

魏巧宜、馬麗菁 (2017), 『利用混合模式預測國內股票型基金績效及多年期軌跡』, 中華民國資訊管理學報, 第二十四卷, 第一期, 頁 69-96。

利用混合模式預測國內股票型基金績效及多年期軌跡

魏巧宜

國立聯合大學資訊管理學系

馬麗菁*

國立聯合大學資訊管理學系

摘要

隨著經濟發展，大眾對於投資理財的需求與日俱增，如何協助不同投資時間長短偏好的投資者，找出較佳短、中及長期績效的基金，是一項受關注的議題。本研究以自組織映射圖結合倒傳遞類神經網路及基因演算法，協助不同投資時間長短偏好的投資者，找出較佳短期、中期、長期績效的基金。此外，過去的研究大多是分析單一期間的基金績效，本研究進一步以自組織映射圖結合案例式推理法及基因演算法，進行多年期群集軌跡預測。研究結果顯示無論短、中及長期績效預測結果，皆是以本研究所提出的混合模式預測結果最好。在多年期軌跡的預測分析方面，本研究提出的多年期群集預測方法，在風險趨勢預測效果上，亦優於單一期間預測分析結果。

關鍵詞：混合模式、自組織映射圖、案例式推理、多年期軌跡、基金績效

* 本文通訊作者。電子郵件信箱：lcma@nuu.edu.tw
2013/08/29 投稿；2016/01/18 修訂；2016/09/19 接受

Wei, C.Y. and Ma, L.C. (2017), 'Using hybrid models to predict domestic equity fund performance and multi-year trajectory', *Journal of Information Management*, Vol. 24, No. 1, pp. 69-96.

Using Hybrid Models to Predict Domestic Equity Fund Performance and Multi-Year Trajectory

Chiao-Yi Wei

Department of Information Management, National United University

Li-Ching Ma*

Department of Information Management, National United University

Abstract

Purpose – As economic growth continues, the needs for investments are increasing. How to assist investors with various preferences in finding out mutual funds with better performance is an important issue. This study aims to propose a framework to analyze the relationships between fund attributes and performance for different investment horizon and to predict multi-year trajectory.

Design/methodology/approach – This study combines self-organizing map, genetic algorithms and back-propagation neural network to analyze the relationships between fund attributes and performance for short, middle and long time horizon. Moreover, this study incorporates the concept of self-organizing map, case-based reasoning and genetic algorithms to conduct multi-year trajectory analysis.

Findings – The results show that the proposed hybrid approach yields the best prediction performance for all investment horizons. In addition, the proposed multi-year trajectory analysis is better than single-period analysis in predicting risk trends.

Research limitations/implications – Because data period adopted in the empirical study is only five years, long-term stability of the proposed framework has not been verified.

* Corresponding author. Email: lcma@nuu.edu.tw
2013/08/29 received; 2016/01/18 revised; 2016/09/19 accepted

Practical implications—This paper provides several implications for investors, fund managers and researchers. Investors can get more knowledge about the relationships between fund attributes and performance for different investment horizon. Fund managers and researchers can pay more attention to multi-year trajectory analysis.

Originality/value—This study proposes an approach to analyze the relationships between fund attributes and performance for different investment horizon. Rather than analyzing single-period performance only, this study incorporates several methods in business intelligence to conduct multi-year trajectory analysis.

