

# 以範例學習法預測台灣股市行爲 \*

楊建民 \*\* 林震岩 \*\*\* 劉立倫 \*\*\*\*

## 提 要

本研究係以人工智慧（Artificial Intelligence）機器學習之技術，探討台灣股票市場行爲，並預測證券交易所總體指數之漲跌。有關股市預測系統之知識庫，係採用範例學習（Learning From Examples）的技術建構。本研究以民國 73 至 75 年每個營業日的證交資料作為例子，透過範例學習系統 dBID III，導出預測台灣股票市場證交指數之漲跌的法則。本研究總計使用了成交量，證交指數變動百分比等相關的 33 個線索，預測證交指數漲跌之命中率達 59.9%；而技術分析指標中，以每日成交筆數之線索的預測能力最強。研究發現台灣股票市場，投資者多傾向短期操作，本系統所用以預測隔日股市證交指數漲跌之效果最佳。本研究亦發現股市預測系統的命中率，會因學習例子數的減少或隨著時間間隔的久遠而降低。因此，系統在學習初始，須有足夠的例子以涵括各種股市行爲之態樣，方能使所歸納學習出的法則具有普遍性；而且每經過一段時間，應加入一些新的例子，供系統學習以修正既有之法則，使系統能不斷適應市場結構的變遷。此學習能力亦為本系統之所以異於傳統系統而最具價值的部份。

\* 本研究接受行政院國科會專案研究補助 (NSC 79-0301-H004-23) 特此誌謝

\*\* 作者為本校資訊管理系專任副教授兼電子計算機中心主任

\*\*\* 作者為私立中原大學企管系副教授

\*\*\*\* 作者為國防管理學院會計系副教授

## 壹、導論

由於國內外政、經環境的迅速變遷，資本市場受到各種因素的影響日趨複雜。如何建全資本市場以有效地導引資本投入經濟建設；以及如何確立投資者的正確理財觀念，以改善其本身之財務狀況，在在都需要專家之專業諮詢與輔導。然而，由於專家的稀少性和昂貴性，使一般人不易取得專業的諮詢與輔導。因此，若能將專家之專業技能（Expertise）存留在電腦中，讓社會大眾充份利用，實在不失為解決上述問題最有效的途徑之一。

近年來，由於人工智慧（Artificial Intelligence，簡稱AI）技術之不斷突破，已經成功地運用到許多領域之中。其最主要的一個分支——「專家系統」（Expert Systems，簡稱ES）應用在醫療診斷，地質探勘，化學分子結構辨識等方面，業已有了很好的績效。由於這些弱結構化（ill-structure）問題的解決，須仰賴各該領域之專業知識（domain-specific knowledge）以及專家之經驗判斷。因此，專家系統所採之擬人化推論方法，便提供了解決這類問題一種較佳之途徑。由於部份企業管理之決策問題——如企業診斷、會計審計、投資評估、租稅規劃、人才選訓等——具有類似上述問題之非結構性特徵，故成為近年來專家系統應用發展最令人振奮的領域之一。

典型之專家系統主要係由推論策動機（Inference Engine）及一個巨大的專業知識庫（Knowledge Base）所構成。目前，對於前者已有很多專家系統之發展工具（或稱殼架shell）可資利用，在應用系統的發展上不會有太大的困難；然而對於後者，在知識庫建構所需專業知識的獲取（Knowledge Acquisition），往往曠日費時，須要許多人費上好幾年的時間才能完成，成為專家系統在推廣應用上難以突破之瓶頸。因此，若專家系統能有效地獲取其所需之推論知識，以建立其專業知識庫，並能隨時更新、增益其所具備之知識，則必能加速專家系統在各個領域之應用。

機器學習（Machine Learning）技術的發展，為專家系統知識之建構，提供了另一嶄新的途徑。傳統專家系統之建立，須仰賴人類專家提供推論法則以建立知識庫。然而現實社會中，有些問題是新近發生的，專家們並沒有這方面的經驗，來歸納經驗法則；有些問題是因時而變遷的，專家們也許尚未調適過來，而所用的法則或已不合時宜。而機器學習能夠直

接由已發生之案例歸納出法則來；同時所發展出之系統，能夠因應知識之變遷而演進，不致因退化、落伍而至廢棄。在機器學習領域中，目前又以範例學習——又稱由例子中學習（Learning From Examples，簡稱 LFE）這一支發展的較為完整。然而值得注意的是，採用範例學習方法發展專家系統，並非完全排除專家的協助，因為在例子（Examples）和線索（Cues）的選取以及法則的解釋等方面，仍需仰賴人類專家的經驗、專業知識和判斷的協助，而不能完全由系統所取而代之的。

在機器學習領域，有關從實例中來歸納法則的技術，最早始於Winston [1975] 所設計之教導積木堆建的系統。後來 Lenat [1976] 又設計出提供基本數學證明的 AM ( Automated Mathematician ) 系統。而 Buchanan et. al. [1976, 1978] 所發展出來之 Meta-DENDRAL 系統，則是專門用於分析大量光譜資料以辨識化學分子之結構。這些研究有助於以範例學習的方式來建立專家系統的知識庫。

Quinlan [1979] 提出 ID3 (Iterative Dichotomizer 3) 之演算法來解西洋棋的棋局（Chess Endgames），引起了學界對範例學習方式廣泛的重視。ID3 基本上是利用分類樹（Classification Tree）的方法，依據事物的屬性（或線索）分出類型，以便從例子中學習經驗法則。後來 Quinlan [1987] 將 ID3 進一步延伸，加入修剪（Pruning）的技巧，演進為 C4 演算法。Carter and Catlett [1987] 採用 ID3 與 C4 兩種演算法來研究信用卡評等，結果發現 ID3 之正確率高於現存的評等方法，而 C4 之正確率又高於 ID3。可見這種由例子中獲取專家般的知識和經驗，以建立專家系統的方法，是十分值得開發的領域。

