}$$

AGD 代表所有群集之間的平均距離，而每個群集用於計算距離的虛擬代表點 $(\bar{k}、\bar{j})$ 的座標為同一群集的所有商品屬性指標的平均值。在解讀意義上，希望不同群集之間的差異愈大愈好。

本研究若只考慮第一階段的最小化目標，其最佳化方向會建議每個商品自成一個類別，$aTDG$ 等於零。但是這並非最適合的分類結果，一來將失去合併銷售記錄提升預測準確度的效果，二來將無法提供資訊給後續分析使用，若有銷售記錄較短的商品加入分類，無法將它放入適當的類別。

因此加入第二階段最大化目標，當一群待分類的商品被愈分愈細，群集之間的距離總平均將愈來愈近，與第一階段目標產生拮抗的作用，當這兩階段目標達成平衡時，才是最適當的分類結果。

構成本目標函式的基本成分——距離 d_{ij}——的計算公式不為線性(平方和)，且每個類別所包含的商品個數在搜尋最佳解的過程中也不固定，無法以最佳化方法直接算出極限值。最小化的過程即不斷嘗試不同的商品分類指派，如果一個商品從一個類別換到另一個類別，代表他與新類別中所有商品的平均距離小於與原本類別中的其他商品平均距離。

第四節　成果評估流程

本研究根據第二章所整理的評估方法，將採用 MAPE 做為預測準確度的量化評估標準。

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{A_t - Y_t}{A_t}\right|$$

A_t 代表第 t 期的已知歷史記錄，Y_t 代表第 t 期的預測值，兩者的差再除以已知歷史記錄變成比值，最後計算整個觀察期間，共 T 期的誤差絕對百分比平均值。

在本研究中，MAPE 會首先用於尋找每項商品最適合用來表示銷售量歷史記錄的長期趨勢關係與季節性指標，評估不同參數設定所產生的結果，如圖 3-3 所示。

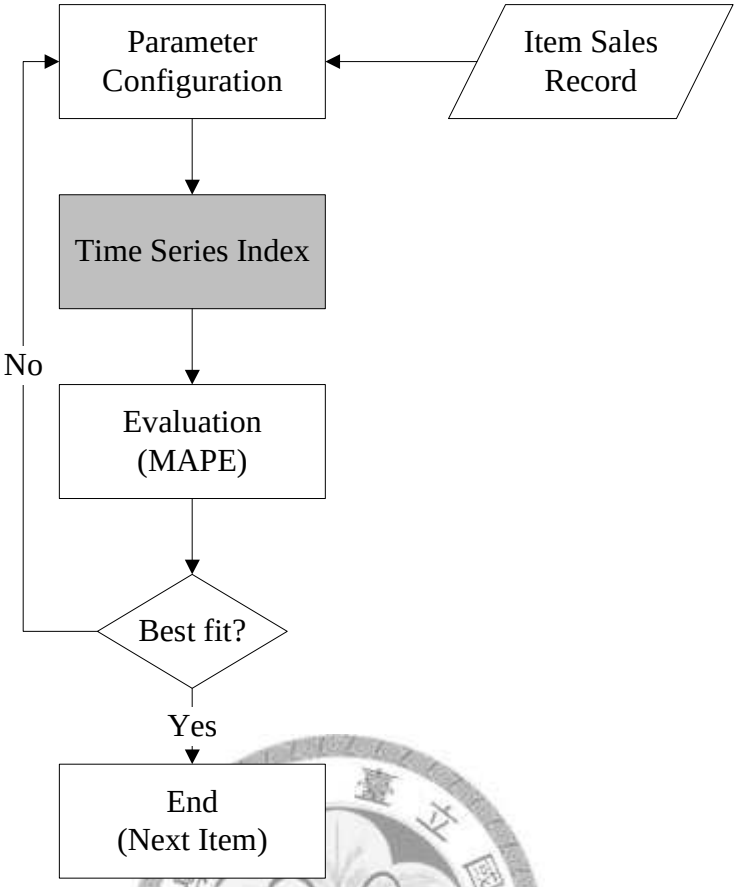

圖 3-6：MAPE 用於時間序列分析階段

(資料來源：本研究整理)

然後在完成商品分類之後，整合同類別商品銷售量記錄，並進行預測，最後將類別商品預測量分配給單項商品，在此步驟之後以 MAPE 評估預測準確度，做為不同商品分類架構的比較標準，如圖 3-4 所示。然而，預測準確度並非所有商品皆一致，大部份的企業會只考量「重要商品」預測的準確度。重要商品之定義通常以營業額為標準，佔整體營業額 50%以上或是營業額排名於前 20%之商品可稱為重要商品。這些影響營業額甚大之商品，其預測之準確度會影響整體系統之營運結果。因此，本研究在最後評比商品架構優劣時，同樣只考量重要商品。

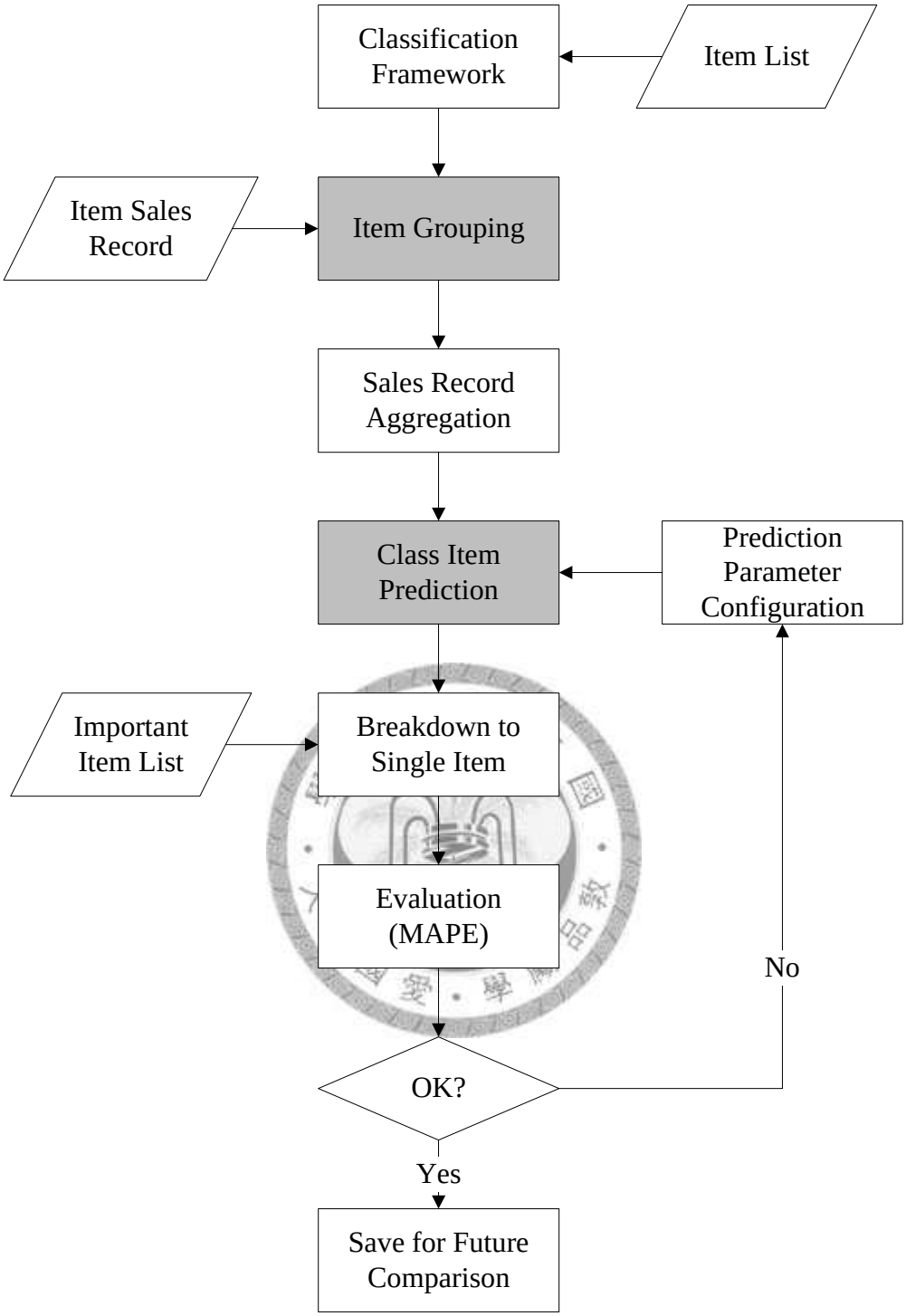

圖 3-7：MAPE 用於銷售量預測

(資料來源：本研究整理)

本研究為了處理隨時間累積的銷售歷史記錄，並且使得不同商品可以在相同的架構下互相比較其銷售發展趨勢的異同，將同一商品不同時間的銷售記錄整合為單一資料列，並且採用時間序列趨勢分析，以長期趨勢指標與季節性指標來描述每項商品的發

展趨勢特徵，避免記錄累積時間長度不同的問題。本研究採用MAPE做為評比標準，決定描述商品發展趨勢的最適參數設定(一次線性趨勢或者二次線性；無季節性、四期季節性或十二期季節性)，使用不同參數的商品視為不同類別。最後以兩階段目標函式定義欲搜尋的最佳商品分類結果。此商品分類是否適合商品銷售量預測，同樣採用 MAPE 評估重要商品的預測準確度，與其他商品分類比較。

第四章　商品依銷售資料分類啟發式演算法

第一節　商品依銷售資料分類演算法概述

本研究將分析商品基本資訊與銷售歷史記錄，進而建構一商品分類架構。依此分類架構，每個商品會被歸屬於一個類別，然後同類別的商品銷售記錄會在後續的預測分析中整合形成單一的類別銷售量進行預測，最後再將類別預測銷售量分配給單項商品做為最終的銷售量預測。

在過往的研究中，研究需求管理、改進銷售預測準確度者並未考慮商品分類架構是否適合預測所需，僅依靠廠商或專家所提供的解釋性分類架構進行銷售記錄整合，可能造成個別商品之間的銷售發展趨勢差異模糊扭曲類別商品整體的銷售發展趨勢；研究分類演算法者僅專注於處理研究對象的靜態屬性，尚未將研究對象延伸至隨時間累積變動的時間序列資料。

本研究試圖改進前述兩者不足之處，找出符合第三章所提出的最小距離模型的「最佳分類架構」。但是其目標函式並非線性(平方和)，且解集合的形式(商品 i 屬於類別 C_i)近似於整數規劃問題，因此無法透過線性規劃的方法找尋最佳解，必須在每個整數點上進行全域搜尋(Global Search)。這樣的做法需要相當長的時間，例如僅僅是將兩百個品項的商品區分為兩類，就必須搜尋多達 $2^{200}(\approx 1.6\text{x}10^{60})$種排列組合，幾乎無法利用有限的運算資源求出解。為了更有效率地解決最適產品分類架構問題，本研究將採用啟發

式演算法，以期在可接受的時間範圍內求得最佳解或近似最佳解。

本研究最主要的特色在於將時間序列資料——商品銷售記錄——轉換為數個描述長期趨勢與季節性變動的「指標」，使得上市日期不同、記錄單位不同的商品有共同的屬性，適合引入資料探勘方法進行相似度的計算，加以分類；而在最關鍵的分類架構建立時，採用基因演算法為基礎的搜尋方法，大幅降低找到最佳解或近似最佳解所需的時間。

第二節　演算法主要流程

本研究的演算法主要流程分為三大部分，如圖 4-1 所示：第一部分為前置作業，將單一商品每日銷售記錄轉換為時間序列指標；第二部分為分類演算法，根據前一部分所轉換的指標將銷售發展趨勢相似的商品加以群集分類，建構一分類架構；第三部份為預測效果評估，依照丁[1]所提出的銷售預測流程，使用第二部分所建構的分類架構整合銷售記錄，再次計算分類商品最佳參數並產生預測值，最後以 MAPE 做為預測準確度的量化標準。

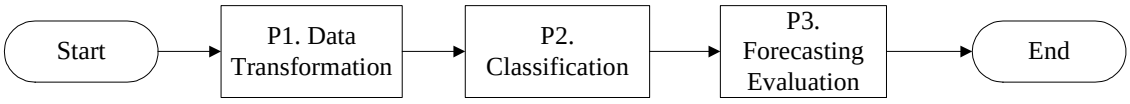

圖 4-8：演算法主要流程

(資料來源：本研究整理)

本章接下來將分數小節分別探討上述各部分演算法的詳細步驟，其中：第三節將說明第一部分前置作業；第四節說明第二部分的分類演算法。

第三節　前置作業

各商品因為上市日期前後不一，所擁有的銷售歷史記錄量也因此不同，僅以銷售歷史記錄無法形成品質良好的資料引入資料探勘方法模型進行分類。為了去除這個問題，本研究採用如圖 4-2 所示的資料轉換流程，將各商品的銷售歷史記錄轉換為時間序列指標。

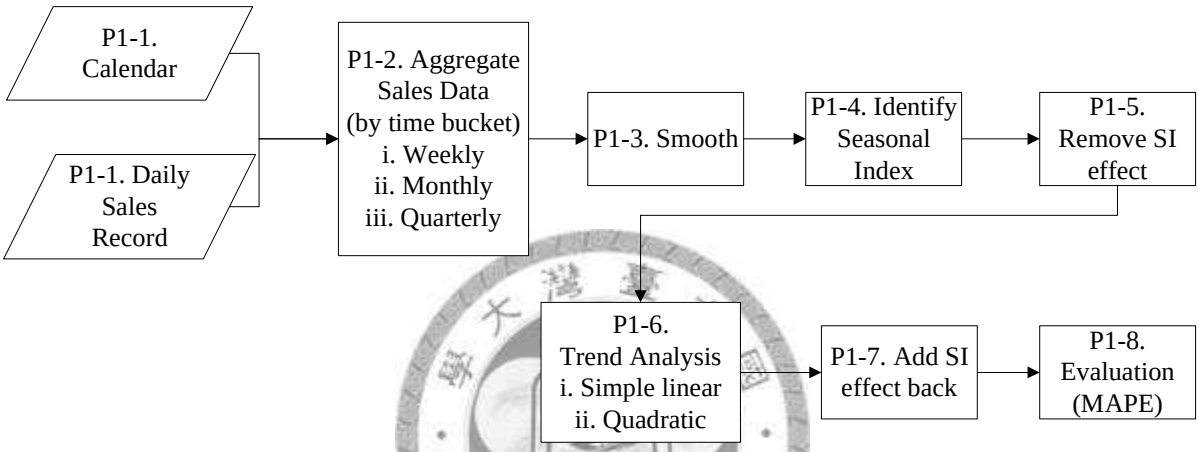

圖 4-9：資料轉換流程

(資料來源：本研究整理)

P1-1. 建立統一的銷售週期表與整理銷售歷史記錄

在任何跟時間有關的商業應用分析中，我們經常需要將資料依照某些日曆週期整合，以得出更加有意義且容易解釋的規則或結論。但是「週(Week)」、「月(Month)」、「季(Quarter)」、「年(Year)」的定義隨著產業或使用情境的不同也有所不同，為了讓大家「說同樣的語言(speak the same language)」以及降低應用程式開發的複雜度，必須先建立一套統一的銷售週期表(Sales Calendar)，明確定義某一段從開始日期到結束日期的期間所屬的週別、月別、季別。

本研究定義每一週從星期一(Monday)開始，星期日結束(Sunday)；由開始日期的月別為準；從一月開始，每三個月為一季，

如表 4-1 所示，週期編號 10000 代表 1999 年的第一週，從 1999/1/4 開始，1999/1/10 結束，屬於該年的第一個月，第一季。

表 4-4：Sales Calendar (Partial)

CALNDR_NO	DISPLAY_NO	FROM_DATE	TO_DATE	MONTH_NUM	SEASON_NUM	YEAR_NUM
10000	1999-1	1999/1/4	1999/1/10	1	1	1999
10001	1999-2	1999/1/11	1999/1/17	1	1	1999
⋮	⋮	⋮	⋮	⋮	⋮	⋮
10012	1999-13	1999/3/29	1999/4/4	3	1	1999
10013	1999-14	1999/4/5	1999/4/11	4	2	1999
⋮	⋮	⋮	⋮	⋮	⋮	⋮

(資料來源：本研究整理)

在銷售歷史資料的準備方面，如第三章所述，將同一商品於不同日期的銷售記錄整合成以時間為軸的時間序列資料。本研究採用的銷售記錄為商品銷售數量，所以連續日期的當日總銷售量成為該商品的一個屬性，如表 3-3 所示。