Keywords: hybrid model, self-organizing map, case-based reasoning, multi-year trajectory, fund performance

壹、緒論

隨著經濟發展，生產力與國民所得提升，大眾對於投資理財的需求與日俱增。傳統上，大眾多選擇低風險的銀行存款為主要理財方式，但是自 2004 年以來，台灣央行的存款利率調降至 2% 以下，加上消費者物價指數不斷的增長，銀行存款的相對價值大幅下降；因此，銀行存款已無法滿足人民對於理財的需求，許多投資者開始尋找其他的投資管道 (Li & Kuo 2008)。

目前投資理財商品繁多，一般大眾常受限於時間或專業知識的不足，無法親自操作，使得由基金經理人代為投資操作且風險較低的共同基金廣受大眾歡迎。然而，基金市場是充滿不確定性的，目前共同基金的評價大多是以星期或月份為單位進行定期資料更新，且資料僅用於比較歷史紀錄。在充滿不確定性的市場中，評估系統應該要能夠根據投資者的需求，不斷地進行更新 (Wang & Huang 2010)，可見依據不同投資者的需求與偏好，提供不同的投資選擇與服務的重要性。因此，當投資人選擇基金投資時，如何協助不同投資時間長短偏好的投資者，找出較佳短期、中期、長期績效的基金，是一個值得探討的問題。

過去有關基金績效的相關研究，大多運用迴歸分析法 (regression analysis; RA) 預測基金績效 (Prather et al. 2004)，但迴歸分析法有許多限制，如自變數需互相獨立等，因此，一旦自變數不具獨立性時，常會影響預測之準確性 (尹相志 2009)。近幾年來，有許多研究利用資料探勘方法，如類神經網路 (Wang & Huang 2010) 及基因演算法 (Tsai et al. 2011) 等預測基金績效。類神經網路具有平行處理能力，基因演算法具有多點搜尋的優點，均很適合用於分析大量的基金資料。

本研究希望協助不同投資時間長短偏好的投資者，找出較佳短期、中期、長期績效的基金，並提供多年期軌跡的預測分析，作為投資決策的參考。大多數的研究僅著重在單期間的績效預測，即不分年度，在資料期間利用時間序列分析，找出各變數的權重，再以前期資料預測次期報酬率。然而，僅以前期資料預測次期資料，存在很大的風險，若前期資料受特殊因素影響 (如金融風暴)，產生較大變異，單期間資料預測方式的正確率可能會偏低。此外，邱永和等 (2008) 採用非傳統式資料包絡法 (DEA)，針對民國 91 年至 95 年國內股票型基金進行績效評估，並以各年之 Super SBM 效率值排名，進行 Spearman 相關性檢定，以分析基金績效是否具有其持續性，結果顯示各年度皆為正相關，亦表示基金績效具有持續性。因此，本研究除了希望找出預測單期間基金績效較好的模式外，並提出了一個分析多年期軌跡的方法，先將資料依年度區分，找出單年度資料的群集關係，並進一步分析多年期間的群集軌跡變化，以預測多年期軌跡趨勢，希望能

改善單期間預測法的限制。

在單期間基金績效預測方面，根據 Thomassey 與 Fiordaliso (2006) 的研究，將資料先以相似性分群再加以分類的方式，能有效提升紡織業原料銷售的預測正確率。本研究亦利用分群加分類的混合模式，預測單期間基金績效，先利用自組織映射圖 (self-organizing map; SOM) 將資料依相似性分群，接者根據分群結果進行倒傳遞類神經網路 (back-propagation neural network; BPN) 與基因演算法 (genetic algorithms; GA) 分析。在多年期軌跡分析方面，先利用自組織映射圖找出各年度資料的群集關係，再利用案例式推理法 (case-based reasoning; CBR) 及基因演算法，找出最相似的群集軌跡及分析各年度的權重，以預測多年期軌跡趨勢。最後結合單期間及多年期軌跡分析結果，預測與分析短、中及長期基金績效。

本研究結合倒傳遞類神經網路、基因演算法、自組織映射圖及案例式推理法等方法，進行資料探勘與分析。倒傳遞類神經網路擁有處理非線性問題及平行處理之能力，可快速處理大量資料，因而能快速且有效地預測基金績效；基因演算法有多點搜尋的優點，可避免落入局部最佳解；自組織映射圖能有效地聚類出特質相似性較高之基金，且擁有較高的模式可靠度；案例式推理法能比較基金多年期軌跡的案例特徵，並分析出最相似的案例軌跡。本研究結合此四種方法之優點，以建構出單期間基金績效預測及多年期軌跡分析模式。本研究之主要目標如下：