Michalski and Chilausky [1980] 以伊利諾州某一農場之大豆疾病為研究對象，搜集了 290 個病例來歸納法則，並用 340 個病例來測試；同時以範例學習及專家告知（Learning From Being Told 簡稱 LFBT）兩種方法進行，結果發現 LFE 比 LFBT 的正確率更高，且高達 97.6%，他們使用的軟體是 AQ11。

至於使用歸納法須有多大的樣本，其法則之正確率才可接受，亦有相關之研究。Quinlan [1979] 及 Shapiro and Niblett [1982] 均曾以小樣本來歸納法則（都是西洋棋局的法則），結果相當成功。Quinlan 以佔母體 0.07% 的樣本來歸納法則，結果其正確率高達 99.67%；當樣本擴充至母體之 0.36% 後，其法則正確率更提高至 99.92%。此結果為 Shapi-

ro and Niblett [1982] 更進一步確認，他們使用 ACLS (Analog Concept Learning System) 自母體0.2%的樣本中歸納出法則，其正確率高達99%。O'Rorke [1982] 曾比較 AQ11 及 ACLS 之歸納邏輯，發現二者並無明顯優劣之分。

將範例學習應用在股票市場上的研究，則是最近幾年才出現的。Braun and Chandler [1987] 以 ACLS 系統研究股價指數的預測。他們訪談了一位在過去幾年績效相當不錯的投資專家。從其經驗中，選出 20 個影響股市行情的線索，並搜集了兩年紐約股市的實例，以 LFE 之方法歸納出法則。研究發現，此法則預測真實市場變動情形比預測人類專家的看法還要準確（命中率為 64.4% 比 51.5%），可見人類專家之預測的穩定性較差。Braun and Chandler 之研究，只採納了一位專家的意見，同時直接採用 Paterson, Blake and Shapiro 等人在 1982 年所撰寫之 ACLS 程式，並未針對股票市場之特性，發展一套所需之演算法，因此值得進一步的探討。

國內相關之研究，目前尚處於起步的階段，文獻不多。劉仲偉 [民國 76 年] 由專家之訪談建立了有關市場、行業的知識庫與個別股分析之知識庫以研究台灣之股票市場。此研究係以 LFBT 之方法建立知識庫。但卻未做驗證 (Validation) 故而無法得知其預測效果。至於以 LFE 之方法研究股票市場行為，據筆者所知目前國內還付諸闕如。本研究係首次將此範例學習之技術，用以研究台灣之股市行為。

## 貳、效率市場假說與股價預測

Fama [1970, 1976] 依市場的效率性將資本市場分為弱式 (weak form)、半強式 (semistrong form) 及強式 (strong form) 三種型態。在弱式效率資本市場下，目前的證券價格，已充份反應了過去的股價或報酬率等所提供的資訊，因此投資者無法藉著分析證券價格來獲取超額利潤。在半強式效率資本市場中，目前證券的價格係反映所有公開可得的 (publicly available) 資訊；也就是說，當新的資訊會影響資產價格時，市場便會迅速的反映此一資訊的經濟意義，因此，投資人亦無法藉著分析這些公開可得的資訊而獲得超額報酬。而在強式效率市場中，投資人不僅無法藉助分析公開可得資訊提昇其績效，亦無法藉助分析內部資訊 (inside information) 以獲致超額報酬。因此，如果效率資本市場成立，則

投資人的各種分析方法均無法使其獲致超額報酬。

有關弱勢效率資本市場是否存在的研究，曾有學者進行不同的研究，並獲致不同的結論，如 Fama & Blume [ 1966 ] 採濾嘴法則進行實證，發現弱勢效率市場存在；Seelenfreund, Parker & Van Horne [ 1968 ] 以移動平均法進行實證，亦發現弱勢效率市場成立；但 Jensen [ 1967 ] 以相對強弱指標來進行實證，卻發現弱勢效率市場值得懷疑。國內的研究亦出現類似的結果：林煜宗 [ 民國 67 年 ] 以序列相關來檢定股票市場，認為弱勢效率市場成立；簡仁德 [ 民國 70 年 ] 驗證股票市場隨機漫步假定，亦發現弱勢效率市場成立；但徐世豪 [ 民國 68 年 ] 以濾嘴法則進行股票市場實證，卻發現超額報酬的存在。

在半強式效率市場上的研究，向為財務學者所關切，故相關的實證亦多。例如 Fama et. al. [ 1969 ] 探討股票分割對股價的影響，認為半強式效率市場成立；Ibbotson [ 1975 ] 研究新上市股票的股價，發現新股在上市一日內的確有超額報酬存在；Randleman, Jones & Latane [ 1982 ] 研究上市公司每股盈餘宣告對股價的影響，也肯定了半強式效率市場的存在；而 Rogers [ 1988 ] 研究益本比和規模效應對股價的影響之後，則提出相反的結論。國內有關的研究，如李廣進 [ 民國 69 年 ] 研究新上市股票是否有超額報酬，發現半強式效率市場不存在；金柏西 [ 民國 69 年 ] 在探討股票股利對股價的影響後，發現半強式效率市場無法成立；王官品 [ 民國 75 年 ] 研究上市公司每月盈收公告對股價的影響，則認為半強式效率市場成立；林芸萱 [ 民國 76 年 ] 在研究油品價格調整對股價的影響之後，亦獲致類似的結論。