P1-2. 依照不同長度的時間刻度整合銷售歷史記錄

根據第三章所提出的假設，本研究所探討的銷售預測時距以週為單位，因此前一步驟所整合的每日銷售量記錄要更進一步依據銷售週期的定義整合為週銷售量記錄。

表 4-5：商品 X 從 1999/1/4 至 1999/1/24 的每日銷售量記錄

Date	SalesQty	Date	SalesQty	Date	SalesQty
1999/1/4	0	1999/1/11	12	1999/1/18	21
1999/1/5	2	1999/1/12	19	1999/1/19	23
1999/1/6	3	1999/1/13	25	1999/1/20	24
1999/1/7	0	1999/1/14	20	1999/1/21	27
1999/1/8	4	1999/1/15	27	1999/1/22	30
1999/1/9	7	1999/1/16	15	1999/1/23	15
1999/1/10	3	1999/1/17	5	1999/1/24	26

(資料來源：本研究整理)

表 4-6：商品 X 的日銷售量經整合為週銷售量

CALNDR_NO	FROM_DATE	TO_DATE	SalesQty

10000	1999/1/4	1999/1/10	19
10001	1999/1/11	1999/1/17	123
10002	1999/1/18	1999/1/24	166

(資料來源：本研究整理)

另外，第三章所提及的季節性指標不一定將整年分作四季(三個月為一季)，亦可分作十二季(一個月為一季)進行分析。為了尋找最適合描述該商品銷售發展趨勢的參數設定，週銷售量記錄更進一步整合為月銷售量記錄以及季銷售量記錄，如表 4-4 所示。

表 4-7：商品 X 銷售量記錄依週月季整合

CALNDR_NO	SalesQty_W	SalesQty_M	SalesQty_Q
10000	19	516	1,957
10001	123		
10002	166		
10003	208		
10004	208	563	
10005	98		
10006	187		
10007	70		
10008	178	878	
10009	212		
10010	211		
10011	138		
10012	139		

(資料來源：本研究整理)

經過此步驟，所有商品的日銷售量歷史記錄都會變成以週月季三種週期長度為刻度整合的記錄。

P1-3. 平滑銷售記錄

為了去除銷售量隨機波動的誤差所造成的影響，也使得可能存在的季節性因素更容易被辨認，本研究假設銷售量為銷售期數的函數，採用簡單線性迴歸分析(simple linear regression analysis)將前一步驟的銷售歷史記錄加以平滑化，而且不會遭遇資料點縮減的問題。例如表 4-5 為商品 X 連續三年的季銷售量記錄，利用

簡單線性迴歸分析得出季銷售量(Y_t)與期數(t)的關係為 $Y_t = 0.0 + 3664.016t$ (假設商品 X 在第 0 期上市，常數項為 0)

表 4-8：商品 X 連續三年的季銷售量記錄與迴歸分析預測值

t	YEAR_NUM	SEASON_NUM	SalesQty_Q (Y_t)	$\hat{Y}_t = 0 + 3664.016t$
0	1999	1	1,957	0.0
1	1999	2	4,038	3,664.02
2	1999	3	2,968	7,328.03
3	1999	4	27,003	10,992.05
4	2000	1	2,760	14,656.05
5	2000	2	14,143	18,320.08
6	2000	3	3,884	21,984.09
7	2000	4	56,529	25,648.11
8	2001	1	2,901	29,312.13
9	2001	2	24,204	32,976.14
10	2001	3	6,835	36,640.16
11	2001	4	86,623	40,304.17

(資料來源：本研究整理)

每個商品的月銷售量記錄與季銷售量記錄都必須進行此步驟，惟週銷售量記錄被視為季節性因素不存在，故不須平滑化。

P1-4. 辨認季節性指標

本研究對時間序列分析模型採用乘法模式，因此季節性指標(SI_t)為一代表實際銷售量與隱含的長期趨勢之間的縮放比率，由 $Y_t/\hat{Y}_t$ 計算得出；當期的季節性指標等於每年該期的指標平均值；最後還必須將季節性指標常態化(normalize)，使得四期季節性指標的總和為 4，或十二期季節性指標總和為 12。表 4-6 與表 4-7 為前述商品 X 的季節性(假設擁有四期季節性)分析步驟。

表 4-9：商品 X 季銷售量季節性分析(Step 1)

t	Y	$\hat{Y}_t$	$SI_t = Y_t / \hat{Y}_t$
0	1,957	0.0	-
1	4,038	3,664.02	1.1021
2	2,968	7,328.03	0.4050
3	27,003	10,992.05	2.4566
4	2,760	14,656.05	0.1883

5	14,143	18,320.08	0.7720
6	3,884	21,984.09	0.1767
7	56,529	25,648.11	2.2040
8	2,901	29,312.13	0.0990
9	24,204	32,976.14	0.7340
10	6,835	36,640.16	0.1865
11	86,623	40,304.17	2.1492

(資料來源：本研究整理)

表 4-10：商品 X 季銷售量季節性分析(Step 2)

t	**1**	**2**	**3**	**4**	**SUM**
SI_t	0.1436	0.8693*	0.2561	2.2699	3.5390
SI_t **(Normalized)**	0.1624	0.9826**	0.2894	2.5656	4

*0.8693 = (1.1021 + 0.7720 + 0.7340) / 3

**0.9826 = 0.8693 * 4 / 3.5390

(資料來源：本研究整理)

P1-5. 移除季節性效應

得知季節性指標之後，在此步驟，重新回到週銷售量記錄的調整。將原本的週銷售量記錄除以前一步驟所算出的季節性指標，根據最開始所定義的銷售週期表來指派應該調整的幅度。例如表 4-1 所示，週期編號 10000 屬於第一季，因此商品 X 該週的銷售量應該調整為 SalesQty_W / 0.1624；又週期編號 10013 屬於第二季，商品 X 該週的銷售量則應該調整為 SalesQty_W / 0.9826。

表 4-11：商品 X 週銷售量記錄移除四期季節性效應 (Partial)

CALNDR_NO	**SEASON_NUM**	**SI**	**SalesQty**	**SalesQty (SI effect free)**
10000	1	0.1624	19	117.03
10001	1	0.1624	123	757.60
⋮	⋮	⋮	⋮	⋮
10015	2	0.9826	365	371.47
⋮	⋮	⋮	⋮	⋮
10027	3	0.2894	302	1,043.41
⋮	⋮	⋮	⋮	⋮
10040	4	2.5656	1,894	738.22

⋮	⋮	⋮	⋮	⋮

(資料來源：本研究整理)

P1-6. 分析長期趨勢

根據本研究的假設，每一商品經過前述五個步驟，去除不同刻度的季節性效應之後所產生的三組週銷售量記錄應該只剩下長期趨勢因素。針對每一組週銷售量記錄，本研究採用兩種迴歸模式分析：簡單線性(simple linear)與二次函式(quadratic)。以簡單線性迴歸模型為例，令平滑之後的銷售量 $\hat{y}_t = b_0 + b_1 t$，代入已知共 T 期銷售量記錄可得出下列的矩陣方程式

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_T \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_T \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}, \qquad \boldsymbol{\hat{Y}} = \boldsymbol{Ab} \ (1)$$

則長期趨勢指標(矩陣 $\boldsymbol{b}$)可藉由 $\boldsymbol{A}^{-1}\boldsymbol{\hat{Y}}$ 計算得出，但是矩陣 $\boldsymbol{A}$ 在絕大多數的情況下並非一個方陣(square matrix)，因此其反矩陣($\boldsymbol{A}^{-1}$)無法計算。若將矩陣方程式(1)改為 $\boldsymbol{A}^T\boldsymbol{\hat{Y}} = \boldsymbol{A}^T\boldsymbol{Ab}$ (2)，則 $\boldsymbol{b} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{\hat{Y}}$。經由矩陣維度檢驗，$\boldsymbol{A}$ 為一 T*2 的矩陣，$\boldsymbol{\hat{Y}}$ 為 T*1，$\boldsymbol{b}$ 為 2*1，則矩陣方程式(2)為(2*T)(T*1) = (2*T)(T*2)(2*1) = (2*1)，又矩陣 $\boldsymbol{A}^T\boldsymbol{A}$ 為一方陣(2*2)，表示($\boldsymbol{A}^T\boldsymbol{A}$)$^{-1}$ 存在，當矩陣 $\boldsymbol{A}^T\boldsymbol{A}$ 的行列式值不為 0，則此方程式有解。若令 $\hat{y} = b_0 + b_1 t + b_2 t^2$，則

$$\boldsymbol{A} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_T & t_T^2 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$$

以商品 X 前五期移除季節性效應的銷售量記錄為例，採用簡單線性迴歸模型分析：

$$\boldsymbol{A}=\begin{bmatrix}1 & 0\\ 1 & 1\\ 1 & 2\\ 1 & 3\\ 1 & 4\end{bmatrix},\qquad \boldsymbol{A}^T\boldsymbol{A}=\begin{bmatrix}1 & 1 & 1 & 1 & 1\\ 0 & 1 & 2 & 3 & 4\end{bmatrix}\begin{bmatrix}1 & 0\\ 1 & 1\\ 1 & 2\\ 1 & 3\\ 1 & 4\end{bmatrix}=\begin{bmatrix}5 & 10\\ 10 & 30\end{bmatrix},$$

$$(\boldsymbol{A}^T\boldsymbol{A})^{-1}=\begin{bmatrix}0.6 & -0.2\\ -0.2 & 0.1\end{bmatrix},$$

$$(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T=\begin{bmatrix}0.6 & 0.4 & 0.2 & 0 & -0.2\\ -0.2 & -0.1 & 0 & 0.1 & 0.2\end{bmatrix}$$

$$\boldsymbol{b}=\begin{bmatrix}b_0\\ b_1\end{bmatrix}=(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\widehat{\boldsymbol{Y}}=\begin{bmatrix}0.6 & 0.4 & 0.2 & 0 & -0.2\\ -0.2 & -0.1 & 0 & 0.1 & 0.2\end{bmatrix}\begin{bmatrix}117.03\\ 757.60\\ 1{,}022.46\\ 1{,}281.15\\ 1{,}281.15\end{bmatrix}=\begin{bmatrix}321.52\\ 285.18\end{bmatrix}$$

P1-7. 加回季節性效應

截至之前的步驟，已計算得出所有參數值，從這個步驟開始將所有參數值套入本研究所假設的時間序列資料描述模型中。以商品 X 共 156 週銷售量記錄為例，假設簡單線性迴歸與四期季節性，b_0 = 397.9662、b_1 = 11.7110、S_1 = 0.1624、S_2 = 0.9826、S_3 = 0.2894、S_4 = 2.5656。

表 4-12：商品 X 引入所有參數值所產生的預測值

t	$y_t=(b_0+b_1t)\times S_t$
0	64.6114
1	66.5127
2	68.4141
3	70.3154
4	72.2167
⋮	⋮

(資料來源：本研究整理)

P1-8. 評估模型適度

本研究以預測誤差作為評估所設定的模型是否適合的標準，而誤差的量化方法採用平均絕對百分比誤差(MAPE)，其計算公式如下所示：

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{A_t - Y_t}{A_t}\right|$$

其中 A_t 代表第 t 期已知的銷售量記錄，Y_t 代表第 t 期依照前述步驟所產生的預測值，T 為所擁有的歷史記錄長度。如此將可避免預測值高估低估之間互相抵銷的問題，而且採用比值也可去除數據級數大小的影響。

以商品 X 前五期預測值為例。

表 4-13：評估模型適度(MAPE)

CALNDR_NO	SalesQty (A_t)	SalesQty_Pred. (Y_t)	$\lvert A_t - Y_t \rvert / A_t$
10000	19	64.61	2.4006
10001	123	66.51	0.4592
10002	166	68.41	0.5879
10003	208	70.32	0.6619
10004	208	72.22	0.6528
		MAPE = 0.9525	

(資料來源：本研究整理)

MAPE 值愈小表示預測誤差愈小，也代表所選擇的參數設定愈適合用於描述該商品的銷售量發展趨勢。

第四節　分類演算法(DMAPC)

經過前置作業之後，每一個商品的銷售量歷史記錄會被轉換成六組參數設計的時間序列指標：在長期趨勢部分，包括簡單線性與二次函式迴歸模式；在季節性分析部分，包括無季節性、四期季節性、十二期季節性，共六種組合。

本研究的分類演算法將創建一階層式分類架構，前兩層為規則基礎的分類，最後一層則以第三章所定義的距離為標準，將相近(相似)的商品群集成為一類，如圖 4-3 所示。

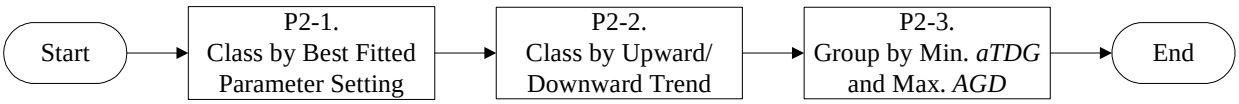

圖 4-10：階層式分類架構建立流程

(資料來源：本研究整理)

在第三章提及，不同的時間序列指標設計模式下所產生的參數值彼此之間並不存在合理的轉換方式，因此分類架構的第一層(P2-1.)即為依照各商品最適合的分析模式組合予以區分。例如商品 X、Y、Z 採用簡單線性迴歸模式與四期季節性分析時，與其他五種分析模式組合相比，皆產生最小的 MAPE 值，則在此分類階層將商品 X、Y、Z 歸為同一類。

本研究所希望避免的不適當的分類結果，其中之一即是銷售發展長期趨勢向上成長與向下衰減的商品被歸為同一類，將造成分類商品整合的銷售發展趨勢互相抵銷，長期趨勢因此顯得趨緩，無法產生準確的銷售預測。基於前述理由，在此分類階層(P2-2.)將前一階層所區分的各個子分類中的所有商品依照長期趨勢指標表示成長或衰減再予以區分。例如代表商品 X、Y、Z 銷售發展長期趨勢的參數為 b_1——因為最適模式為簡單線性迴歸，依此參數值正負判斷商品 X、Y、Z 是否應該在此分類階層分作不同類別。若最適模式為二次函式迴歸模式，則觀察參數 b_2。

分類階層第三層(P2-3.)是最主要的分類步驟，使用資料探勘的方法，藉由定義的「距離」遠近來判斷兩商品的銷售發展趨勢是否相似，而適合聚集為一個分類，為後續的銷售預測流程創造良好的基礎。

為了去除長期趨勢指標值數量級距對距離計算的比重影響，本研究捨棄用於描述趨勢線截距的指標(b_0)，並且標準化其他長期趨勢指標，將指標值除以所有待分類商品的指標平均值($\bar{b}_1$、$\bar{b}_2$)，使所有商品該指標的平均值等於 1，與每個季節性指標值產生相同

的權重。所以，代表每個商品銷售趨勢的特徵向量，以一採用簡單線性迴歸與四期季節性描述的商品為例，如下所示：

$$V = \left(\frac{b_1}{\bar{b}_1}, s_1, s_2, s_3, s_4\right)$$

為了找出一個分類結果最佳化第三章所提出的目標函式，並且將搜尋解所需的時間縮短至可接受的時間內，本研究採用基因演算法為基礎的最適分類搜尋方法。大致來說，針對同一群商品，在數個分類結果中，經過評估選出較好的分類結果進行調整來產生新的分類結果。如此一來改善了純粹使用規則基礎(rule-based)的分類方法缺乏彈性的缺點——一旦決定了，就無法再調整分類結果，即使根據目前的規則只能找到區域最佳解(local optimal)而不是全域最佳解(global optimum)。

此方法的流程如圖 4-4 所示，共包含兩個迴圈：一開始指定分群數目 m，將分類結果進行編碼，使之成為「染色體」的形式，然後創造出兩條起始的染色體。每次產生新的染色體都要進行評估，捨棄最差的染色體並找出最優質的染色體進行基因演算法中的交配(crossover)或以突變(mutation)來產生子代，也就是新的分類結果。評估然後繁衍的循環不斷重複直到終止條件被滿足為止。內層迴圈的最佳分類結果進行第二階段評估，然後重新設定 m，再次進行內層基因演算法搜尋，直到外層迴圈的限制被滿足。以下詳細說明各細部流程：

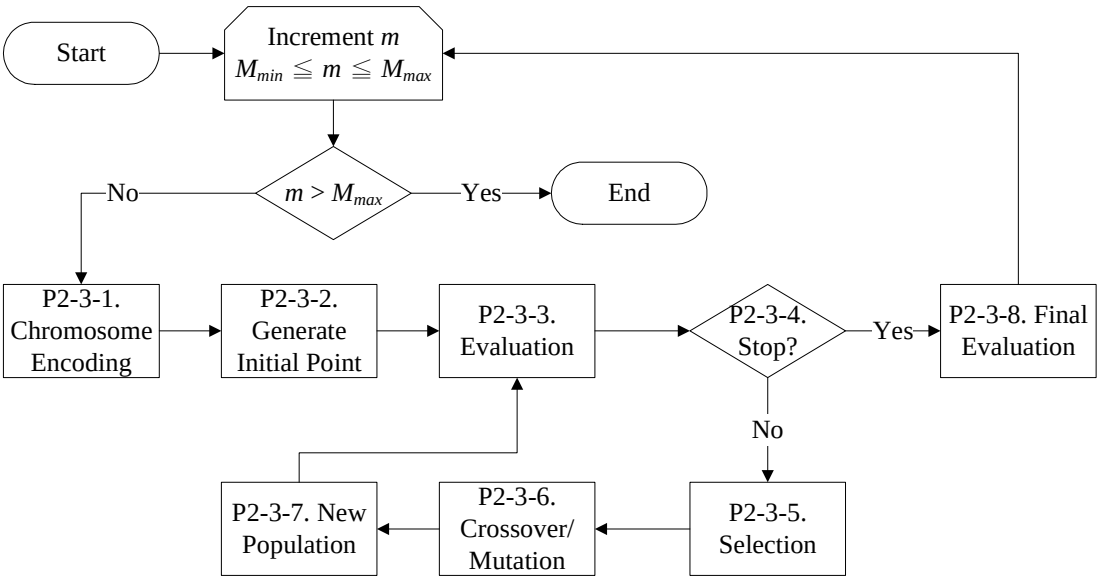

圖 4-4：最小距離群集分類演算法流程

(資料來源：本研究整理)