1. 單期間基金績效預測：分析不同群集間，基金屬性與基金短、中及長期績效之關係，並預測不同群集之基金績效；
2. 多年期軌跡分析：分析多年期間的群集軌跡變化，以預測多年期軌跡趨勢。

貳、文獻探討

一、基金分類方式與屬性

基金的種類繁多，目前有許多種不同的分類方式，舉例來說，基金若依據發行地點來區分，可分為國內、海外基金，國內基金是指國內基金公司所發行的各種基金，銷售對象以中華民國人民為主。根據 Chan 與 Covring (2012) 所提出國內投資者在進行海外投資時，常因文化、語言及地理環境等因素影響，且受限於信息不對稱的情形下，對海外投資市場卻步。再加上過去幾年，國內基金數呈現逐漸增長的趨勢，如圖 1 所示，國內基金數量自 2001 年以來，除了 2008 年因金融海嘯而略減外，皆呈現逐漸增長的情形，故本研究以國內基金為主。若依據基金的發行方式區分，可分為開放型、封閉型基金，開放型基金是指投資者可隨時

依據基金淨值報價，向基金公司買進、轉換或贖回。若以基金的投資標的加以分類，可分為股票、債券、貨幣、貴金屬、指數、特殊產業及保本型基金，依據中華民國證券投資信託暨顧問商業同業公會（SITCA）統計資料顯示，以股票型基金為大宗（約佔 61%）。本研究採用國內開放型股票型基金作為本研究之研究標的。