至於在強式市場的研究則較為少見，多針對相互基金、專業證券商及公司內部人員是否能夠獲致超額利潤而進行研究。如 Niederhoffer & Osborne [ 1966 ] 曾對紐約證券交易所的專業證券商 ( specialists ) 進行研究，發現他們確可獲得超額報酬；Jaffe [ 1974 ] 針對上市公司內部人員進行研究，亦發現他們可獲得超額報酬。

由以上的評論不難發現，效率資本市場是否存在，仍難有一致的定論。雖然許多財務學者認為技術分析不可信，但在華爾街卻仍充斥許多技術分析的專家 ( Brady [ 1984 ] )，亦有藉著技術分析而獲致鉅額報酬的專業投資人 ( Liscio [ 1987 ] )。這些現象無異是說明，仍有許多人認為技術分析可以擊敗弱式的效率市場，而在國內股票市場亦可發現類似的現

象。近年來，在投資人口快速成長的同時，技術分析及其操作策略亦逐漸影響投資人的投資行為，並影響台灣股票市場的動向。故本研究嘗試以範例學習法，採用技術面之指標為主要線索，導出台灣股市行為之型態，據以預測台灣證券交易所總指數之漲跌，並以實例驗證預測之命中率。

## 參、股市預測範例學習系統

Quinlan [ 1979 ] 所提出之 ID3 學習演算法，係將學習視為一種分類法則（Classification Rules）的建構，他以決策樹之方式來發展分類法則，每個節點（Node）以該領域的屬性（如：顏色、形狀等）來分枝。此系統之優點在於他發現了一種建構最小樹（a minimum tree）的方法，可使系統之運作變得非常經濟。

本研究所使用之範例學習系統，係以 Quinlan [ 1979 ] 所發展之 ID3 為基礎，依股市行為之特性改進而成之 dBiD III 系統。本系統在 Aston Tate 公司之 dBASE III PLUS [ 1986 ] 之版本上開發，並經由 Clipper 編譯器之編譯，以增強程式執行之速度。本系統可在 IBM PC / AT 或其他相容的機型上執行。並可在中文的環境下執行（係採用倚天中文系統）。同時系統採用交談式（Interactive）和功能表導向（menu – driven）的方式設計，可提昇使用者之親和度。本研究範例學習系統可以分為下列五大部份：

1. 例子集（Examples）：又稱訓練集（Training Set）或例子資料庫（Example DataBase，簡稱 EXDB），亦即用來歸納法則的案例資料，收集的例子應儘可能的完備，以涵括問題領域中各種的態樣。
2. 線索（Cues）：或稱屬性（attributes），用來描述例子之特徵，一個EXDB中可採用許多線索來建立分類樹，這些線索可以是描述性的分類（Catagorial）變數，或稱計質的（qualitative）變數；亦可以為計量的（quantitative）變數，例如整數、實數等。
3. 目標（Goals）：或稱類別（Classes），表示專家所做之決策或事件在不同狀況下發生之結果（Outcomes）。亦即目標係各線索在不同值組合之下的結果，而每個例子應只有一種結果。
4. 演算法（Algorithm）：即用以導出分類樹並將例子歸納出法則之程式。

5. 法則 (Rules)：由演算法導出之分類樹的每一完整分枝 (Branch) —— 或稱路徑 (Path) —— 即為一條法則。

本演算法，最主要之功能，在於分類樹的導出。由根點 (Root) 出發，對每一節點選取一個「最佳」之線索，並找出「最佳」之分割點 (Thresholds)，而將節點擴張出數個後繼節點 (successors)，如此反覆進行，一直到後繼節點皆滿足停止準則 (Termination Rule) 為止。

選取「最佳」之線索和分割點之目的，在於優先採用較具區別能力線索，期能縮減分類樹之層級以儘速達到分類的目的。而選取標準則係根據 Shannon [1975] 所發展出來之資訊理論 (Information Theory) 中之熵 (Entropy) 來衡量，熵亦稱為樹的資訊內容 (Information Content of Tree，簡稱 ICT)。熵代表的是經由該線索分類後，該節點中例子集之亂度或不確定性。熵愈高表示該節點的亂度愈高，分類後的節點之混雜程度愈高，亦即分類的效果愈差。因此「最佳」線索和分歧點的選取，目的在於使分類後，熵能降低的最多；亦即使後繼節點之純度愈高，而混雜的資訊內容愈少。

停止分割準則之設定，往往取決於分類精確度與成本效益二者間之取捨 (tradeoff)。理想狀況下，係希望分類樹節點之熵能很快地降至零，亦即每個葉子 (leaf) 都達到百分之百的純度，如此自然就可停止樹的繼續分割。然而對於複雜的現實社會之分類問題，往往所有的線索皆已用完，但卻還無法使熵降到一個滿意的水準（此水準可以預先設定）。此時，唯一的方法，便是設法再發掘新的線索，使得樹能繼續成長。至於新線索的發掘，目前仍需仰賴領域專家的協助。在實用上，有時候熵雖可以繼續降低，但下降的幅度甚緩，幾乎可以忽略；此時由成本與效益比的觀點考慮，此分類樹已不值得再繼續分割，我們亦可停止該演算法之執行。