P2-3-1. 染色體編碼

由於基因演算法皆以染色體(chromosome)為執行之個體，在演算法開始之前，必須先定義染色體的編碼方式，用以表示不同的分類結果，然後依此建立資料結構，以便進行後續繁衍的動作。本研究的染色體編碼方式如表 4-11 所示，假設第二個分類階層的某個子分類包含有 N 個商品，則染色體會有 N 個基因；位置 i 的基因代表任意排列的商品清單中列於第 i 個商品所屬的群集 C_i，染色體即 C_i 所組成的序列。

表 4-14：染色體編碼範例

i	1	2	3	4	5	6	7	8	9	10
Item No.	A	B	C	D	E	F	G	H	I	J
C_i	1	1	2	1	1	3	2	3	1	2

(資料來源：本研究整理)

以表 4-11 為例，第 2 個基因為 1，代表商品 B 屬於群集 1；第 7 個基因為 2，代表商品 G 屬於群集 2；這條染色體為 1121132312。群集的編號僅具標示區別的意義，並不表示順序，C_i 相同表示所對應的商品屬於同一群。

P2-3-2. 產生初始染色體

當決定了染色體的表現方式之後，要產生初始的染色體以啟動演算法，做為「祖先」繁衍第一代。因為效率是本研究很重要的評量指標，因此初始染色體的品質對最後產生的結果有很大的影響。

本研究採用兩種方式各產生一組初始染色體，隨機產生與規則基礎：(假設要將 N 個商品分為 m 群)

1. 隨機產生：染色體的每個基因透過擲一個公平的 m 面骰子來決定 C_i，表示每個商品屬於任一群集的機率相等。
2. 規則基礎：藉由觀察商品的屬性值，令距離較遠的商品屬於不同的群集，然後各自招攬較近的商品與之形成同一群集。以最適時間序列分析模型為簡單線性迴歸與四期季節性，且長期趨勢為向上成長的十件商品為例，表 4-12 所列為各商品的參數值。其詳細流程如下：

表 4-15：商品時間序列指標資訊範例

Item No.	b_1	s_1	s_2	s_3	s_4
A	1.2515	0.7802	0.5036	1.7631	0.9531
B	1.7058	0.0840	0.8374	0.9872	2.0914
C	0.9258	0.4154	1.1289	1.1267	1.3290
D	1.9570	1.4980	0.7807	0.8379	0.8835
E	0.4035	2.0600	0.5874	0.9560	0.3966
F	0.1431	0.5739	1.7908	1.1252	0.5101
G	0.6074	0.5278	0.5976	1.0644	1.8102
H	1.7204	1.3413	0.6842	0.0727	1.9018
I	0.0481	0.5601	0.8491	2.1942	0.3966
J	1.2791	1.0859	0.9356	0.9982	0.9803

(資料來源：本研究整理)

i.計算商品之間的距離

距離 (d_{ij}) 係指商品 i 與商品 j 的所有指標值差方和，如第三章所述。以商品 A 與商品 B 為例，

$$d_{ij} = (1.2515 - 1.7058)^2 + (0.7802 - 0.0840)^2 + (0.5036 - 0.8374)^2$$
$$+ (1.7631 - 0.9872)^2 + (0.9531 - 2.0914)^2 = 2.700$$

表 4-13 為上述十項商品兩兩之間的距離。

表 4-16：商品距離範例

d_{ij}	A	B	C	D	E	F	G	H	I	J
A	0	2.700	1.176	1.951	3.325	3.531	1.710	4.325	2.111	0.867
B	2.700	0	1.404	3.547	8.536	6.110	1.546	2.477	7.304	2.430
C	1.176	1.404	0	2.638	4.169	1.746	0.632	3.125	2.878	0.750
D	1.951	3.547	2.638	0	3.017	5.386	3.706	1.712	6.604	0.688
E	3.325	8.536	4.169	3.017	0	3.766	4.399	5.306	3.977	2.179
F	3.531	6.110	1.746	5.386	3.766	0	3.335	7.346	2.051	2.521
G	1.710	1.546	0.632	3.706	4.399	3.335	0	2.900	3.652	1.570
H	4.325	2.477	3.125	1.712	5.306	7.346	2.900	0	10.200	2.029
I	2.111	7.304	2.878	6.604	3.977	2.051	3.652	10.200	0	3.570
J	0.867	2.430	0.750	0.688	2.179	2.521	1.570	2.029	3.570	0

(資料來源：本研究整理)

ii. 找出尚未分開且距離最大的兩個商品，加入極點集合

根據表 4-13，可以看出仍屬於同一群的商品 H 與商品 I 的距離(=10.20)最大，所以這兩個商品應該優先區分屬於不同群集，並將他們稱為極點(polar point)。

iii. 依極點集合，區分所有商品

在此步驟計算所有其他商品與極點商品的距離(d_{ij})，與哪個極點距離較近就將之歸為同一群集。表 4-14 為其他商品與商品 H、I 的距離，可以觀察得出，商品 H、B、D、G、J 形成一個群集，商品 I、A、C、E、F 形成另一個群集，如表 4-15 與圖 4-5 所示。

表 4-17：商品 i 與極點距離表範例

i	d_{iH}	d_{iI}
A	4.325	2.111
B	2.477	7.304
C	3.125	2.878
D	1.712	6.604

E	5.306	3.977
F	7.346	2.051
G	2.900	3.652
J	2.029	3.570

(資料來源：本研究整理)

表 4-18：規則基礎初始染色體分類結果範例

C_1	H	B	D	G	J
C_2	I	A	C	E	F

(資料來源：本研究整理)

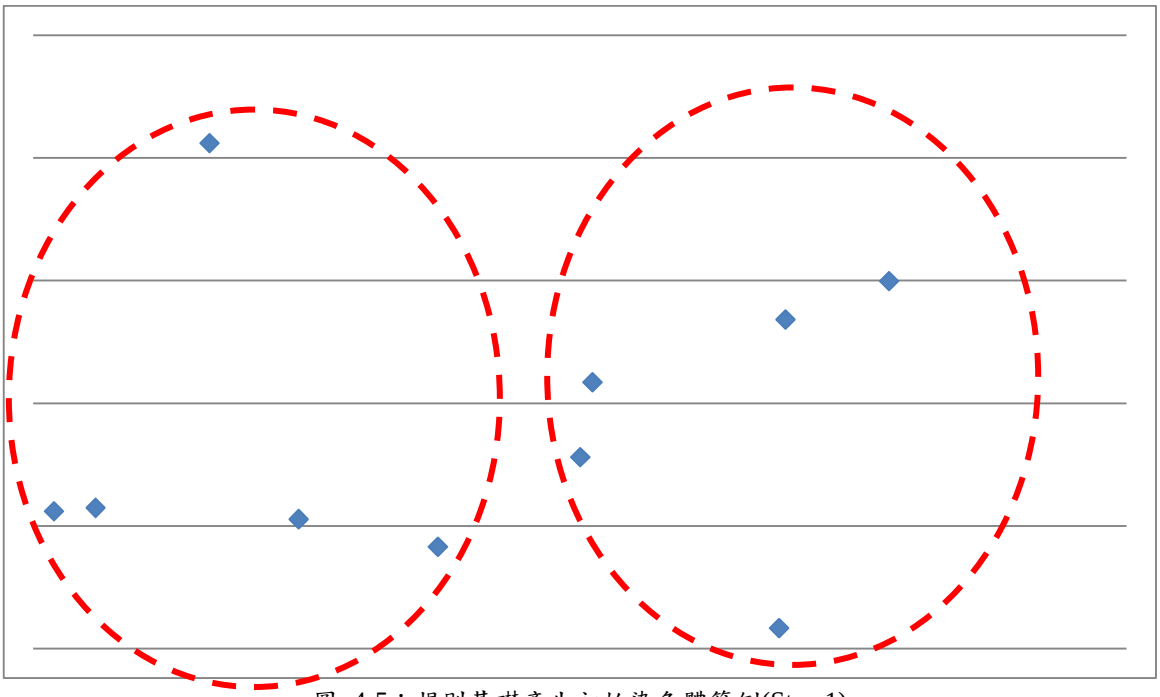

圖 4-5：規則基礎產生初始染色體範例(Step 1)

(資料來源：本研究整理)

iv. 檢查已區分群集數是否達到指定 *m*

若經過前述步驟已取得 *m* 個不為空集合的群集分類結果，則分類演算法終止，儲存目前的分類結果做為規則基礎的初始染色體。若群集數仍不足，例如經過前述步驟僅將十項商品分為兩群，但是 *m* 指定為 3，則回到步驟 ii.，繼續搜尋尚未分開但是季節性差異最大的兩個商品，將之加入極點集合，重新進行步驟 iii.。若群集數超過指定 *m*，繼續下一個步驟。

v. 重新合併被分開的群集

新極點通常會成對加入，以前述的十項商品為例，下一個加入極點集合的為商品 C 與商品 E，然後形成四個群集；但是若 *m* 指定為 3，則不符合要求。此時，計算每個群集的虛擬代表點，也就是屬於同群集的所有商品參數值的平均，然後計算群集中心兩兩的距離，選擇中心距離最近的兩個群集予以合併，以符合要求。例如商品 C、E 加入極點集合之後所形成的群集成員如表 4-16 與圖 4-6 所示，各群集的中心點與距離則如表 4-17 所示。

表 4-19：規則基礎初始染色體分類結果範例 2

C_1	H	D				
C_2	I					
C_3	C	A	B	F	G	J
C_4	E					

(資料來源：本研究整理)

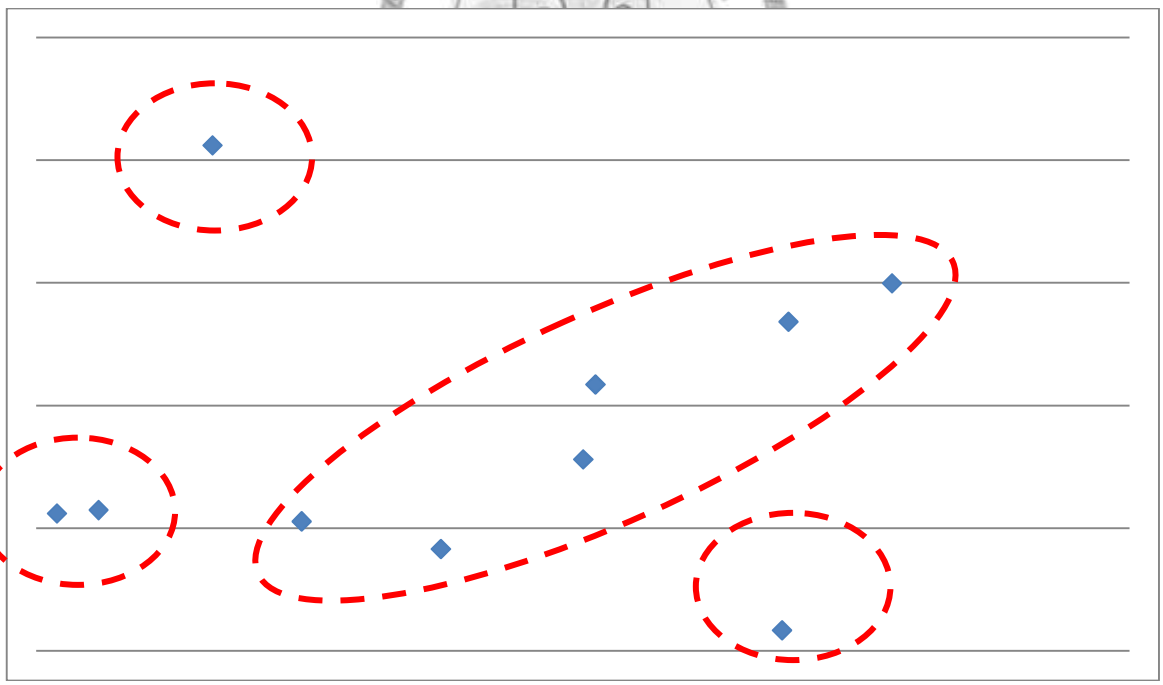

圖 4-6：規則基礎產生初始染色體範例(Step 2)

(上圖僅試圖以二維平面展示範例之多維空間結果)

(資料來源：本研究整理)

表 4-20：群集中心點距離範例

	b_1	s_1	s_2	s_3	s_4	d_{iC1}	d_{iC2}	d_{iC3}

C_1	1.8387	1.4196	0.7324	0.4553	1.3926			
C_2	0.0481	0.5601	0.8491	2.1942	0.3966	7.974		
C_3	0.9855	0.5779	0.9656	1.1775	1.2790	2.025	2.705	
C_4	0.4035	2.0600	0.5874	0.9560	0.3966	3.734	3.977	3.506

(資料來源：本研究整理)

因此，C_1 與 C_3 應該合併，成為三個群組的分類結果，$\{C_1, C_2, C_3\}=\{(A, B, C, D, F, G, H, J), (E), (I)\}$，如圖 4-7 所示。

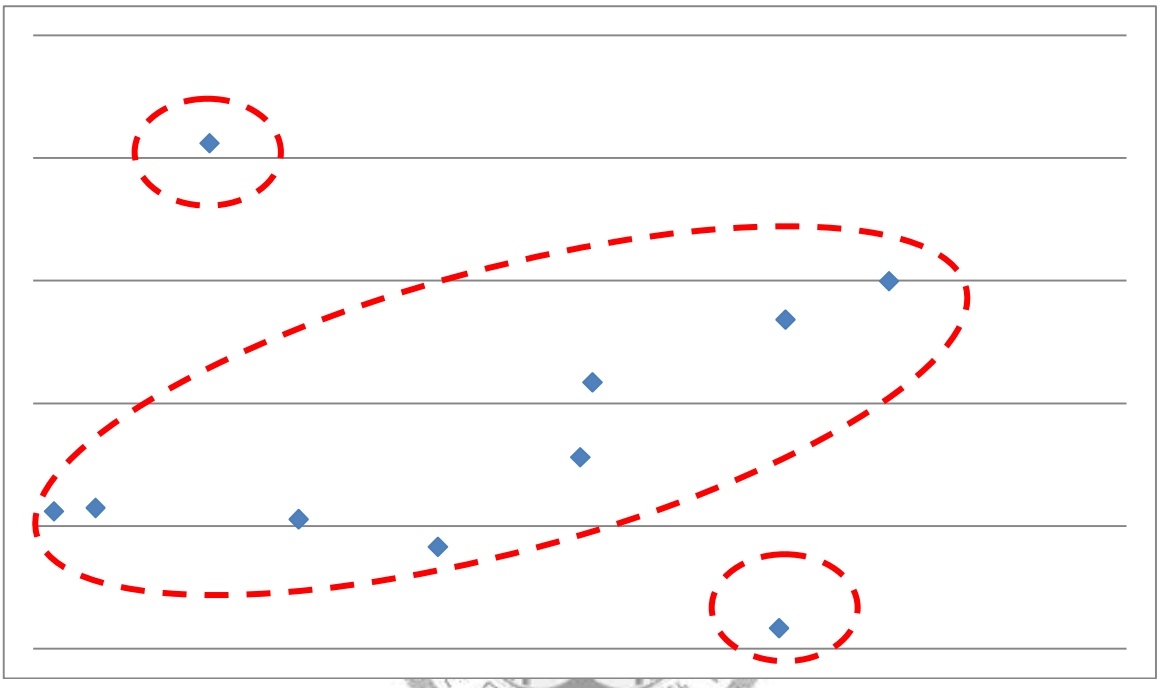

圖 4-7：規則基礎產生初始染色體範例(Step 3)

(資料來源：本研究整理)

上述規則基礎初始染色體的設計優點在於將距離的因素加入判斷規則，優先將應該分開的個體分為不同群集，並且避免了一般距離基礎分類方法的缺點——優先將彼此距離相近的個體綁在一起，在重新群集的過程中保證原本因為距離較遠而分開的個體不會被歸為一類。