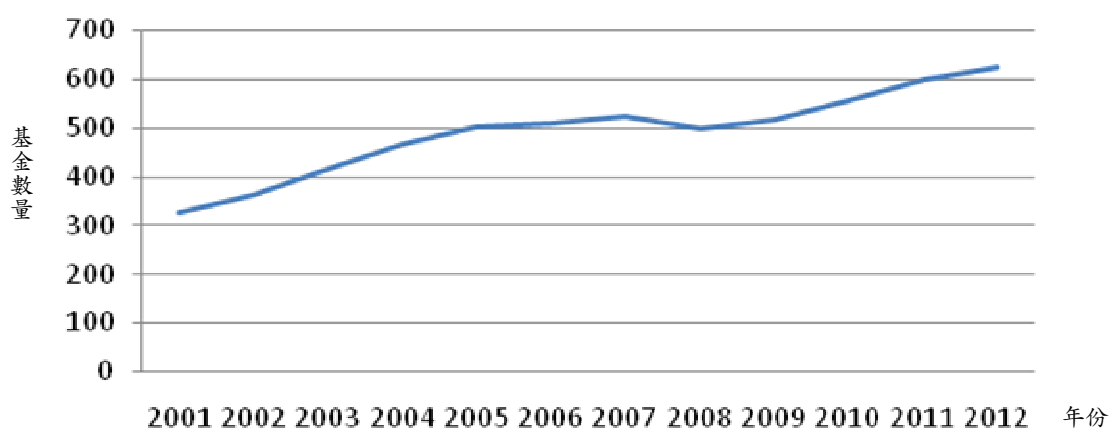


圖 1：台灣共同基金數量圖（資料來源：SITCA）

過去許多學者指出，某些基金屬性常會影響基金績效，例如：Indro 等（1999）曾提到年平均週轉率、年化標準差及一年報酬率可能會影響基金績效；Babalos 等（2012）以總費用率、年化標準差作為輸入變數對共同基金績效進行分析；Bauer、Koedijk 與 Otten（2005）以基金成立年數、總費用率作為評估基金績效的標準；Lai 與 Lau（2010）研究 1990~2005 年馬來西亞 79 種伊斯蘭基金的表現，面對短期與長期兩種不同的投資期間，發現平均而言私募基金 Beta 值比政府資助的基金組合低，但是私募基金的表現是優於市場指數和政府基金的；Shu、Yeh 與 Yamada（2002）研究投資者對台灣開放型股票型共同基金的買入和賣出行為，分析變數包括季報酬率、半年報酬率及一年報酬率等共 12 個變數，結果顯示小額投資者大多投資大型基金，並傾向於能在短期內獲得績效的基金，而大額投資者傾向於投資小型基金，且對短期績效較不敏感。本研究亦採用基金成立年數（Bauer et al. 2005）、基金淨資產（Prather et al. 2004）、年平均週轉率（Indro et al. 1999）、總費用率（Babalos et al. 2012）、年化標準差、Beta 值（Lai & Lau 2010）、季報酬率（Herrmann & Scholz 2013; Shu et al. 2002）、半年報酬率、一年報酬率（Badrinath & Gubellini 2011）、二年報酬率（Javier 2013）、三年報酬率及五年報酬率等十二項常用基金屬性，作為可能影響基金績效的變量詳如表 1 所示。根據 SITCA 民國 92 年年報，指出評估基金績效所慣用的期間，可分

為九種評估期。其中，又可分為短期（一年內）、中期（三年）、長期（三年以上）績效，故本研究分別以一年報酬率、三年報酬率及五年期報酬率，代表短、中、長期基金績效。另外，值得注意的是投資者多希望兼顧風險最小化與報酬率最大化，故常採用納入風險溢酬的基金績效指標如 Sharp Ratio 等，但因本研究輸入變數已包含年化標準差及 Beta 值風險指標，若預測輸出值設定為 Sharp Ratio，恐會產生輸入變數與輸出變數相關性過高的問題，故未採用 Sharp Ratio 為本研究輸出基金績效指標。

表 1：基金屬性變數對照表

變數	涵義	說明	資料類型	參考文獻
A ₁	基金成立年數	基金從成立日起至今的累積年數	數值⇨轉為低、中、高（三區間）	Bauer et al. (2005) ; Prather et al.(2004)
A ₂	基金淨資產	市場收盤價所計算出的總資產價值扣除基金當日的各類成本及費用	數值⇨轉為低、中、高（三區間）	Prather et al. (2004) ; Yan (2008)
A ₃	年平均週轉率	平均一年內基金買進、賣出股票的次數	數值⇨轉為低、中、高（三區間）	Indro et al. (1999) ; 李顯儀等 (2011)
A ₄	總費用率	關於基金的總費用支出	數值⇨轉為低、中、高（三區間）	Bauer et al. (2005) ; 李顯儀等 (2011)
A ₅	年化標準差-12 個月	指一年內基金淨值波動的情況	數值⇨轉為低、中、高（三區間）	Indro et al. (1999) ; 馬麗菁等 (2011)
A ₆	Beta 值	衡量基金相較於市場報酬率波動的幅度	數值⇨轉為低、中、高（三區間）	Lai & Lau (2010)
A ₇	季報酬率	基金最近三個月報酬率	數值⇨轉為低、中、高（三區間）	Shu et al. (2002) ; Herrmann & Scholz (2013)
A ₈	半年報酬率	基金最近半年報酬率	數值⇨轉為低、中、高（三區間）	Shu et al.(2002) ; 馬麗菁等 (2011)
A ₉	一年報酬率	基金最近一年的報酬率	數值⇨轉為低、中、高（三區間）	Indro et al. (1999); Shu et al. (2002)
A ₁₀	二年報酬率	基金最近二年的報酬率	數值⇨轉為低、中、高（三區間）	Carhart (1997) ; Javier (2013)
A ₁₁	三年報酬率	基金最近三年的報酬率	數值⇨轉為低、中、高（三區間）	Carhart (1997) ; Badrinath & Gubellini (2011);
A ₁₂	五年報酬率	基金最近五年的報酬率	數值⇨轉為低、中、高（三區間）	Carhart (1997) ; Badrinath & Gubellini (2011)