由於股票市場行爲預測需要用到很多計量性的線索（如成交量、證交指數等），這些線索的可能值，在某一區間內為連續性之數值。因在搜尋最佳分割點，所需的時間將隨著分隔的密度，呈快速的成長，產生所謂的組合爆炸 (Combinatorial Explosion) 的現象。例如若某一線索有一百種可能值，欲分成五個區隔，亦即須在此 100 個值中，找出四個最佳分割點，去掉兩個端點，則仍有  $C(98,4) = 3,612,280$  種可能組合。而對每一個組合，必須計算

一次熵的值來比較；此外，對每一種線索在分類樹每一階層，皆須作一次同樣的運算。隨著應用系統複雜度的增加，演算法將很快的無法在可容忍的時間內完成分類。

因此，對於計量性線索可能的分割點的選取，我們採用了下列三種方法：

方法一：以資料分佈的全距（Range）分割成等距離的 10 群，並取得相對應的點。

方法二：以資料分佈的頻率（Frequency）分割成例子數相等的正負各 10 群。

方法三：以資料分佈的頻率，分割成例子數相等的正（或負）的 10 群，而負（或正）的值，由正（或負）的值對應而來。

初步的研究顯示，方法二的效果較佳（參見第五節表 5-1）。主要是因為當極端值發生較大的偏離現象時，方法一將無法有效地分隔例子群；亦即絕大多數的例子，可能落在少數的組距之內。而股市預測的線索中有許多變動百分比的變數，當分母相當小時，便很容易出現極端值。因此，後續的研究中，均以方法二來產生計量線索的分割點。

由於影響股市的因素錯綜複雜，幾乎不可能掌握所有影響股價的因素，因此無法將股市行爲之態樣完全區別開來，亦即無法將分類樹所有的葉子都達到 100% 純度的要求。因此本研究將演算法的停止分割準則，預先設定於 70% 純度。有些葉子在結束時，早已超過了此純度；也有些葉子用盡了所有的線索，仍達不到 70% 的純度。

根據 Breiman, Friedman, Olshen & Stone [1984] 等人的研究，採用修剪（Pruning）技巧，能得一個較佳的分類樹，本研究中係以葉子中例子的個數，作為分類樹修剪的準則。亦即當分枝後所有的後繼節點的例子數皆小於 5 時，則不論其個別的純度為何，皆將此分枝修剪回原節點。修剪策略的使用，可避免無用的分枝產生；也可避免因過早停止分割而遺漏了有用的分枝。

關於系統之驗證，我們以股市實際發生之例子來測試，並計算命中率（hit rate），即測試能夠正確預測的例子數所佔的比率。此外我們還計算強態率（strong pattern rate），即純度超過 70% 的葉子數佔總葉子數的比例。預測時，當例子落入此種強態的葉子中時，該次預測的命中之機率會高於平均水準。因此，實際使用本系統預測股市證交指數時，可以採用選擇性預測之策略；亦即在預測時，只有在測試例子落入強態葉子中時，才採用該預測結果，如此可以大幅地提高預測的命中率。

## 肆、研究設計與實例說明

在本節中，我們將分別探討，以範例學習法預測證交所總指數之漲跌時有關線索之選取，例子資料庫的範圍與建立，以及目標之設定等。

### 一、線索之選取

在技術分析中，可以作為股價預測的線索有許多，較常見的有  $k$  線，移動平均線，人氣指標——成交量值，趨勢線——抵抗線，圖形研判等。基於「範例學習」特性的考慮，我們修正技術分析中，部份無法與本研究採取的研究程序配合之線索；同時，也發展出一些新的線索，以增強系統的預測能力。

由於範例學習演算法，能夠優先選取出較具區別能力之線索（Quinlan [ 1979 ]），因此線索之選取，僅需要專家儘量提供與指數變動有關的線索即可，而將線索的取捨交由系統來決定。本研究在第一階段中訪問了專業投資機構（如鴻運基金）經理人、證券投資專家等。經過了多次討論後獲得了「每日成交筆數」等 43 個線索（見表 4-1）。其中線索 1 至 13 和線索 50（重要事件）係以實際歷史資料輸入例子資料庫；而線索 14 至線索 42 的資料，則是依據前述線索的資料加以計算得來。

經過幾次初步的測試，系統發現第一階段 43 個線索中有每日成交總值等 22 個線索被優先選取的頻率甚低；亦即在演算法建立分類樹的過程中該線索的區別能力較差。因此決定刪除這些線索（表 4-1 中有「 $\times$ 」記號的 22 個線索）。所保留的線索中，有每日成交筆數，每日成交量，證交所每日股價指數，加上這三個線索之變動百分比和 6 日變動百分比，以及證交所股價指數 13 日與 25 日變動百分比。其中成交筆數用以反映市場的人氣；而成交量的變化，則顯示資金的進出。這種供需的消長，實是股價變動的原動力；證交所股價指數之變動，則可顯現出整個市場漲跌的趨勢和脈動，是很多指標的基礎。

每日漲跌家數和其 6 日變動百分比亦予以保留，此亦為影響大盤漲跌的氣勢。另外融資餘額、融券餘額、財氣指標和 OBV 人氣指標雖未予保留，但其變動百分比則予保留。此乃因我們主要在預測指數的漲跌，故趨勢的變動比率值，反而較其絕對值重要。