P2-3-3. 評量染色體

基因演算法的精神來自於模擬自然界「物競天擇，適者生存」的規則，所有產生的染色體都必須經過設計的適度函式(fitness function)評量，以決定誰應該被淘汰，誰應該留下來繁衍下一代。本研究採用的適度函式即第三章所述的目標函式(1)，群集內總平

均距離

$$aTDG = \frac{2}{\sum_{c=1}^{m} N(G_c^m)(N(G_c^m)-1)} \sum_{i=1}^{N(P)-1} \sum_{j=i+1}^{N(P)} d_{ij}, \qquad \forall g \in G^m$$

以前述染色體 1111211131 為例，根據表 4-12 所列的商品距離，$aTDG$ = 73.860*2/(8*7+1*0+1*0) = 2.6378。

適度函式值愈小表示該染色體品質愈好，因為他最符合本研究的目標。

P2-3-4. 檢查終止條件是否滿足

演算法是否停止繁衍與評量新的染色體將會在這個步驟決定。本研究選擇兩種終止條件：其一為跟效率息息相關的染色體繁衍代數，若可用的代數消耗殆盡，表示分類演算法已耗盡可接受的時間長度，因此將目前最好的染色體分類結果傳出；其二為最佳的分類結果連續維持某特定代數，若該分類結果在多少次產生後代的的過程中，仍然沒有比他更好的分類結果，演算則終止。

P2-3-5. 篩選染色體

經過評量之後，且演算法尚未達到終止條件時，為了保持族群的品質，並保留新成員加入的機會，會將適度最差($aTDG$ 最大)的前兩名染色體從族群中移除。

P2-3-6. 繁衍子代

為了尋求更好的分類結果，在這個步驟透過染色體交配或突變來產生新的染色體。

1. 交配：交配的目的在於希望子代(offspring)保留優良親代(parent)的部分特徵，也就是分類結果，然後因此得到更好的適度表現。在進行交配時，以經過評量之後適度最好的兩條

染色體做為親代，在 N 個基因裡隨機選擇一個點，做為交配的分界點進行單點交配(one-point crossover)。假設表 4-18 所示為族群中適度最好與第二好的染色體，若隨機選擇的交換點為 7，則新的染色體的前半段基因保留適度最好的染色體基因序列，交換點以後取適度第二好的基因序列，形成一個子代。

表 4-18：染色體交配範例

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Parent1	1	1	2	1	1	3	2	3	1	2
Parent2	1	1	1	1	2	1	1	1	3	1
Offspring	1	1	2	1	1	3	1	1	3	1

(資料來源：本研究整理)

2. 突變：突變則不保留任何跟族群裡現有染色體相關的訊息，當演算法決定在此代產生突變，將如同隨機產生初始染色體的做法產生一個突變種(mutant)，希望能產生跳躍的效果，一舉將解的搜尋區域帶往另一個更好的地帶。

P2-3-7. 更新族群成員

完成染色體篩選與子代的繁衍之後，移除適度最差的染色體，將經由交配或突變產生的子代加入族群，以形成新一代的族群，然後再次進行評量的循環。

P2-3-8. 終止後評量

當分類階層第三層的演算法達到終止條件，將用另一個函式評量最後產出的染色體，也就是在第三章中所提的目標函式(2)，群集間距離總平均

$$Max\ AGD = \frac{2}{m(m-1)} \sum_{k=1}^{m-1} \sum_{j=k+1}^{m} d_{\bar{k}\bar{j}}, \qquad \text{for } M_{min} \le m \le M_{max}$$

將屬於同群集的所有商品參數值平均，做為代表該群集的中心點

$(\bar{k}、\bar{j})$，計算群集之間的總平均距離。以表 4-17 所示的群集中心點為例，AGD = 23.921*2/(4*3) = 3.9868。

在此分類階層搜尋最適分類結果有一個限制，就是必須在固定分群數(m)之下進行搜尋，然而 m 也限制了分類結果的樣貌，因此也必須調整 m——從 M_{min} 到 M_{max}——進行搜尋，最後以 AGD 做為比較基準，能夠產生最大 AGD 值的 m 與其對應的染色體就是那 N 個商品在這分類階層中最適當的分類結果。

第五節　複雜度分析

本研究之演算法主要包過前置作業、分類架構建立與預測效果評估。其中前置作業在每次執行本演算法時只會做一次，預測效果評估也只會做一次，對整體演算法效率影響極小；最複雜的部分即是以基因演算法為基礎的分類架構建立演算法，本節將針對此部分做複雜度分析。

複雜度分析所使用的參數詳細說明可參考第三章第三節第二子節，本節將沿用重要參數如下：

$N(P)$：表示待分類商品數目。

T_i：商品 i 所擁有的銷售歷史記錄總期數。

m：表示欲區分的群集數目。

M_{min}：表示群集數目之下限。

M_{max}：表示群集數目之上限。

系統相關參數：

GC：產生有效演化的代數

p：產生突變的機率

在前置作業的部分，必須將每個待分類的商品的銷售歷史記錄轉換為時間序列指標，每個商品有 T_i 期記錄，共 $N(P)$個商品，

所需的矩陣運算的時間複雜度為 $O(T*N(P))$。在分類演算法中，以亂數產生初始染色體所需的時間極短，而以規則基礎產生的初始染色體需要建立商品兩兩之間的距離表，以及多次搜尋距離表中最大值，其時間複雜度為 $O(N(P)^2)$；在染色體演化的過程中，每一代的交配與突變動作皆相當簡單，其時間複雜度可寫作 $O(N(P))$；最不理想的狀況即是必須完全耗盡系統所設定的有效演化代數 (GC)才能達成此迴圈停止條件，產生品質足夠好的解。前述分類動作所預設的區分群集數目 (m)必須從 M_{min} 試到 M_{max}，才能決定 P 集合中的商品的最佳分類結果，因此分類架構建立的時間複雜度為 $O(N(P)^2 + (M_{max} - M_{max})*GC*N(P))$。

經過上述前置作業與分類架構建立的時間複雜度分析，所有變數皆不在指數項，因此本研究所提出的演算法不為 NP 演算法。

第五章　系統說明與模式分析

第一節　分類架構建立系統說明

為了驗證本研究所提出之啟發式演算法 DMAPC 的確有其可行性與實質效益，本研究實際建置一商品分類架構建立系統，並引入實際案例進行測試。本節將於第一子節中介紹此系統所使用的資料結構，在第二子節中展示系統畫面。

本研究的系統開發與測試皆使用 Microsoft Windows Server 2003 做為作業系統，Microsoft SQL Server 2005 做為資料庫，Microsoft Visual Studio 2005 做為開發環境；硬體方面的 CPU 時脈為 Intel® Core™2 Duo E8300 @2.83GHz，記憶體大小為 2.99GB。

5-1-1 資料結構

本研究所使用的資料結構主要可分為五大類：

一、商品銷售資訊

此分類之資料為前置作業所需之輸入資訊，包括日曆主檔、商品主檔、商品銷售記錄。

1. 日曆主檔(CALENDAR)

此資料主檔記錄時間週期定義的資訊，是所有與時間相關分析的基礎，包含週期索引、顯示編號、起始時間、終止時間、所屬月份、所屬季別、年份，如表 5-1 所示。

表 5-21：日曆主檔

Field Name	Description	Type	Length
CALNDR_NO	Calendar Number	int	
DISPLAY_NO	Semantic Calendar Number	String	10
FROM_DATE	The starting time of the period	DateTime	
TO_DATE	The end time of the period	DateTime	
MONTH_NUM	The month number of the period	int	
SEASON_NUM	The season number of the period	int	
YEAR_NUM	The year number of the period	int	

(資料來源：本研究整理)

2. 商品主檔(ItemMaster)

此資料主檔記錄所有待分類商品的基本資料，包括商品標準代碼、產品名稱、製造商名稱、建議售價、上市日期、最後銷貨日期，如表 5-2 所示。

表 5-22：商品主檔

Field Name	Description	Type	Length
ItemNo	Universal Product Number	varchar	13
ProdName	Product Name	varchar	100
Manufacturer	Manufacturer	varchar	100
Sugst_price	Suggested Price	decimal	11, 2
Issue_Date	The date when the product became available in the market	DateTime	
Off_Shelf_Date	The date when the product were no longer for sale	DateTime	

(資料來源：本研究整理)

3. 商品銷售記錄(Sales_FACT)

此資料主檔記錄商品在各商店每日的售出記錄，是銷售記錄最原始的資料，包括標準商品代碼、銷售數量、銷售金額、銷售日期、商店編號，如表 5-3 所示。

表 5-23：商品銷售記錄

Field Name	Description	Type	Length
ItemNo	Universal Product Number	varchar	13
SalesQty	Sales Quantity	decimal	18, 4
SalesAmt	Sales Amount	decimal	18, 4

SalesDate	Sales Date	DateTime	
Store	Store Number	varchar	10

(資料來源：本研究整理)

二、商品銷售整合記錄

此類資料為前述商品銷售記錄經過前置作業，依各商店所屬地區整合各個商品銷售記錄的中繼資料表，依不同的週期長度，分別整合成商品週地區銷售記錄(SA_Item_CALNDR)及商品月季年地區銷售記錄(SA_Item_YSM)，並記錄前置作業中產生的銷售量平滑值、預測值，以計算季節性指標與長期趨勢指標，包括標準商品代碼、商品名稱、週期編號、地區編號、月份編號、季節編號、銷售數量、去除季節性效應後的銷售量、使用不同參數設定並加回季節性效應的銷售量預測值，如表 5-4 及 5-5 所示。

表 5-24：商品週地區銷售記錄

Field Name	Description	Type	Length
ItemNo	Universal Item Number	varchar	13
Geographic	Sales Region	String	15
CALNDR_NO	Calendar Number	int	
MONTH_NUM	Month Number	int	
SEASON_NUM	Season Number	int	
SQty_total	Total Sales Quantity within given period	decimal	18, 4
SQty_total_12sf	Sales Quantity after removing monthly seasonal effect	decimal	18, 4
SQty_total_4sf	Sales Quantity after removing quarterly seasonal effect	decimal	18, 4
SQty_total_Pred_W_sl	Sales Quantity Predict value using simple linear regression model and assume no seasonality	decimal	18, 4
SQty_total_Pred_M_sl	Sales Quantity Predict value using simple linear regression model and assume 12-period seasonality	decimal	18, 4
SQty_total_Pred_S_sl	Sales Quantity Predict value using simple linear regression model and assume 4-period seasonality	decimal	18, 4
SQty_total_Pred_W_qu	Sales Quantity Predict value using quadratic	decimal	18, 4

	regression model and assume no seasonality		
SQty_total_Pred_M_qu	Sales Quantity Predict value using quadratic regression model and assume 12-period seasonality	decimal	18, 4
SQty_total_Pred_S_qu	Sales Quantity Predict value using quadratic regression and assume 4-period seasonality	decimal	18, 4

(資料來源：本研究整理)

表 5-25：商品月季年地區銷售記錄

Field Name	Description	Type	Length
ItemNo	Universal Item Number	varchar	13
Geographic	Geographic	String	
MONTH_NUM	Month Number, when 0 represents the whole season	int	
SEASON_NUM	Season Number, when 0 represents the whole year	int	
YEAR_NUM	Year Number	int	
SQty_total	Total Sales Quantity within given period	decimal	18, 4
SQty_total_smoothed	Smoothed Sales Quantity using simple linear regression model	decimal	18, 4

(資料來源：本研究整理)

三、商品銷售發展趨勢指標(TSINDEX)

此類資料表為前置作業的輸出，且為後續以基因演算法為基礎之最適分類架構搜尋演算法的輸入。記錄每一商品採用不同參數設定時所得出的參數值與預測準確度——做為該模型參數設定對該商品的適用度。包括標準商品代碼，參數組合代碼、長期趨勢指標、季節性指標、預測誤差量化標準(採用 MAPE)，如表 5-6 所示。

表 5-26：商品銷售趨勢指標

Field Name	Description	Type	Length
ItemNo	Universal Item Number	varchar	13
Mode	Parameter Configuration Code	String	3
b_0	Trend Index 1	decimal	18, 6
b_1	Trend Index 2	decimal	18, 6

b_2	Trend Index 3	decimal	18, 6
s_1	Seasonal Index 1	decimal	18, 6
s_2	Seasonal Index 2	decimal	18, 6
s_3	Seasonal Index 3	decimal	18, 6
s_4	Seasonal Index 4	decimal	18, 6
s_5	Seasonal Index 5	decimal	18, 6
s_6	Seasonal Index 6	decimal	18, 6
s_7	Seasonal Index 7	decimal	18, 6
s_8	Seasonal Index 8	decimal	18, 6
s_9	Seasonal Index 9	decimal	18, 6
s_10	Seasonal Index 10	decimal	18, 6
s_11	Seasonal Index 11	decimal	18, 6
s_12	Seasonal Index 12	decimal	18, 6
MAPE	Quantitative Fitness Evaluation score using MAPE	decimal	18, 6

(資料來源：本研究整理)

四、基因演算法記錄(GA_RECORD)

此類資料記錄經過基因演算法演化後所找到最佳的染色體，也就是最適合目標函式的分類結果。包括待分類商品數目、參數組合代碼、指定分群數目、指定演化代數、目標函式值一、目標函式值二、耗費時間、商品清單、最佳染色體基因序列，如表 5-7 所示。

表 5-27：基因演算法記錄

Field Name	Description	Type	Length
No_Item	Number of Item to be classified	int	
Mode	Parameter Configuration Code	String	3
M	Number of Group	int	
No_Generation	Number of Generation	int	
Min_aTDG	Fitness Value	decimal	
AGD	Objective Function Value	decimal	
RunTime	Total Time Consumed	decimal	
ItemList	Item Number Sequence	String	
BestChromosome	Best Gene Sequence	String	

(資料來源：本研究整理)

五、分類結果資訊

此類資料記錄經由前述步驟所得出的分類結果，並將其轉換為一階層式分類架構，包括分類主檔、類別階層關係、類別商品配對表。

1. 分類主檔(Classification)

此資料主檔記錄各類別基本資料，包括分類架構編號、類別編號、類別階層，如表 5-8 所示。

表 5-28：分類主檔

Field Name	Description	Type	Length
Class_Type_No	The number of Class Type	int	
ClassID	Class ID	varchar	15
Level	Class Level	int	

(資料來源：本研究整理)

2. 類別階層關係(ClassRelationship)

此資料記錄類別階層之間父類別與子類別的從屬關係，包括分類架構編號、父類別編號、子類別編號，如表 5-9 所示。

表 5-29：類別階層關係

Field Name	Description	Type	Length
Class_Type_No	The number of Class Type	int	
ParentClassID	The ClassID of Parent Node in Classification Tree	varchar	15
ChildClassID	The ClassID of Child Node in Classification Tree	varchar	15

(資料來源：本研究整理)

3. 類別商品對照(ClassItemMap)

此資料記錄每一商品與最小分類的對照關係，包括分類架構編號、最小分類編號、標準商品代碼，如表 5-10 所示。

表 5-30：類別商品對照

Field Name	Description	Type	Length
Class_Type_No	The number of Class Type	int	
ClassID	Class ID	varchar	15
ItemNo	Universal Item Number	varchar	13

(資料來源：本研究整理)

5-1-2 系統畫面與執行步驟

本研究所開發的啟發式演算法系統功能主要分為兩個部分：第一部分為前置作業(Data Transformation)與分類架構建立(do Classification)，如圖 5-1 所示。

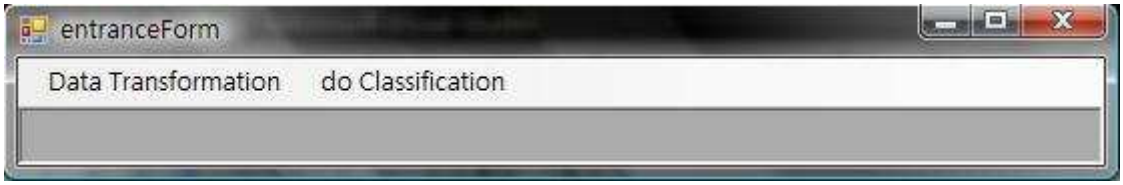

圖 5-11：系統主要功能畫面

(資料來源：本研究整理)

執行前置作業時，如圖 5-2 所示，依序執行下列四個步驟：(1)銷售記錄平滑化(Smooth)，將各商品依不同週期長度整合的銷售記錄以簡單線性迴歸加以平滑化；(2)去除季節性效應(Remove SI effect)，辨識各商品的銷售記錄中可能存在的季節性指標數值，並計算去除季節性效應之後的銷售量；(3)長期趨勢分析(Trend Analysis)，計算各商品的迴歸趨勢線係數，並以此趨勢線產生預測值並加回季節性效應；(4)評估參數設定適度(Evaluate)，計算各商品對不同模型參數設定的適度，以預測誤差做為比較標準，然後保留其中最適合的參數設定值。