二、資料探勘方法

本研究所採用或比較的分析方法可分迴歸分析法、倒傳遞類神經網路、基因演算法、自組織映射圖、案例式推理法及混合方法，分述如下。

迴歸分析法是常被使用於預測或比較的統計方法，例如，Bianco、Manca 與 Nardini (2009) 應用不同的迴歸模型於義大利的電力消費預測；Ng、Skitmore 與 Wong (2008) 運用基因演算法結合線性迴歸分析模型探討未來建築業的供給和需求預測。迴歸分析法的優點為對簡單問題與小型數據是有效且不耗時的方法，且迴歸係數容易解釋 (Antonio et al. 2011)。迴歸分析的缺點為資料必須呈現常態分配、自變數必須具獨立性及誤差項的變異數必須相等 (蕭文龍 2009)。本研究亦將迴歸分析法作為分析與比較的方法之一。

類神經網路起源於生物神經網路的結構或功能，是由許多可以相互連結的人工神經元組合而成的非線性信息處理器 (Gan 2012)。倒傳遞類神經網路是透過輸入層、隱藏層及輸出層不斷反覆進行訓練的方式，並以實際目標值與倒傳遞類神經網路的輸出結果間的誤差，不斷地進行權重值修正，直到收斂為止。Leigh、Purvis 與 Ragusa (2002) 提出類神經網路是目前決策和預測方法中應用最廣泛的分類與預測工具，其中以倒傳遞類神經網路模式最具代表性 (Leigh et al. 2002)。倒傳遞類神經網路的應用非常廣泛，例如 Tsang 等 (2007) 應用倒傳遞類神經網路於香港股市，並建立一個股票買入與賣出警報系統；Law (2000) 採用倒傳遞類神經網路法於非線性的旅遊需求預測；Chen 等 (2010) 採用倒傳遞類神經網路法進行洪水預報。倒傳遞類神經網路具有以下的優點，首先，倒傳遞類神經網路具有豐富的假設空間可以分析非線性關係 (Berardi et al. 2004)。第二，倒傳遞網路具有普遍逼近任何目標函數，並且處理大量資料仍然可以達到所需的準確度。第三，由於倒傳遞類神經網路是屬於多層網路，因此可以處理輸出與輸入變數間複雜的關係 (施雅月與賴錦慧 2008)。然而，倒傳遞網路具有以下的限制，Rumelhart、Hinton 與 Williams 等 (1986) 和 Akarachai 與 Daricha (2007) 曾提及倒傳遞網路是屬於單點局部搜尋方式，因此容易陷入局部最佳解，類神經網路的原始概念為將目標值與實際輸出值之間的誤差函數最小化，透過調整網路的連結加權值，得到最佳加權值，因此，若沒有足夠的樣本，則可能受到初始權重隨機產生的影響，易陷入局部最佳解。此外，與基因演算法的多點搜尋 (多個起始解) 方式相比較，倒傳遞類神經網路僅有單一起始解，並且不似基因演算法具有突變運算元得以跳脫局部最佳解。基於倒傳遞類神經網路的諸多優點，本研究亦採用倒傳遞類神經網路進行單期間基金績效預測模式分析，並且進一步針對其可能陷入局部最佳解的情況進行改進。基因演算法是目前應用廣泛的最佳化搜尋方法之一，是由 Holland (1975) 依據達爾文「適者生存，不適者淘汰」的觀點所