表 4-1 線索總表

*	1	每日成交筆數
*	2	每日成交量
*	3	每日成交總值
*	4	證交所每日股價指數
*	5	工商時報每日股價指數
*	6	每日融資餘額
*	7	每日融券張數
*	8	每日上漲家數
*	9	每日下跌家數
*	10	每日 DJIA 股價指數
*	11	騰落指數
*	12	財氣指標
*	13	OBV 人氣指標
*	14	成交筆數變動百分比
*	15	成交量變動百分比
*	16	成交總值變動百分比
*	17	證交所股價指數變動百分比
*	18	工商時報股價指數變動百分比
*	19	融資餘額變動百分比
*	20	融券張數變動百分比
*	21	DJIA 股價指數變動百分比
*	22	財氣指標變動百分比
*	23	OBV 人氣指標變動百分比
*	24	成交筆數 6 日平均值
*	25	成交量 6 日平均值
*	26	成交總值 6 日平均值
*	27	證交所股價指數 6 日平均值
*	28	漲跌家數 6 日平均值
*	29	財氣指標 6 日平均值
*	30	乖離率 6 日平均值
*	31	RSI 6 日平均值
*	32	成交總值 6 日變動百分比
*	33	證交所股價指數 6 日變動百分比
*	34	每日漲跌家數 6 日變動百分比
*	35	成交總值 13 日平均值
*	36	證交所股價指數 13 日平均值
*	37	乖離率 13 日平均值
*	38	RSI 13 日平均值
*	39	成交總值 13 日變動百分比
*	40	證交所股價指數 13 日變動百分比
*	41	證交所股價指數 13 日平均值
*	42	證交所股價指數 25 日變動百分比
	43	成交筆數變動百分比型態
	44	成交量變動百分比型態
	45	證交所股價指數變動百分比型態
	46	融資餘額變動百分比型態
	47	融券張數變動百分比型態
	48	財氣指標變動百分比型態
	49	OBV 人氣指標變動百分比型態
	50	重要事件
	51	成交筆數 3 日變動百分比
	52	成交量 3 日變動百分比
	53	證交所股價指數 3 日變動百分比
	54	融資餘額 3 日變動百分比
	55	融券張數 3 日變動百分比

由於股市有追漲殺跌的情形，因此形成股價的各種波動現象。然而消極必長，當股價偏離平均線太遠時，便隨時有反轉的現象，因此，乖離率便可作為一個判斷的因子。我們保留了 6 日與 13 日平均值之乖離率，同時也保留 RSI 6 日平均值之線索。

股價的預測，往往須觀察較長的一段時間，亦即所謂的「趨勢」。在前述所保留的線索中，雖有變動百分比的線索，但只能觀察到兩期間變化的方向和幅度。因此第二階段的測試中，我們新增了七個變動型態的線索（表 4-1 中線索 43~49），由此可觀察到前三期指標連續變動的情形。

政治、經濟等重要事件往往會對股價造成衝擊。我們將利率、匯率和油價之變動等利多或利空因素，以計質性的指標列為新增的線索（表 4-1 中線索 50）。

本研究初期的測試，發現台灣投資人採取非常短線的操作策略。因此，我們又新增了 5 個「3 日變動百分比」的線索（表 4-1 中線索 51~55）。

## 二、例子資料庫（EXDB）：

本研究以 1984 年（民國 73 年）至 1986（民國 75 年）作為研究期間，並以證交所每一個交易日發生的實例，作為系統學習的例子的樣本。總計收錄 858 個例子。因為股市在這段期間尚稱穩定，股市證交指數變動的範圍不大。本研究依時序將 858 個例子分為二群，第一群為民國 73 年及 74 年之資料，計有 573 個例子，用以產生規則庫（Rule Base）；而第二群為民國 75 年之資料，計有 285 個例子，則作為本系統預測能力的驗證之用。

## 三、目標（Goal）設定

本研究係以台灣證券交易所的加權股價指數為預測標的，採逐日預測的方式進行。由於股市在本研究期間每日實際變動的幅度亦不大，因此，研究人員便以 0.3% 為分割點，設定股價指數變動的方向，分別是：

變動率  $> 0.3\%$ ，設定為「+」，表示上漲

變動率  $\leq |0.3\%|$ ，設定為「○」，表示無變動

變動率  $< -0.3\%$ ，設定為「-」，表示下跌

為使得本系統有效學習並進行預測，研究人員便將設定為「○」的例子刪除，僅以設定為「+」的及設定為「-」的例子供系統學習及預測之用。經過研究人員過濾之後，使得原先的 858 個例子減為 650 個例子。同樣的，研究人員仍將其分為二群，一群（433 個）作為系統學習之用，另一群（217 個）則供系統驗證效度之用。

#### 四、實例說明

在此舉一例說明以本演算法導出之分類樹與預測法則。假如從過去的例子中，發現成交量之變動、指數之變動以及 DJIA 指數之變動為重要線索，並希望以此線索預測證交指數之漲或跌（目標值）。假設例子資料庫收集了 10 個例子，而其所建立之 EXDB 如表 4-2 所示，根據本演算法，能夠獲得圖 4-1 分類之結果，並可表示成圖 4-2 之分類樹，而將 EXDB 中 10 個例子，依照三個線索的值 100% 的區別出來。

一般而言，在分類樹中，愈重要的線索愈會被事先選取，由圖中可看出三個線索對目標值分類之重要性依次為成交量、指數變動與 DJIA 變動，甚至不需要使用 DJIA 之線索，即可將所有的例子依漲跌之樣態，正確的分類出來。

#### 伍、研究結果分析與討論

##### 一、最佳分割方法：

首先，我們以計量線索的三種分割方法，並以例子資料庫（EXDB）中 433 個例子導分類樹，而以 217 個例子測試命中率，停止分割的純度訂為 80%。我們發現方法二得到命中率 59.90%，強態率為 42.9%，為三種方法中之最佳者。其分析之結果如表 5-1 所示。