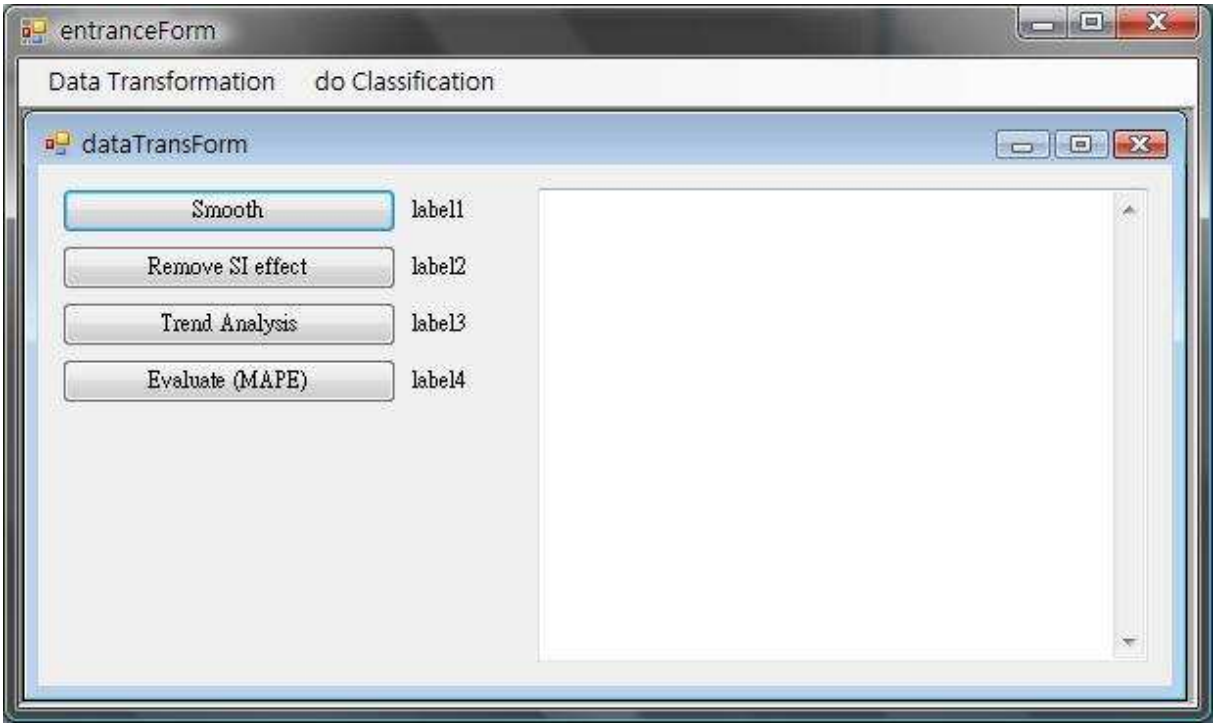

圖 5-12：前置作業系統畫面

(資料來源：本研究整理)

以某商品經共 90 週的銷售記錄為例，經統整為 20 個月、7 季，使用簡單線性迴歸模型平滑之後的銷售量如圖 5-3、5-4 所示，詳細資料請參閱附錄表 A-2。

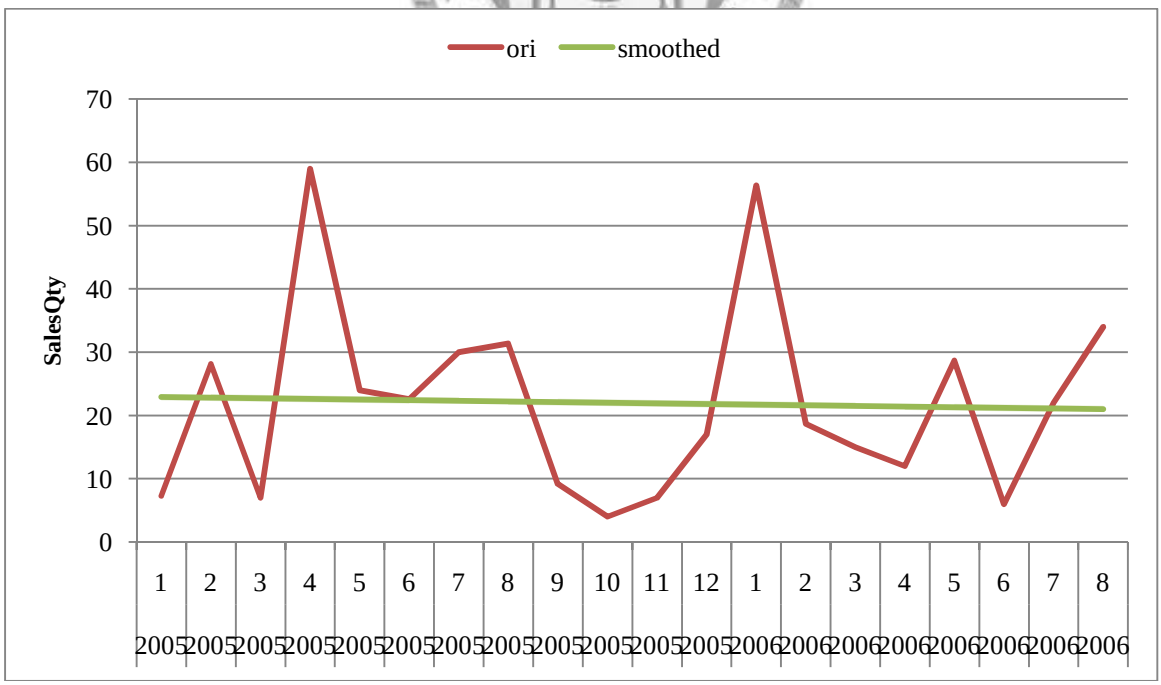

圖 5-13：商品月銷售記錄與平滑效果

(資料來源：本研究整理)

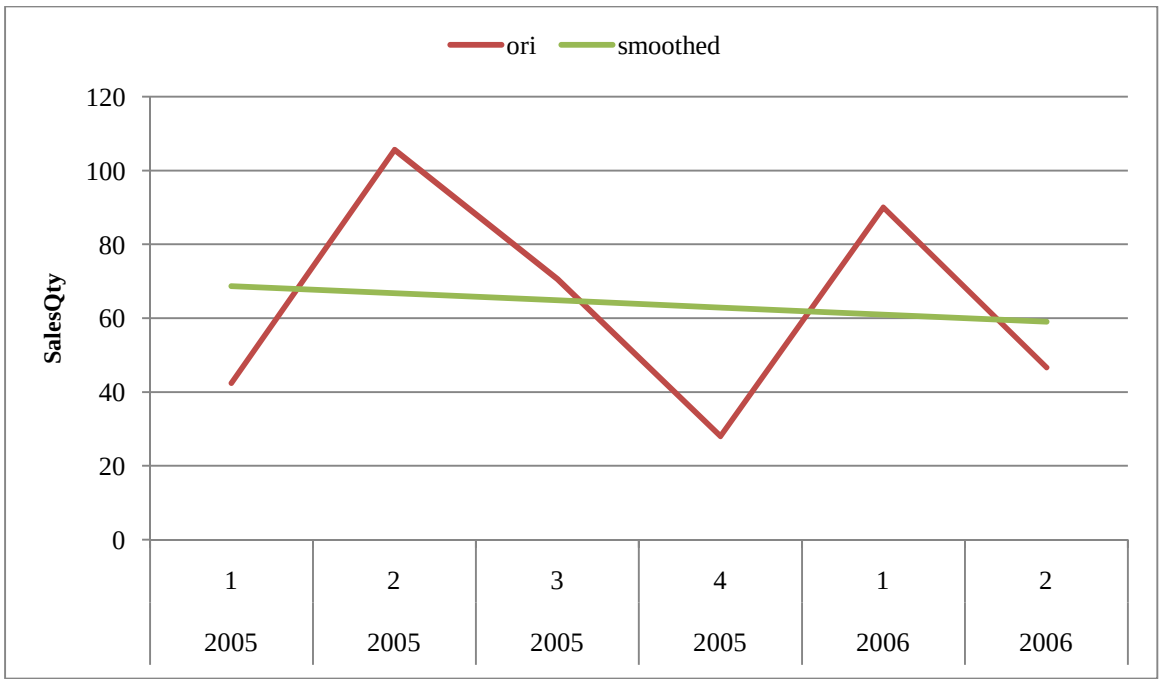

圖 5-14：商品季銷售記錄與平滑效果

(資料來源：本研究整理)

使用原始銷售記錄與平滑化之後的數值比較計算得出可能存在的季節性指標，如表 5-11 所示，且去除季節性效果如圖 5-5 所示。

表 5-31：商品季節性指標

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}
12 期季節性	1.611	1.1597	0.5561	1.7532	1.3346	0.7134	1.3201	1.6764	0.4592	0.201	0.3533	0.8621
4 期季節性	1.1117	1.2598	1.1559	0.4727								

(資料來源：本研究整理)

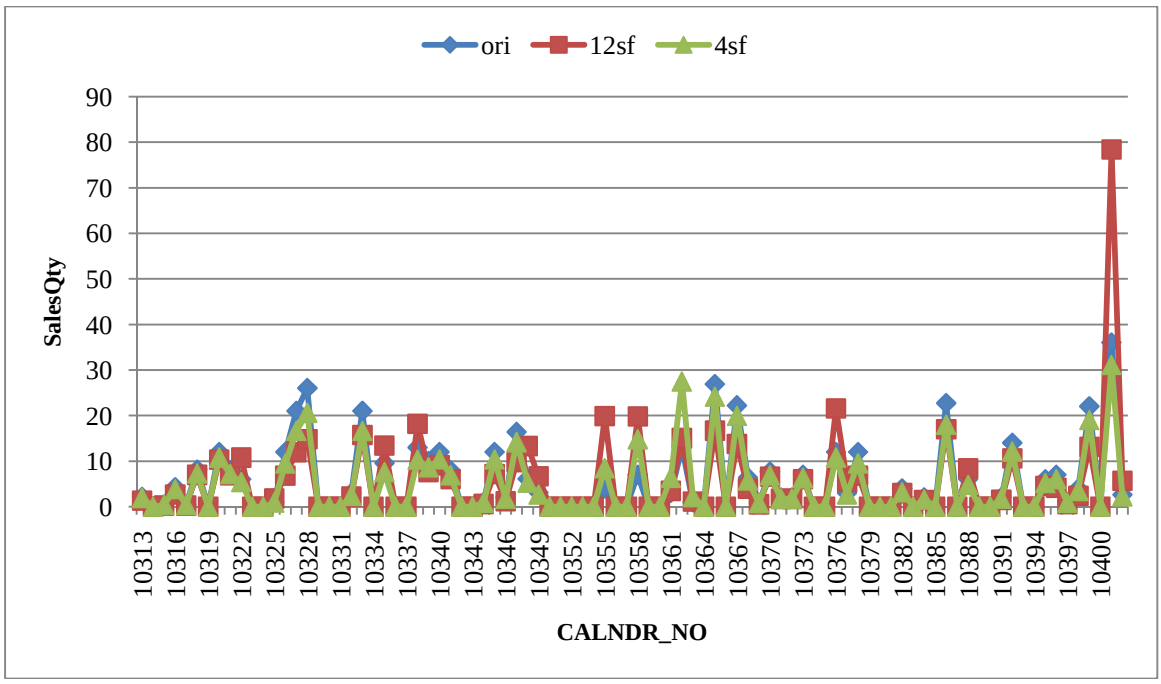

圖 5-15：商品週銷售記錄與去除季節性效果

(資料來源：本研究整理)

然後使用去除季節性效果的銷售記錄進行長期趨勢分析，綜合前述三種時間週期分析與兩種迴歸模式，每一商品會得到六種不同模式設定下所產生的預測值，如圖 5-6 所示。

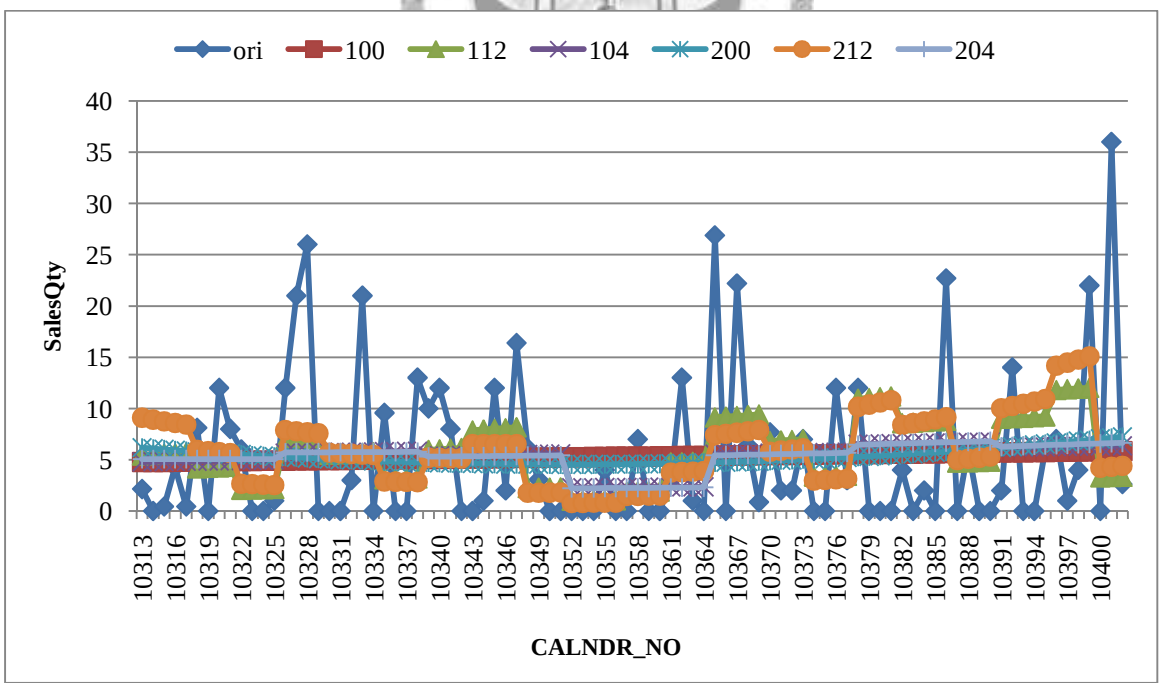

圖 5-16：商品週銷售記錄與六種預測模式組合預測值

*Mode 代碼說明：第一碼 1 = simple linear regression，2 = quadratic regression

第二，三碼 00 = 無季節性，04 = 四期季節性，12 = 十二期季節性

(資料來源：本研究整理)

所有商品經前述步驟分析轉換為指標數值的形式儲存，例如

某商品採用四期季節性與簡單線性迴歸分析，其指標(b_0, b_1, s_1, s_2, s_3, s_4) = (4.310024, 0.013465, 1.111700, 1.259800, 1.155900, 0.472700)。於附錄表 A-9 中列舉有 11 項商品經資料轉換的指標數值結果。最後每個商品會從六種模式中選擇一種預測誤差值最小的保留。

完成前置作業之後，執行最適分類結果搜尋(do Classification)，如圖 5-7 所示。首先依序針對各個區塊進行設定：(1)設定本次引入分類演算法的商品條件，即最適參數組合；(2)設定本次搜尋群集數目範圍(M_{min}~M_{max})與基因演算法所需之基本資訊，包括族群大小(populationSize)，表示族群每一代可容納的最大個體數目；突變機率(mutationRate)，表示每一代會產生一個突變種的機率；以及其他初始化亂數產生器的種子資訊；(3)設定搜尋停止條件，包括最多演化代數(# of Generation)，當基因演算法演化代數抵達這個上限，則搜尋停止；最佳染色體停留代數(stayBest)，當同一個分類結果也就是同樣的基因序列保留在族群裡，且沒有任何已存在或新出現的個體比它更好，時間長達設定的代數，則搜尋停止；或突出的個體(outstandingBest)，表示當某一最佳個體的適度函式值優於第二好的個體的適度函式值高達設定的倍數，則搜尋停止。

完成設定之後執行讀取資料與設定(set'N'readData)，然後即可開始主要搜尋步驟(START!)。讀取到的待分類商品數目與搜尋建議分類結果與對應目標函式值皆會顯示於畫面中指定的欄位，最後執行分類架構建立(buildup Classification)。