由表 5-1 中之葉子數，我們亦可發現方法二所導得之分類樹為最小者，亦即最有效率者；其平均途徑（path）——由根點至葉子之距離，亦為三者中最短者；也就是說方法二能使用較少的線索，而得到較佳的分類效果。

由上述三種方法所導出之分類樹，亦可看出各線索相對之重要性。在方法一與方法二中，首先被選出來的線索是「每日成交筆數」，亦即系統發現此線索對於股市漲跌的預測，

以範例學習法預測台灣股市行爲

表 4-2 股市預測實例的例子資料庫 (EXDB)

樣本	證交指數 (goal)	成交量變動 (cue1)	指數變動 (cue2)	DJIA 變動 (cue3)
1	跌	減少	漲	跌
2	跌	持平	跌	跌
3	漲	持平	漲	漲
4	跌	減少	漲	漲
5	漲	增加	漲	漲
6	漲	增加	漲	跌
7	漲	持平	漲	跌
8	漲	增加	跌	跌
9	跌	持平	跌	漲
10	跌	減少	跌	跌

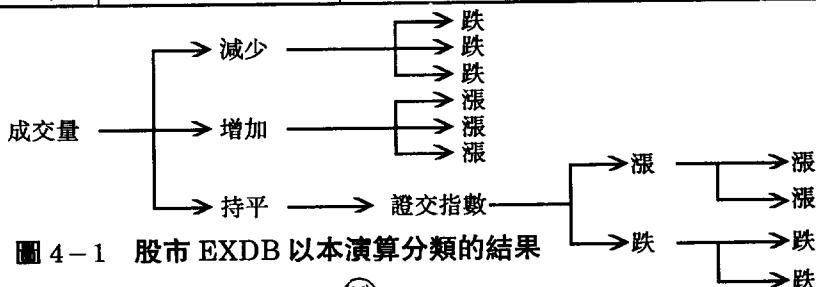
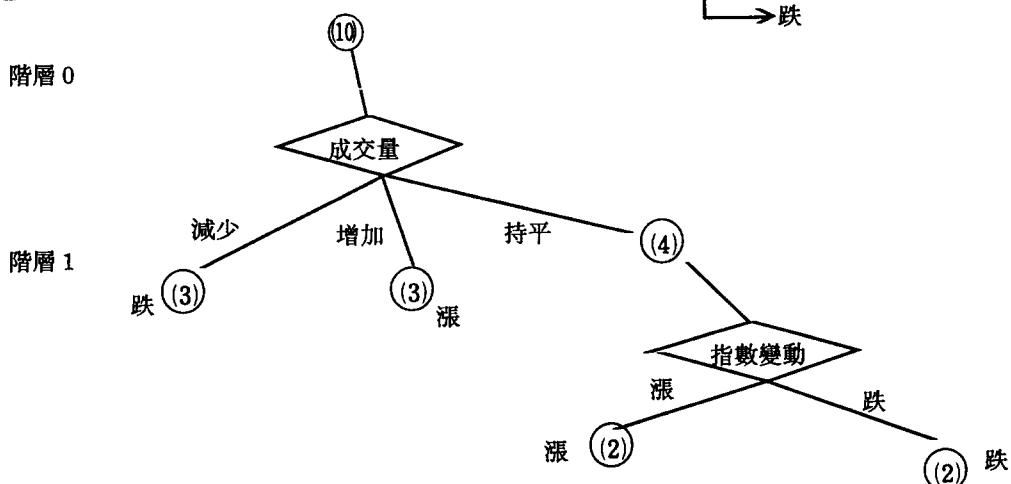


圖 4-1 股市 EXDB 以本演算分類的結果



註1. 節點中之數字代表例子數

2. 若將每一條路徑皆拆成一條法則，則可表示成下列四條獨立的法則：

- (1) IF 成交量 = 減少 THEN 下跌
- (2) IF 成交量 = 增加 THEN 上漲
- (3) IF 成交量 = 持平 and 指數變動 = 漲 THEN 證交指數將上漲
- (4) IF 成交量 = 持平 and 指數變動 = 跌 THEN 證交指數將下跌

圖 4-2 股市預測之分類樹圖

表 5-1 三種分割方法效率之比較

測試	分割方法	葉子數	命中率	純度高於 70% 葉子數	強態率
1	方法一	48	56.8%	16	33.3%
2	方法二	35	59.9%	15	42.9%
3	方法三	54	56.5%	15	28.0%

具有最大的判別能力。依股市「量先於價」的常識而言，這個結果並不讓人覺得特別意外。有趣的是，同樣是量的指標，每日成交筆數却優於每日成交量的線索。這或許可以說明，除了量的效果外，市場人氣之熱絡與否，也是影響股價變動重要的因素。

此外，較優先入選的線索，分別有成交筆數變動百分比，融券張數變動百分比，OBV 人氣指標變動百分比，6 日平均值乖離率，每日漲跌家數 6 日變動百分比、6 日平均值乖離率，證交所股價指數 25 日變動百分比，成交筆數變動百分比型態，重要事件，成交量 3 日變動百分比等線索。在三次測試中，新增的七個變動百分比型態和 5 個 3 日變動百分比的線索，雖亦出現在分類樹中，但並沒有特別顯著的表現。

## 二、預測能力的時效：

由於市場的行為，會隨時間而變化，以歷史資料所導出的規則，通常會漸漸失去其時效，而使得預測的能力降低。為了驗證以本系統所歸納出的法則，其預測能力所能持續之時效，我們將1986年的測試例子，分為四季，分別測試系統的命中率和強態率，結果如表 5-2 所示：