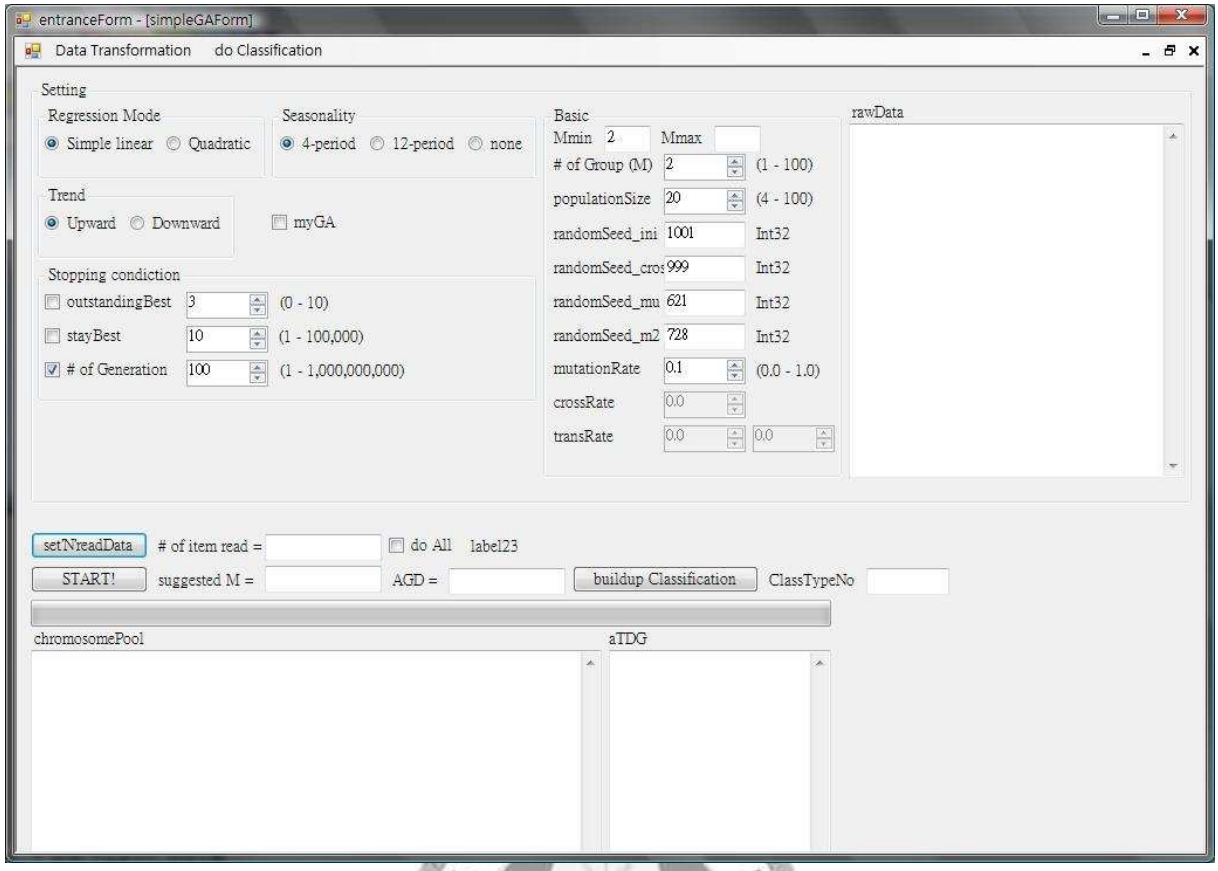

圖 5-17：最適分類結果搜尋系統畫面

(資料來源：本研究整理)

以前述 11 項商品為例，因為其最適預測模式皆為簡單線性迴歸與四期季節性(代碼 104)，所以同時引入進行最適分類結果搜尋。搜尋範圍設定(M_{min}, M_{max}) = (2, 10)，搜尋停止條件為最大演化代數為 1000，相關基本設定如圖 5-7 所示。經過搜尋之後，結果如表 5-12 所示，系統建議將此 11 項商品區分為兩個小分類，可以得到最大的群集間距離總平均(AGD = 12.81201)，其類別與商品對照可參照附錄表 A-7。

表 5-32：最適分類結果搜尋記錄

# of Item	M	# of Generation	Min aTDG	AGD	runTime
11	10	1000	0.011873	5.228166	5278.047
11	9	1000	0.041677	5.67619	756.3438
11	8	1000	0.067675	6.024996	25.21875
11	7	1000	0.074766	6.607518	3.8125
11	6	1000	0.088094	7.364112	1.078125
11	5	1000	0.244022	7.394477	0.484375
11	4	1000	0.290311	6.302402	0.328125

11	3	1000	0.372993	7.181878	0.28125
11	2	1000	0.735051	12.81201	0.4375

(資料來源：本研究整理)

完成 DMAPC 之後，即進入預測效果評估階段。將經過前述步驟所建立的分類架構匯入需求預測學習系統中，執行一次完整的預測流程。其主要步驟包括：

(1)銷售資料整理，清除為零或負的銷售記錄，計算並去除已知的特殊事件與促銷效果，系統畫面如圖 5-8 所示。

圖 5-18：銷售資料整理系統畫面

(資料來源：本研究整理)

(2)統整銷售記錄，將原本僅記錄單項商品於某日於某分店的銷售記錄依銷售區域以及商品分類架構與不同的時間週期長度加以合併儲存，形成某一段時間週期的某一區某一類別的商品總共的銷

售記錄，做為後續預測學習分析的基本資料。執行步驟如圖 5-9 所示，產生單品週期資料、產生單品年季月資料、產生分類週期資料、產生類別年季月資料。

圖 5-19：銷售記錄統整系統畫面

(資料來源：本研究整理)

(3)計算分配比率，為了將類別商品的預測值適當地分配給同類別中各個商品，在此步驟計算各個從長週期給短週期、大區域給小分店，或類別商品給單品的分配比率，系統畫面如圖 5-10 所示。

圖 5-20：計算分配比率系統畫面

(資料來源：本研究整理)

(4)選擇最佳參數，如圖 5-11 所示，設定要分析的銷售區域與指定分類商品，並且選擇列入考慮的預測模式方法，包括有迴歸模型、指數平滑法、趨勢指數平滑法、移動平均法、前期同季平移法、前期同季平均法。每個類別商品最後會得到一組建議的最佳預測模式，包括預測方法、使用參數、預測時間類別、與對應產生的預測誤差。以前述 11 項商品範例，經過學習之後選擇的最佳預測模型為：使用月的週間資料，採指數平滑法並設 $\alpha = 0.1$ 與使用週的週末資料，採指數平滑法並設 $\alpha = 0.3$ 在合併成整週的預測銷售量，其預測誤差 MAPE 為 10.53%

圖 5-21：選擇最佳預測模式系統畫面

(資料來源：本研究整理)

(5)預測未來銷售，如圖 5-12 所示，設定預測區域、預測天數與預測類別商品，從系統日期今天開始往後產生數天的預測銷售量。此一步驟所產生之預測值方具有實質規劃意義，未來將更進一步轉換為實際的訂單，但是本研究並不需要後續的步驟。

圖 5-22：預測未來銷售系統畫面

(資料來源：本研究整理)

(6)計算暢銷商品 MAPE，如圖 5-13 所示，分別針對學習階段與預測階段所產生的預測值與資料庫中的歷史記錄比較，評估預測準確度。

圖 5-23：計算暢銷商品 MAPE 系統畫面

(資料來源：本研究整理)

前述預測模型學習系統的執行步驟在變更分類架構或取得新的資訊時，皆必須重新計算學習，尤其變更分類架構之後，必須重新統整合併銷售記錄並計算分配比率，以保證學習結果的正確性、有效性與預測結果的準確性。

第二節　實例分析

本節將使用真實案例進行本研究之啟發性演算法的測試，藉以證明本演算法之可行性。

5-2-1 驗證方法與環境

本研究的最終應用目的是為了提高商品銷售量預測的準確度，針對擁有足夠銷售記錄的一般商品，使用[2]所建置的需求預測學習系統進行驗證。首先以銷售歷史記錄中移除最後一個月的資料做為訓練資料集進行最佳預測模型學習，最後模擬產生最後一個月的預測銷售量，與資料庫中的實際值比較，以 MAPE 做為預測誤差的量化標準。

使用不同的分類架構下——包括廠商所提供的分類架構與 DMAPC 所建立的分類架構——會導致經統整得出的類別銷售量擁有不同的趨勢發展，然後可能適合不同的預測模式。最後仍然必須以單項商品的預測誤差做為標準，比較使用不同分類架構所產生的差異。

5-2-2 案例簡介

一、某知名茶飲料商

此案例所提供商品與銷售歷史資訊共包括 319 個商品，原本以三層的分類架構區分各項商品——產品型態、產品口味、產品包裝規格，總共有 58 個小分類。茶飲料商品具有生命週期長，銷售量大且季節性波動趨勢明顯，同類商品在長期趨勢與季節性表現上相似，因此不調整分類架構即可得出可接受的預測結果。

二、某連鎖藥粧店

此案例在其擁有的通路店面販售不同品牌的彩粧商品及保養品，為了管理方便，廠商以品牌及產品用途予以分類，共有 39 個小分類。此案例的商品銷售量起伏變動大，且經常有短期的異常銷售高峰出現；另外，商品生命週期短，款式細且雜，下市之後通常由新商品取代舊商品的地位，因此較難以預測。

第三節　實例分析結果

5-3-1 案例一：某知名茶飲料商

經過 DMAPC 分析之後，產生 30 個小分類，少於原本的 58 個。其中以選擇了 12 期季節性的商品佔了半數，選擇 4 期季節性佔了三分之一，其餘才是無明顯季節性。

仔細比較廠商所提供的分類架構與 DMAPC 所分析的分類結果，可以發現原本屬於同類別的商品，其實在長期趨勢的正負成長與季節性波動描述上皆有所不同，如表 5-13 所示，應該將其分開。

表 5-33：分類結果差異分析範例

ItemNo	Trend (P: Positive, N: Negative)	Seasonality
A	P	12
B	N	12
C	P	12

D	P	12
E	N	4
F	N	0
G	N	12
H	N	12
I	P	12
J	P	12

(資料來源：本研究整理)

在學習階段的評量結果，以累積銷售金額排名前百大商品為比較對象，原本的分類架構的平均預測誤差為 113.99%，而使用 DMAPC 所建立的分類架構的平均預測誤差為 63.54%，表現最差的是隨機任意群集的結果，平均預測誤差高達 195.46%。

使用學習後建議的最佳預測模式與參數實際模擬預測往後 30 天的銷售量，同樣以累積銷售金額排名前百大商品為比較對象，原本的分類架構所產生的平均預測誤差為 67.90%，而使用 DMAPC 分析所得的平均預測誤差為 81.61%，最差的仍然是使用隨機指派的結果，為 119.52%。若個別比較則可以發現，前百大暢銷商品中，使用 DMAPC 有 73 項商品的預測誤差低於使用原本的分類架構，或是不大於 10%。

藉由此案例可以發現，DMAPC 面對長期趨勢與季節性明顯且記錄時間長、銷售量穩定的商品集合時，可以有效分辨出應該群集在一起的商品，進而有效提升預測準確度。

5-3-2 案例二：某連鎖藥粧店

經 DMAPC 分析後，此案例共約 450 項商品被分作 30 個小分類，略少於原本所分的 39 個小分類。個別商品在分析過程中所選擇的長期趨勢與季節性組合模式以無季節性與 12 期季節性佔了絕大多數。因為本案例的銷售記錄特性，比較結果發現，DMAPC 的效果不大，多數商品因為記錄過短的關係無法分析出有效的季節

性波動，所以原本同類的商品在 DMAPC 的分類結果中仍然屬於同類。

本研究在此案例中，再加入單項商品直接進行學習預測的情境與其他分類架構比較。預測百大暢銷商品未來 30 天銷售量，以 DMAPC 所產生的預測值最為準確，平均預測誤差為 32.95%，優於原本以品牌分類的 39.63%;若不做銷售記錄合併直接進行學習，所產生的平均預測誤差最大，達 57.59%。

第四節　適用性分析

5-4-1 效率分析

本研究所提出的啟發式演算法以第二階段的最適分類結果搜尋最為耗時，其所要花費時間受系統參數設定影響，包括搜尋範圍(M_{min}~M_{max})與搜尋停止條件(演化總代數與最佳解停留代數)。

假設有 N 個待分類的商品，且 N 大於 30，完整的搜尋範圍應該在兩個不分類的極端狀況之內，也就是將(M_{min}, M_{max})設為(2, N-1)。但是經過實驗發現，最適分類結果的分群數都不曾出現於超過 N/2，甚至不曾超過 10，如圖 5-14 所示。因此建議將搜尋範圍設定為(2, N/2-1)；若 N 的數目不多，即可將所有非極端分類群集數搜尋過。

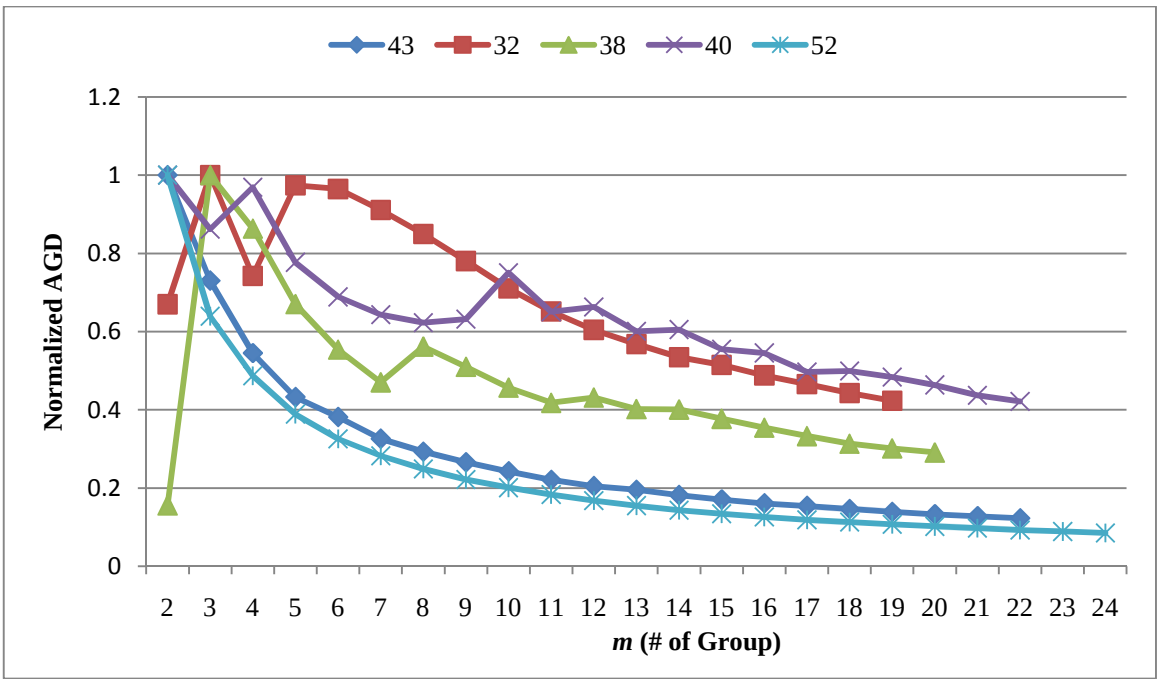

圖 5-24：建議搜尋範圍 (數列代表不同的待分類商品數目)

(資料來源：本研究整理)

以基因演算法為基礎的最適解搜尋方法以競爭為原則進行，從這一代進行到下一代必須以一個新的可行競爭者出現才算數。經實驗發現，在第四章所描述規則基礎產生的初始染色體其品質優良可以保留在族群裡數十代，或是在前期就經由交配產生更好的染色體。因此建議將演化總代數設為 200 即可。若要更加縮短搜尋時間則可設定最佳解停留代數，惟此代數必須小於總演化代數的一半才有意義。

使用前述建議系統參數進行分類，兩個實際案例都可以在十分鐘內完成分類架構建立。

第六章　結論

第一節　總論

本研究將分析商品的銷售發展趨勢，以時間序列指標——長期趨勢及季節性波動——做為該商品的特徵，藉以群集相似商品建立分類架構，達成改進商品預測準確度的目標。

過去研究時間序列資料預測方法者，皆未探討分類架構對整合商品銷售記錄的影響，最嚴重的情況就是把長期趨勢發展相反的商品歸為一類，使得整個類別商品的長期發展趨勢遭受扭曲或模糊；研究資料探勘者，也尚未將觸角延伸至隨時間累積的時間序列資料上，因為不同的觀察對象並非在所有的觀察時間內都有觀察值，導致如此的資料型態無法引入資料探勘的分析模型中。本研究提出一模型將商品銷售歷史記錄轉換為數個時間序列指標形成描述該商品銷售發展趨勢的特徵向量，使得原本擁有不同數量屬性的商品可以適用於資料探勘分析模型加以計算彼此之間的相似度。

本研究針對最適分類結果提出一兩階段最佳化目標模型：在內層固定分群數目時，尋求群集內樣本距離總平均最小化；然後以不同群集之間距離總平均最大者為外層目標。在這個模型中，必須搜尋的可行解集合大小將以待分類商品數目為指數的速度成長，以全域搜尋法在商品數目達到實際應用規模前，就無法使用有限的運算資源於可接受的時間範圍內找到最佳解。本研究針對

此限制提出一啟發式演算法，以基因演算法為基礎，大幅縮短近似最佳解搜尋時間，並且足以處理具有實際應用規模的商品數量。

本研究的啟發式演算法流程為：資料轉換，使用時間序列分析方法將各商品的銷售量轉換為數個指標；然後以啟發式演算法搜尋最適群集分類結果，並依此建立分類架構；最後匯入需求預測方法學習系統進行效果評估。

本研究可以在合理的時間內找出一組可行的分類結果，並且經過實例分析驗證將有效提升第三階段學習系統的品質，為類別商品選擇較好的預測模式提升預測準確度。

在實務層面上，適合引入本研究的商品類型並未限制，只要擁有足夠的銷售歷史記錄進行資料轉換即可，因此本研究所提出的方法可以適用於各種產業商品。本研究建議商品項目龐雜的使用者採用此分類分析方法，來區分不知道如何分類的商品集合，以得到較好的需求預測結果，並且同時藉此觀察商品在銷售發展趨勢上的異同，發現不適合歸為同類的商品或跨類別卻具有相似銷售發展趨勢的商品，協助訂補貨的策略調整。

第二節　未來研究方向

本研究針對長銷型商品，希望是擁有長達一年以上銷售記錄的商品，才具有足夠的資訊進行有效的季節性與長期趨勢分析，但是當面對銷售不滿一年的商品，或銷售量起伏變動劇烈，無法有效以長期趨勢與季節性來描述時，將降低本研究所提出的效果。

因此在資料轉換之前，未來可以加入其他商品銷售趨勢分析因素，例如去除流行性影響，使整體銷售趨勢更加明顯；另外研

究另一套方法處理銷售記錄短的商品與其他商品之間的相似度定義。

本研究所提出之方法亦可用於協助改進供應鏈管理其他功能，例如存貨重要性管理、訂補貨策略制定。

參考文獻

[1] 丁恬文，流通業協同規劃預測補貨解決方案，國立台灣大學資訊管理研究所碩士論文，民國 96 年。

[2] 陳靜枝與蔣明晃，需求預測模式之研究期末報告，財團法人工業技術研究院，民國 94 年。

[3] Cardoso, G. and F. Gomide, “Newspaper demand prediction and replacement model based on fuzzy clustering and rules,” Information Sciences, Vol. 177, Issue 21, 2007, pp. 4799-4809.

[4] Carvalho, D. R. and A. A. Freitas, “A hybrid decision tree/genetic algorithm method for data mining,” Information Sciences, Vol. 163, Issues 1-3, 2004, pp. 13-35.

[5] Chiu, C., “A case-based customer classification approach for direct marketing,” Expert Systems with Applications, Vol. 22, Issue 2, 2002, pp. 163-168.

[6] Chopra, S. and P. Meindl, Supply Chain Management: Strategy, Planning, and Operation, Second Edition, Pearson Education International, USA, 2004.

[7] Forina, M., S. Lanteri, and S. Rosso, “Confidence intervals of the prediction ability and performance scores of classifications methods,” Chemometrics and Intelligent Laboratory Systems, Vol. 57, Issue 2, 2001, pp. 121-132.

[8] Guyon, I. and A. Elisseeff, “An Introduction to Variable and Feature Selection,” Journal of Machine Learning Research, Vol. 3, 2003, pp. 1157-1182.

[9] Han, J. and M. Kamber, Data Mining: Concepts and Techniques, Second Edition, Morgan Kaufmann Publishers, USA, 2006.

[10] Hanczar, B., M. Courtine, A. Benis, C. Hennegar, K. Clément, and J. D. Zucker, “Improving Classification of Microarray Data using Prototype-based Feature Selection,” ACM SIGKDD Explorations Newsletter, Vol. 5, Issue 2, 2003, pp. 23-30.

[11] Jain, A. K., M. N. Murty, and P. J. Flynn, “Data Clustering: A Review,” ACM Computing Surveys (CSUR), Vol. 31, Issue 3, 1999, pp. 264-323.

[12] Kahn, K. B., “Benchmarking Sales Forecasting Performance Measures,” The Journal of Business Forecasting Methods & Systems, Vol. 17, No. 4, Winter 1998/1999, pp. 19-23.

[13] Keller, G., Statistics for Management and Economics, Seventh Edition, Thomson Brooks/Cole, USA, 2005.

[14] Kim, D., S. Lee, J. Chun, and J. Lee, “A Semantic Classification Model for e-Catalogs,” Proceeding of the IEEE International Conference on e-Commerce Technology (CEC 2004), 2004, pp. 85-92.

[15] Kotsiantis, S. B., “Supervised Machine Learning: A Review of Classification

Techniques," Informatica, Vol. 31, 2007, pp. 249-268.
[16] Lee, Y. and C. K. Lee, "Classification of multiple cancer types by multicategory support vector machines using gene expression data," Bioinformatics, Vol. 19, No. 9, 2003, pp. 1132-1139
[17] Li, R. and Z. Wang, "Mining classification rules using rough sets and neural networks," European Journal of Operational Research, Vol. 157, Issue 2, 2004, pp. 439-448.
[18] Li, X. B., "A scalable decision tree system and its application in pattern recognition and intrusion detection," Decision Support Systems, Vol. 41, Issue 1, 2005, pp. 112-130.
[19] Liu, H. and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," IEEE Transactions on Knowledge and Data Engineering, Vol. 17. Issue 4, 2005, pp. 491-502.
[20] Lo, V. S. Y., "The True Lift Model – A Novel Data Mining Approach to Response Modeling in Database Marketing," ACM SIGKDD Explorations Newsletter, Vol. 4, Issue 2, 2002, pp. 78-86.
[21] Mayr, E., Principles of Systematic Zoology, Second Edition, McGraw-Hill, New York, 1991.
[22] Mohanty, B. K. and B. Bhasker, "Product classification in the Internet business—a fuzzy approach," Decision Support Systems, Vol. 38, Issue 4, 2005, pp. 611-619.
[23] Moshkovich, H. M., A. I. Mechitov, and D. L. Olson, "Rule induction in data mining: effect of ordinal scales," Expert Systems with Applications, Vol. 22, Issue 4, 2002, pp. 303-311.
[24] Nauck, D., and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," Artificial Intelligence in Medicine, Vol. 16, Issue 2, 1998, pp. 149-169.
[25] Pawlak, Z., J. Grzymala-Busse, R. Slowinski, and W. Ziarko, "Rough Sets," Communication of the ACM, Vol. 38, Issue. 11, 1995, pp. 89-95.
[26] Shaw, M. J., C. Subramaniam, G. W. Tan, and M. E. Welge, "Knowledge management and data mining for marketing," Decision Support Systems, Vol. 31, Issue 1, 2001, pp. 127-137.
[27] Sheikh, K., Manufacturing Resource Planning (MRP II) with introduction to ERP, SCM, and CRM, International Edition, McGraw-Hill, Singapore, 2002.
[28] Sousa, I. and D. Wallace, "Product classification to support approximate life-cycle assessment of design concepts," Technological Forecasting and Social Change, Vol. 73, Issue 3, 2006, pp. 228-249.
[29] Swiniarski, R. W. and A. Skowron, "Rough set methods in feature selection and recognition," Pattern Recognition Letters, Vol. 24, Issue 6, 2003, pp. 833-849.
[30] Thabtah, F., P. Cowling, and S. Hammoud, "Improving rule sorting, predictive accuracy and training time in associative classification," Expert Systems with Applications, Vol. 31, Issue 2, 2006, pp. 414-426.
[31] Wakaki, T., H. Itakura, and M. Tamura, "Rough Set-Aided Feature Selection for

Automatic Web-Page Classification," Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence (WI 2004), 2004, pp. 70-76.

[32] Yuan, H., S. S. Tseng, W. Gangshan, and Z. Fuyan, "A Two-phase Feature Selection Method using both Filter and Wrapper," Proceeding of the IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC '99 Conference), Tokyo, Japan, 1999, pp. 132-136.

附錄　　系統執行步驟範例資訊

表 A-34：範例商品銷售週期資訊

CALNDR_NO \ ItemNo	24553	26718	27664	28713	30030	30433	30850	31996	32396	41954	45956
10312				180.78	11.43	64.09		24.82	49.78		
10313	2.14	27.53	2.68	512.2	96.45	306.42	9.01	91.84	142.1		
10314	0	215.23	5.35	580.74	153.6	289.73	12.72	50.89	152.95		
10315	0.43	0	17.4	396.2	167.18	220.38	14.31	91.84	187.34		
10316	4.27	1521.63	33.45	985.22	165.04	224.77	28.1	84.39	135.76	151.21	
10317	0.43	35.04	0	676.4	198.61	326.61	33.41	105.5	197.31	0	6.87
10318	8.11	25.03	0	316.36	128.6	44.78	34.46	49.64	141.19	8.49	0
10319	0	1280	0	533	148	215	16	81	182	20	2
10320	12	7	11	690	140	244	40	96	157	50	0
10321	8	19	0	677	193	238	37	97	157	100	0
10322	6	197	0	600	103	307	14	85	184	40	4
10323	0	111	11	320	223	189	30	99	159	0	14
10324	0	1301	8	301	133	190	14	72	378	0	16
10325	1	0	11	652	122	170	18	78	132	460	21
10326	12	80	0	183	114	256	9	78	107	20	0
10327	21	144	5	625	106	278	17	96	105	0	0
10328	26	1461	11	266	170	177	18	47	122	20	0
10329	0	0	11	959	91	179	50	53	128	2	0
10330	0	58	35	878	120	233	6	98	126	0	0
10331	0	111	22	352	80	236	30	81	136	20	0
10332	3	1339	0	614	104	148	14	83	120	0	0
10333	21	7	13	208.06	103.08	159.53	7	114.15	150.9	0	0
10334	0	7.71	26.56	1014.77	149	211.2	34.7	85.93	298.33	183.14	0
10335	9.57	259.74	26.56	670.13	92.54	314.64	9.47	93.63	201.56	22.89	0
10336	0	92	11	282	85	122	11	100	181	0	0
10337	0	1243	22	547	118	280	17	75	214	0	5
10338	13	0	0	725	92	293	51	108	139	0	0
10339	10	45	1	478	68	121	17	60	113	0	0
10340	12	125	1	634	91	277	7	49	116	20	0
10341	8	1216	11	568	60	72	17	92	118	20	0
10342	0	0	23	435	114	163	51	61	152	0	6
10343	0	0	0	657	66	177	11	73	137	40	0
10344	1	143	38	654	76	156	20	67	141	20	0
10345	12	1304	16	430	86	186	21	68	130	40	2

10346	2	3	29	505	179	271	26	47	164	0	0
10347	16.38	3.17	0	506.51	67.79	195.47	22.83	33.16	71.42	0	1.47
10348	6.12	63.47	20.81	260.71	107.04	217.43	20.81	250.35	137.78	276.83	7.33
10349	3.06	212.61	16.85	655.5	191.5	177.71	23.5	47.53	168.81	7.84	0
10350	0	1293	10	288	94	163	10	85	167	0	0
10351	0	8	0	618	144	259	25	73	195	22	10
10352	0	77	22	343	87	308	14	102	166	0	0
10353	0	96	13	177	68	175	15	52	230	27	10
10354	0	1420	14	1140	83	309	25	132	255	100	0
10355	4	0	12	290	156	219	20	58	220	0	4
10356	0	0	22	682	100	290	51	121	184	60	1
10357	0	123	11	642	126	260	12	73	184	20	5
10358	7	1272	0	243	71	114	35	43	224	44	4
10359	0	100	22	234	126	262	25	75	199	0	0
10360	0	0	12	1246	100	307	24	99	225	200	0
10361	3	41	30	805	154	254	15	70	216	60	3
10362	13	148	0	329	230	242	48	56	175	0	0
10363	1	1387	0	319	255	271	19	70	183	62	0
10364	0	0	4	771.1	133.29	280.97	18.53	95.96	232.49	42	0
10365	26.89	0	29.44	685.44	91.44	226.53	7.42	78.19	244.37	254.85	0
10366	0	136.39	29.44	802.19	70.73	264.27	49.3	139	208.17	8.49	0
10367	22.2	2634.07	29.44	277.94	103.6	209.84	71.58	76.95	177.39	4.25	5.15
10368	6.4	112.62	0	599.57	124.31	259.89	20.15	130.32	315.87	16.99	12.03
10369	0.86	0	2.68	141.61	59.3	32.49	17.49	64.54	158.39	0	0
10370	7.68	71.33	14.72	767.54	86.45	181.74	3.18	117.9	122.19	0	5.16
10371	2	189.8	0	738	98.29	289.36	19.53	74.21	150.63	20	0
10372	2	2768	151	808	131	256	10	78	122	0	0
10373	7	0	22	483	212	244	42	88	176	42	0
10374	0	135	33	261	115	361	21	120	236	20	6
10375	0	1003	0	387	144	222	26	81	155	20	0
10376	12	108	0	821	89	151	34	69	161	500	0
10377	3	80	0	632	130	210	21	72	181	0	7
10378	12	11	12	403	96	195	17	78	143	0	0
10379	0	146	33	517	137	271	24	96	157	0	4
10380	0	1390	0	580	129	175	23	98	194	27	5
10381	0	37	2	739	112	244	14	62	158	40	3
10382	4	0	0	932	229	169	18	97	253	0	1
10383	0	64	11	380	91	218	29	115	144	0	0

10384	2	1221.57	1	747.79	88.6	108.03	17.62	122.27	89.06	0	2
10385	0	37.02	0	361.23	135.06	202.36	40.22	75.63	125.65	0	8.58
10386	22.7	137.27	0	457.76	99.66	260.18	33.52	96.12	163.23	20	0
10387	0	36	0	422	184	359	24	90	177	20	0
10388	6	1293	13	725	62	162	11	103	166	0	0
10389	0	62	39	596	107	106	21	115	183	29	0
10390	0	151	0	385	87	199	17	138	209	320	0
10391	2	10	22	577	132	146	9	96	208	0	0
10392	14	123	33	955	78	238	11	69	220	31	0
10393	0	1441	20	696	83	268	19	50	144	0	0
10394	0	196	33	674	74	172	12	74	133	29	13
10395	6	0	11	503	129	236	34	91	156	0	2
10396	7	48	33	1456	87	168	26	97	222	26	6
10397	1	1190	0	58	116	171	45	99	173	20	0
10398	4	161	11	1030	63	189	10	81	122	20	2
10399	22	200	23	727	80	211	9	64	180	200	3
10400	0	35	0	688	273	115	36	88	238	9	0
10401	36	132.85	0	318.72	74.03	244.63	39.67	66.37	198.61	0	2
10402	2.62	1746.4	38.17	182.5	149.86	257.15	12.09	58.59	152.94	7.84	2.94

(資料來源：本研究整理)

表 A-35：商品月銷售記錄(經平滑化)

Year	**Month**	**24553**	**26718**	**27664**	**28713**	**30030**	**30433**	**30850**	**31996**	**32396**	**41954**	**45956**
2005	**1**	22.91	1572.7	44.13	2264.49	574.31	984.19	99.67	348.2	713.66	178.55	11.5
2005	**2**	22.81	1585.12	45.9	2288.23	567.98	980.36	99.58	350.73	717.31	179.9	11.38
2005	**3**	22.71	1597.55	47.68	2311.97	561.66	976.53	99.48	353.25	720.96	181.25	11.26
2005	**4**	22.61	1609.98	49.46	2335.72	555.33	972.7	99.39	355.78	724.61	182.61	11.14
2005	**5**	22.51	1622.41	51.24	2359.46	549	968.87	99.3	358.3	728.26	183.96	11.03
2005	**6**	22.41	1634.83	53.02	2383.2	542.68	965.04	99.21	360.83	731.91	185.31	10.91
2005	**7**	22.31	1647.26	54.79	2406.94	536.35	961.21	99.11	363.35	735.56	186.66	10.79
2005	**8**	22.21	1659.69	56.57	2430.69	530.02	957.38	99.02	365.88	739.21	188.01	10.67
2005	**9**	22.11	1672.12	58.35	2454.43	523.69	953.56	98.93	368.4	742.86	189.37	10.55
2005	**10**	22.01	1684.54	60.13	2478.17	517.37	949.73	98.83	370.93	746.51	190.72	10.43
2005	**11**	21.91	1696.97	61.9	2501.91	511.04	945.9	98.74	373.45	750.16	192.07	10.32
2005	**12**	21.81	1709.4	63.68	2525.66	504.71	942.07	98.65	375.98	753.82	193.42	10.2
2006	**1**	21.71	1721.82	65.46	2549.4	498.39	938.24	98.56	378.5	757.47	194.77	10.08
2006	**2**	21.61	1734.25	67.24	2573.14	492.06	934.41	98.46	381.03	761.12	196.13	9.96
2006	**3**	21.51	1746.68	69.01	2596.89	485.73	930.58	98.37	383.55	764.77	197.48	9.84

2006	**4**	21.41	1759.11	70.79	2620.63	479.4	926.75	98.28	386.08	768.42	198.83	9.72
2006	**5**	21.31	1771.53	72.57	2644.37	473.08	922.92	98.19	388.6	772.07	200.18	9.61
2006	**6**	21.21	1783.96	74.35	2668.11	466.75	919.09	98.09	391.13	775.72	201.53	9.49
2006	**7**	21.11	1796.39	76.12	2691.86	460.42	915.26	98	393.65	779.37	202.89	9.37
2006	**8**	21.01	1808.82	77.9	2715.6	454.1	911.43	97.91	396.18	783.02	204.24	9.25