由表 5-2 發現預測第一季的能力最佳，而後效果較差。唯第二季之命中率較三、四季更差，此並不符合上述假想，這或許是 4~6 月之股市有其季節特性存在，使得預測能力減

弱。

表 5-2 季別的預測效果

	1~3月	4~6月	7~9月	10~12月
命中率	63.04%	57.14%	60.38%	59.62%
強態率	45 %	50 %	39 %	53 %

在 Braun & Chandler [ 1987 ] 的研究中，他們將例子分為兩群，第一年的例子用以導分類樹，第二年則分為四季，用來測試，結果四季的命中率分別是 77%，62%，62% 和 54%，呈明顯之下降趨勢。亦即時間隔得愈久遠，系統將會逐漸「落伍」，而漸漸失去其預測能力。然而，從另一個角度看，這也正是範例學習法最有價值的地方。因為，如果我們不斷地更新例子資料庫，餵以最新的例子，則系統便能依據新的例子資料庫學習出新規則，以適應新環境的變遷。

### 三、EXDB 例子數對預測效果之影響：

必須具備足夠的正例，才能使得所導得之分類樹完備（Completeness）；同樣的，必須具備足夠的負例，才能使得分類樹具一致性（Consistency）（參見 Michalski [ 1983 ]）。以股市而言，我們不可能涵括所有的正負例，因此導分類樹例子的多寡，便對其預測能力造成影響。本研究分別以半年（1985 年 7 月～12 月），一年（1985 年 1 月～12 月），一年半（1984 年 7 月～1985 年 12 月）和兩年（1984 年 1 月～1985 年 12 月）的例子資料庫所導出之分類樹，分別用 1986 年 1～3 月 和 1～12 月的例子來測試，其結果如表 5-3，5-4 所示。

表 5-3 EXDB 例子數對預測效果之影響

	半 年	一 年	一年半	兩 年
命中率	44.93%	44.88%	52.43%	59.90%
強態率	5.9 %	16 %	32 %	42.9 %

註：測試用例子 1986 年 1 月～12 月

表 5-4 EXDB 例子數對預測效果之影響

	半 年	一 年	一年半	兩 年
命中率	52.17%	39.13%	54.35%	63.04%
強態率	33 %	18 %	33 %	45 %

註：測試用例子 1986 年 1 月～3 月

由上述二表顯示，股市預測，至少須用兩年以上的例子所導出之分類樹，才有令人滿意的預測能力。

#### 四、分類樹修剪之效果：

關於分類樹修剪的效果，本研究分別設定純度為（80%，90%，100%）的三種停止分割準則，並將例子個數少於 5 個之葉子修剪掉。然後，以 1986 年之例子測試修剪的效果，其結果如表 5-5 所示：

一般而言，修剪之技術，多用於分類樹發展較完全後再進行。由上表中，修剪後之分類樹，在停止分割純度訂在 80% 時，並無較佳效果；然而在 90% 與 100% 時，則命中率和強

## 以範例學習法預測台灣股市行爲

態率皆能有所增進。

表 5-5 分類樹修剪效果之分析表

停止分割純度	80%		90%		100%		
	修 剪	無	有	無	有	無	有
命中率( % )	59.9	58.5	53.6	58.9	53.1	55.6	
強態率( % )	43	36	28	34	27	33	

## 陸、結論

本研究以範例學習法為基礎，發展出一套適用於股市行為研究之學習系統。並以此範例學習系統導出股市各種態樣之分類樹，以預測台灣證交所總體指數之漲跌。此系統能優先選取出影響指數變動較具區別能力的線索；同時，經由實例的驗證，此系統具備預測股市總指數漲跌的能力。尤其重要的是，本系統具有學習的能力，可不斷地藉由新的例子更新其法則以適應環境的變遷。然而，由於影響股市的因素錯綜複雜，本研究僅以技術面之線索為主，並做較短期的預測，未來應進一步擴展至基本面及較長期的預測，期能對股市之行為有更進一步的了解。此外，範例學習研究法是用以發現一組例子間關係之方法，它應該和傳統的一些著名的統計方法相互比較（例如：區別分析），比較範例學習法與統計方法兩者間之差異，並探討在何種情況下，採取那一種方法較適宜，這些都是相當值得研究的課題，也是我們今後的研究方向之一。

## 參考文獻

### 一、英文部份：

- [1] R. Brady [ 1984 ] , " Charting a Path in the Market " , *National Business*, 16 ( February ).
- [2] H. Braun and S. S. Chandler [ 1987 ] , " Predicting Stock Market Behavior through Rule