(資料來源：本研究整理)

表 A-36：商品(24553)週銷售記錄與六種預測模式下之預測值

	ori	**100**	**112**	**104**	**200**	**212**	**204**
10313	2.14	4.77	5.43	4.79	6.18	9.1	5.07
10314	0	4.78	5.51	4.8	6.09	8.92	5.07
10315	0.43	4.79	5.58	4.82	6.01	8.75	5.06
10316	4.27	4.8	5.65	4.83	5.93	8.59	5.06
10317	0.43	4.81	5.72	4.85	5.86	8.43	5.06
10318	8.11	4.83	4.17	4.86	5.78	5.96	5.05
10319	0	4.84	4.22	4.88	5.71	5.85	5.05
10320	12	4.85	4.27	4.89	5.64	5.75	5.05
10321	8	4.86	4.32	4.91	5.57	5.65	5.05
10322	6	4.87	2.09	4.92	5.51	2.66	5.05
10323	0	4.89	2.12	4.94	5.44	2.62	5.05
10324	0	4.9	2.14	4.95	5.38	2.58	5.05
10325	1	4.91	2.17	4.97	5.32	2.54	5.05
10326	12	4.92	6.92	5.65	5.27	7.89	5.72
10327	21	4.93	7	5.66	5.21	7.78	5.73
10328	26	4.95	7.08	5.68	5.16	7.67	5.73
10329	0	4.96	7.16	5.7	5.11	7.57	5.73
10330	0	4.97	5.5	5.71	5.06	5.69	5.73
10331	0	4.98	5.56	5.73	5.01	5.63	5.74
10332	3	5	5.62	5.75	4.97	5.57	5.74
10333	21	5.01	5.68	5.76	4.93	5.51	5.75
10334	0	5.02	5.74	5.78	4.89	5.46	5.75
10335	9.57	5.03	3.1	5.8	4.85	2.89	5.76
10336	0	5.04	3.13	5.81	4.81	2.86	5.76
10337	0	5.06	3.16	5.83	4.78	2.84	5.77
10338	13	5.07	3.19	5.85	4.75	2.83	5.78
10339	10	5.08	5.97	5.38	4.72	5.2	5.31
10340	12	5.09	6.03	5.4	4.69	5.18	5.31
10341	8	5.1	6.09	5.41	4.67	5.16	5.32

10342	0	5.12	6.15	5.43	4.65	5.14	5.33
10343	0	5.13	7.88	5.44	4.63	6.52	5.34
10344	1	5.14	7.95	5.46	4.61	6.51	5.35
10345	12	5.15	8.03	5.48	4.59	6.51	5.36
10346	2	5.17	8.1	5.49	4.58	6.52	5.37
10347	16.38	5.18	8.18	5.51	4.57	6.53	5.38
10348	6.12	5.19	2.26	5.52	4.56	1.79	5.39
10349	3.06	5.2	2.28	5.54	4.55	1.79	5.4
10350	0	5.21	2.3	5.55	4.54	1.8	5.41
10351	0	5.23	2.32	5.57	4.54	1.81	5.43
10352	0	5.24	1.02	2.28	4.54	0.79	2.22
10353	0	5.25	1.03	2.29	4.54	0.8	2.23
10354	0	5.26	1.04	2.29	4.54	0.81	2.23
10355	4	5.27	1.05	2.3	4.55	0.81	2.24
10356	0	5.29	1.06	2.31	4.56	0.82	2.24
10357	0	5.3	1.88	2.31	4.57	1.46	2.25
10358	7	5.31	1.89	2.32	4.58	1.48	2.26
10359	0	5.32	1.91	2.33	4.59	1.49	2.26
10360	0	5.33	1.92	2.33	4.61	1.51	2.27
10361	3	5.35	4.74	2.34	4.63	3.74	2.28
10362	13	5.36	4.77	2.34	4.65	3.79	2.28
10363	1	5.37	4.81	2.35	4.67	3.84	2.29
10364	0	5.38	4.85	2.36	4.7	3.9	2.3
10365	26.89	5.4	9.14	5.56	4.72	7.4	5.43
10366	0	5.41	9.21	5.58	4.75	7.52	5.45
10367	22.2	5.42	9.28	5.59	4.78	7.64	5.47
10368	6.4	5.43	9.35	5.61	4.82	7.77	5.49
10369	0.86	5.44	9.43	5.62	4.85	7.9	5.51
10370	7.68	5.46	6.84	5.64	4.89	5.78	5.53
10371	2	5.47	6.89	5.65	4.93	5.89	5.55
10372	2	5.48	6.94	5.67	4.97	6	5.57
10373	7	5.49	6.99	5.68	5.02	6.11	5.59
10374	0	5.5	3.37	5.7	5.07	2.98	5.61
10375	0	5.52	3.4	5.71	5.11	3.04	5.63
10376	12	5.53	3.42	5.73	5.17	3.1	5.66
10377	3	5.54	3.45	5.74	5.22	3.16	5.68
10378	12	5.55	10.96	6.53	5.27	10.17	6.46
10379	0	5.57	11.03	6.54	5.33	10.38	6.49

10380	0	5.58	11.11	6.56	5.39	10.59	6.52
10381	0	5.59	11.19	6.58	5.45	10.81	6.55
10382	4	5.6	8.58	6.6	5.52	8.4	6.58
10383	0	5.61	8.63	6.61	5.58	8.57	6.61
10384	2	5.63	8.69	6.63	5.65	8.75	6.64
10385	0	5.64	8.75	6.65	5.72	8.94	6.67
10386	22.7	5.65	8.81	6.66	5.79	9.13	6.7
10387	0	5.66	4.74	6.68	5.87	4.98	6.73
10388	6	5.67	4.77	6.7	5.95	5.09	6.76
10389	0	5.69	4.8	6.71	6.02	5.2	6.79
10390	0	5.7	4.83	6.73	6.11	5.31	6.82
10391	2	5.71	9.01	6.19	6.19	10.03	6.29
10392	14	5.72	9.07	6.21	6.27	10.25	6.32
10393	0	5.73	9.12	6.22	6.36	10.47	6.35
10394	0	5.75	9.18	6.24	6.45	10.69	6.38
10395	6	5.76	9.24	6.25	6.54	10.92	6.42
10396	7	5.77	11.81	6.27	6.64	14.17	6.45
10397	1	5.78	11.89	6.28	6.74	14.47	6.48
10398	4	5.8	11.96	6.3	6.83	14.78	6.52
10399	22	5.81	12.03	6.32	6.93	15.09	6.55
10400	0	5.82	3.31	6.33	7.04	4.22	6.59
10401	36	5.83	3.33	6.35	7.14	4.31	6.62
10402	2.62	5.84	3.35	6.36	7.25	4.4	6.66

(資料來源：本研究整理)

表 A-37：範例商品經過前置作業所得指標數值

ItemNo	Mode	b_0	b_1	b_2	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}
24553	100	4.770163	0.012121													
24553	200	6.182334	-0.08416	0.001081												
24553	112	3.376178	0.044248		1.611	1.1597	0.5561	1.7532	1.3346	0.7134	1.3201	1.6764	0.4592	0.201	0.3533	0.8621
24553	212	5.652077	-0.11093	0.001743	1.611	1.1597	0.5561	1.7532	1.3346	0.7134	1.3201	1.6764	0.4592	0.201	0.3533	0.8621
24553	104	4.310024	0.013465		1.1117	1.2598	1.1559	0.4727								
24553	204	4.564755	-0.0039	0.000195	1.1117	1.2598	1.1559	0.4727								
26718	100	342.5629	1.223175													
26718	200	339.7786	1.413017	-0.00213												
26718	112	320.2249	1.648688		1.4278	1.31	0.8947	0.9863	0.8928	0.932	0.9253	0.8913	0.9555	0.958	0.8924	0.934
26718	212	317.6216	1.826185	-0.00199	1.4278	1.31	0.8947	0.9863	0.8928	0.932	0.9253	0.8913	0.9555	0.958	0.8924	0.934
26718	104	316.6429	1.739975		1.2232	0.9392	0.9113	0.9263								

26718	204	313.7306	1.938539	-0.00223	1.2232	0.9392	0.9113	0.9263								
27664	100	10.36025	0.08143													
27664	200	6.913498	0.316435	-0.00264												
27664	112	10.57188	0.070412		1.3966	1.5541	0.5676	0.6201	1.0508	0.9344	1.1381	1.193	0.8374	1.4153	0.7452	0.5473
27664	212	7.300421	0.293466	-0.00251	1.3966	1.5541	0.5676	0.6201	1.0508	0.9344	1.1381	1.193	0.8374	1.4153	0.7452	0.5473
27664	104	10.44225	0.0755		1.1848	0.88	1.0164	0.9189								
27664	204	6.970564	0.312205	-0.00266	1.1848	0.88	1.0164	0.9189								
28713	100	515.0771	1.049804													
28713	200	528.2603	0.161051	0.009875												
28713	112	505.1391	1.179946		1.2057	1.0436	0.8221	0.8757	1.2127	0.8788	1.0884	1.1865	0.7539	1.0785	0.9598	0.8942
28713	212	494.7263	1.881932	-0.0078	1.2057	1.0436	0.8221	0.8757	1.2127	0.8788	1.0884	1.1865	0.7539	1.0785	0.9598	0.8942
28713	104	503.1975	1.35077		1.0287	1.0203	0.9398	1.0112								
28713	204	523.6518	-0.02817	0.015321	1.0287	1.0203	0.9398	1.0112								
30030	100	123.2807	-0.10404													
30030	200	131.0757	-0.62955	0.005838												
30030	112	119.459	0.008382		1.1148	1.0579	0.9951	0.9145	1.1698	0.8168	0.8372	0.8172	1.0102	0.9415	0.8162	1.5088
30030	212	129.7444	-0.68501	0.007704	1.1148	1.0579	0.9951	0.9145	1.1698	0.8168	0.8372	0.8172	1.0102	0.9415	0.8162	1.5088
30030	104	117.4455	0.063511		1.0761	0.976	0.8489	1.0989								
30030	204	130.2602	-0.8004	0.009599	1.0761	0.976	0.8489	1.0989								
30433	100	219.5367	-0.08318													
30433	200	205.3249	0.874914	-0.01065												
30433	112	212.718	0.156906		1.2072	0.8855	0.9325	0.922	1.0143	0.9591	0.8958	0.9074	0.8451	1.3509	0.9832	1.0971
30433	212	207.7676	0.490639	-0.00371	1.2072	0.8855	0.9325	0.922	1.0143	0.9591	0.8958	0.9074	0.8451	1.3509	0.9832	1.0971
30433	104	214.5315	0.133232		1.0237	0.9735	0.8489	1.1539								
30433	204	216.0229	0.032689	0.001117	1.0237	0.9735	0.8489	1.1539								
30850	100	22.10591	0.018498													
30850	200	20.56832	0.123334	-0.00118												
30850	112	21.31966	0.042691		1.328	1.0169	0.8983	0.8676	1.1635	0.816	0.8956	0.9664	0.7997	1.2616	0.9698	1.0166
30850	212	20.96812	0.06666	-0.00027	1.328	1.0169	0.8983	0.8676	1.1635	0.816	0.8956	0.9664	0.7997	1.2616	0.9698	1.0166
30850	104	21.41407	0.038565		1.0845	0.9332	0.9179	1.0644								
30850	204	20.52726	0.099029	-0.00068	1.0845	0.9332	0.9179	1.0644								
31996	100	78.70016	0.126985													
31996	200	73.0151	0.510247	-0.00426												
31996	112	75.76274	0.18965		1.2531	0.9296	0.9168	0.8161	1.2935	1.09	0.8417	0.8226	1.2351	1.2512	0.7751	0.7751
31996	212	69.80464	0.591321	-0.00446	1.2531	0.9296	0.9168	0.8161	1.2935	1.09	0.8417	0.8226	1.2351	1.2512	0.7751	0.7751
31996	104	76.08038	0.167327		1.0539	1.075	0.9323	0.9388								
31996	204	68.91641	0.650291	-0.00537	1.0539	1.075	0.9323	0.9388								
32396	100	157.2013	0.31768													

32396	200	142.6639	1.297727	-0.01089												
32396	112	152.4772	0.518729		1.2741	0.8027	1.0493	0.7281	1.051	0.9566	0.8736	0.8626	0.8819	1.3849	1.0868	1.0483
32396	212	152.701	0.503642	0.000167	1.2741	0.8027	1.0493	0.7281	1.051	0.9566	0.8736	0.8626	0.8819	1.3849	1.0868	1.0483
32396	104	149.5758	0.607991		1.0622	0.9285	0.8159	1.1934								
32396	204	152.5911	0.404709	0.002258	1.0622	0.9285	0.8159	1.1934								
41954	100	48.05615	-0.08983													
41954	200	50.61698	-0.2706	0.002101												
41954	112	45.157	0.013469		1.1157	0.6324	2.6554	0.274	0.5822	0.9448	0.2466	0.8867	1.5657	0.9479	1.3289	0.8197
41954	212	56.19443	-0.76564	0.009059	1.1157	0.6324	2.6554	0.274	0.5822	0.9448	0.2466	0.8867	1.5657	0.9479	1.3289	0.8197
41954	104	38.73173	0.131825		1.5099	0.6299	0.7861	1.0741								
41954	204	43.35349	-0.19442	0.003793	1.5099	0.6299	0.7861	1.0741								
45956	100	3.096255	-0.01471													
45956	200	4.064142	-0.08384	0.000813												
45956	112	2.072662	0.008448		1.1386	0.3431	3.0688	0.6103	0.5963	0.2267	1.0668	0.7489	1.6245	1.422	0.863	0.291
45956	212	1.921461	0.019249	-0.00013	1.1386	0.3431	3.0688	0.6103	0.5963	0.2267	1.0668	0.7489	1.6245	1.422	0.863	0.291
45956	104	2.135284	0.005856		1.5644	0.6118	0.8611	0.9627								
45956	204	2.596046	-0.02706	0.000387	1.5644	0.6118	0.8611	0.9627								

(資料來源：本研究整理)

表 A-38：範例商品經篩選保留最適描述銷售發展趨勢之預測模式與對應預測誤差

ItemNo	**Mode**	b_0	b_1	s_1	s_2	s_3	s_4	**MAPE**
24553	104	4.310024	0.013465	1.1117	1.2598	1.1559	0.4727	1.219607
26718	104	316.6429	1.739975	1.2232	0.9392	0.9113	0.9263	8.064906
27664	104	10.44225	0.0755	1.1848	0.88	1.0164	0.9189	1.143057
28713	104	503.1975	1.35077	1.0287	1.0203	0.9398	1.0112	0.58986
30030	104	117.4455	0.063511	1.0761	0.976	0.8489	1.0989	0.406864
30433	104	214.5315	0.133232	1.0237	0.9735	0.8489	1.1539	0.366315
30850	104	21.41407	0.038565	1.0845	0.9332	0.9179	1.0644	0.62121
31996	104	76.08038	0.167327	1.0539	1.075	0.9323	0.9388	0.257357
32396	104	149.5758	0.607991	1.0622	0.9285	0.8159	1.1934	0.220455
41954	104	38.73173	0.131825	1.5099	0.6299	0.7861	1.0741	1.353416
45956	104	2.135284	0.005856	1.5644	0.6118	0.8611	0.9627	0.79535

(資料來源：本研究整理)

表 A-39：範例商品指標數值，經標準化後引入最適分類演算法

ItemNo	**b1_norm**	**s1**	**s2**	**s3**	**s4**
24553	0.034222	1.1117	1.2598	1.1559	0.4727
26718	4.422286	1.2232	0.9392	0.9113	0.9263

27664	0.191889	1.1848	0.88	1.0164	0.9189
28713	3.43309	1.0287	1.0203	0.9398	1.0112
30030	0.161418	1.0761	0.976	0.8489	1.0989
30433	0.33862	1.0237	0.9735	0.8489	1.1539
30850	0.098016	1.0845	0.9332	0.9179	1.0644
31996	0.425275	1.0539	1.075	0.9323	0.9388
32396	1.545258	1.0622	0.9285	0.8159	1.1934
41954	0.335044	1.5099	0.6299	0.7861	1.0741
45956	0.014883	1.5644	0.6118	0.8611	0.9627

(資料來源：本研究整理)

表 A-40：範例商品經搜尋後建議分類結果

ClassID	ItemNo
99104P0001	28713
	26718
99104P0002	27664
	24553
	30030
	30433
	30850
	31996
	32396
	41954
	45956

(資料來源：本研究整理)

簡歷

姓　　名：黃聖祐

出生地：台灣省彰化縣

出生日：中華民國七十四年六月二十一日

學　　歷：九十二年九月至九十六年六月

國立台灣大學資訊管理學系

地　　址：台灣省彰化縣彰化市彰益街９１之１號８樓