- Induction : an Application of the Learning – From – Example Approach " , *Decision Science*, Vol.18, No.3.
- [ 3 ] L. Breiman, H. J. Friedman, R. A. Olshen & C. J. Stone [ 1984 ] , *Classification and Regression Trees*, Wadsworth, Belmont, California.
  - [ 4 ] B. G. Buchanan, et. al. [ 1976 ] , " Applications of Artificial Intelligence for Chemical Inference XXII. Automatic Rule Formation in Mass Spectrometry by means of the Meta – DENDRAL Program " , *Journal of the American Chemical Society*, 98.
  - [ 5 ] B. G. Buchanan and E. A. Feigenbaum [ 1978 ] , " DENDRAL and Meta – DENDRAL, their Applications Dimension " , *Artificial Intelligence*, 11.
  - [ 6 ] C. Carter & S. Catlett [ 1987 ] , " Assessing Credit Card Application using Machine Learning " , *IEEE Expert*.
  - [ 7 ] E. F. Fama [ 1970 ], " Efficient Capital Markets: A Review of Theory and Empirical Work " , *Journal of Finance* ( May ) .
  - [ 8 ] E. F. Fama [ 1976 ], *Foundation of Finance, Basic Book*, New York.
  - [ 9 ] E. F. Fama & M. E. Blume [ 1966 ] , " Filter Rules and Stock Market Trading Profits " , *Journal of Business, Special Supplement* ( January ) .
  - [ 10 ] E. F. Fama et. al., [ 1969 ] , " The Adjustment of Stock Prices to New Information " , *International Economic Review* ( February ) .
  - [ 11 ] R. G. Ibbotson [ 1976 ] , " Price Performance of Common Stock New Issues " , *Journal of Financial Economics* ( September ) .
  - [ 12 ] J. F. Jaffe [ 1974 ] , " Special Information and Insider Trading " , *Journal of Business* ( July ) .
  - [ 13 ] M. Jensen [ 1967 ] , " Random Walks : Reality or Myth " , *Financial Analysis Journal*.
  - [ 14 ] J. Liscio [ 1987 ] , " Simple Technical Analysis of Stocks and Index Futures " , *InterMarket*, 4.
  - [ 15 ] D. B. Lenat [ 1976 ] , " AM : an Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search " , *Ph. D. dissertation, Stanford University*.
  - [ 16 ] R. S. Michalski & J. G. Carbonell & T. M. Mitchell, Eds. [ 1983 ] , *Machine Learningfl An Artificial Intelligence Approach*, Tioga Press ( Palo Alto, CA ) .
  - [ 17 ] R. S. Michalski & R. L. Chilausky [ 1980 ] " Learning by Being Told and Learning From Examples : an Experimental Comparison of the two Methods of Knowledge Aquisition in the Context of Developing an Expert System for Soybean Disease Diagnosis " , *Journal of Policy Analysis and Information Systems*, 4.
  - [ 18 ] V. Niederhoffer & M. F. M. Osborne [ 1966 ] , " Market Making and Reversal on the Stock Exchange " , *Journal of the American Statistical Association*.
  - [ 19 ] P. O' Rorke, [ 1982 ] , " A Comparative Study of Two Inductive Learning Systems AQ11 and ID3 Using a Chess Endgame Test Problem " , *Intelligent Systems Group Report 82 – 2*,

## 以範例學習法預測台灣股市行爲

University of Illinois, Department of Computer Science.

- [20] J. R. Quinlan [ 1979 ] " Discovering Rules from Large Collections of Examples : A case study ", *Expert Systems in the Micro Electronic Age*, D. Michie ( Ed. ) Edinburgh University Press.
- [21] J. R. Quinlan, et. al. [ 1987 ], " Inductive Knowledge Acquisition : A Case Study ", in *Applications of Expert Systems* , J. R. Quinlan ( Ed. ) Addison – Wesley.
- [22] R. Randleman, C. Jones & H. A. Latane [ 1982 ], " Empirical Anomalies Based on Unexpected Earnings and the Importance of the Risk Adjustments ", *Journal of Financial Economics*(Nov.).
- [23] J. C. Robert [ 1983 ], " Technical Analysis for the Private Investor ", *Wealthbuilding* ( September ).
- [24] R. C. Rogers [ 1988 ], " The Relationship Between Earnings Yield and Market Value : Evidence from the American Stock Exchange ", *Financial Review*, 23.
- [25] A. Seelenfreund, G. Parker & J. Van Horne [ 1968 ], " Stock Price Behavior and Trading ", *Journal of Financial and Quantitative Analysis* ( September ).
- [26] C. E. Shannon & W. Weaver [ 1975 ], *The Mathematical Theory of Communication*, University of Illinois Press ( Urbana ).
- [27] A. Shapiro & T. Niblett [ 1982 ], " Automatic Induction of Classification Rules for a Chess Endgame ", *Advances in Computer Chess*, Vol. 3, M. R. B. Clarke ( Ed. ), Edinburgh University Press.
- [28] P. H. Winston [ 1975 ], " Learning Structural Descriptions from Examples ", *The Psychology of Computer Vision*, P. H. Winston (Ed.), McGraw Hill, New York.

## 二、中文部份：

- [1] 王官品〔民國 75 年〕，「上市公司每月盈收公告與股價變動關係之研究」，國立中興大學企管研究所碩士論文。
- [2] 李廣進〔民國 69 年〕，「台灣地區新上市股票股利短期投資報酬率之研究」，國立台灣大學商學研究所碩士論文。
- [3] 林芸萱〔民國 76 年〕，「台灣地區油品價格調整對證券市場股票價格影響之實證研究」，國立政治大學企管研究所碩士論文。
- [4] 林煜宗〔民國 67 年〕，「市場因素對台灣證券市場股價變動之影響」，證交資料（6月 25 日）。
- [5] 金柏西〔民國 69 年〕，「台灣股票市場股利對股價影響之實證研究」，大同工學院事業經營研究所碩士論文。
- [6] 徐世豪〔民國 68 年〕，「台灣證券有效性之研究——過濾法投資效益之評估」，國立政治大學企管研究所碩士論文。
- [7] 劉仲偉〔民國 76 年〕，「應用於股票投資基本分析知識庫系統研究與設計」，國立台灣大學商學研究所碩士論文。

- [8] 簡仁德〔民國 70 年〕，「台灣證券市場價格變動習性為隨機漫步假定之實證分析」，淡江大學管理科學研究所博士論文。