提出的。基因演算法是一個全域性的搜索過程，通常在建構最佳化搜尋過程中，以選擇、複製、交配及突變為主要的步驟。過去有許多研究採用基因演算法解決最佳化問題，例如，Renner 與 Ekart (2003) 應用基因演算法於不同領域的工程設計，以提供最優化的設計方案；Chiu (2002) 提出一種基因演算法結合案例式推理的系統，對全球保險直銷公司進行顧客購買行為預測分析。基因演算法的優點是可解決大量資料與非線性空間的問題，且由於基因演算法是採多點搜尋，再加上演算法的特性，因此不易陷入局部最佳解 (Kuo & Lin 2010)。而基因演算法的缺點為目標函數必須針對問題而進行設定，及停止搜尋規則較不可靠 (Sexton & Dorsey 2000)。本研究應用基因演算法於兩個部份，第一部份是用來改善倒傳遞類神經網路可能會落入局部最佳解的問題，第二部份為調整案例式推理法因素的權重，以提升預測準確度。

自組織映射圖 (Kohonen 1990) 是類神經網路中最著名的分群工具，以無監督式網路為主，以特徵映射的方式，將高維度的輸入資料映射至低維度的特徵空間上，可分為輸入層、輸出層及網路連結三部份。許多研究採用自組織映射圖進行分群及預測，例如，Carlos (1996) 應用自組織映射圖進行公司財務資料分析，並開發一個決策支援系統；Hsu (2011) 提出一個結合自組織映射圖與基因規劃法之模式，並應用於股票價格預測；Jardin 與 Severin (2011) 提出一個自組織映射圖模型，並結合整體資料的多年期軌跡進行分析，結果顯示此模型不但擁有高準確度，且能保持其穩定性。自組織映射圖可有效地依資料的相似性進行群集分析。然而，自組織映射圖必須給定網格數，但網格數往往需要透過問題了解進行設定，或者必須不斷地嘗試以達到最好的分群效果 (Tan et al. 2005)。本研究應用自組織映射圖進行群集分析，並建構出基金績效多年期群集軌跡，以追蹤及預測多年期軌跡走向。

案例式推理 (Chang et al. 2006) 源自於動態記憶，進而發展為一種以過去經驗來推理的技術，嘗試從舊有的經驗中找尋最相似之經驗，藉由過去類似經驗的結果判斷可能解決的新問題。不少研究採用案例式推理法進行分析，例如，Chang 等 (2006) 整合基因演算法與案例式推理法，建構出一個高準確性、高效率的滯銷書退書預測系統，以提升退書決策的精準度。案例式推理法最大的優點在所有案例都能根據過去的案例，檢索出合適的解決方法 (Zhuang et al. 2013)。然而，大部分案例式推理法對於各個因素都具有相同的權重，因此，無法反映實際的情況，預測能力也會受到影響 (Chang et al. 2006)。本研究利用案例式推理法檢索過去多年期群集軌跡走向，並結合基因演算法調整不同年度的權重，以提高多年期群集軌跡預測的準確度。在機器學習領域中，混合方法一直被應用於各種領域，以改善單一學習方法的分類或預測績效 (Jain & Avadhnam 2007)。一般而言，混合方法是基於兩種不同的機器學習方法相結合的技術 (Tsai & Chen

2010)。例如：Tsai 與 Chen (2010) 採用四種混和模式預測信用評等之最佳績效及最大利潤，結果顯示分類結合分類的混合模式較其他模式能提供較好的預測結果。Thomassey 與 Fiordaliso (2006) 提出將資料先以相似性分群再加以分類的方式，能有效提升紡織業原料銷售的預測正確率。因此，本研究亦以混合模式為主軸，結合不同方法的優點，以提升預測正確率。

參、研究方法

研究方法分為研究模式、資料探勘流程及研究架構三部份，分述如下。

一、研究模式

經由文獻探討及與財金專家討論，整理出本研究之研究模式，如圖 2 所示。本研究之研究模式主要是以基金屬性作為基金績效構面之預測變項。其中，各基金屬性變數包括基金成立年數、基金淨資產、年平均週轉率、總費用率、年化標準差-12 個月、Beta 值、季報酬率、半年報酬率、一年報酬率、二年報酬率、三年報酬率及五年報酬率等十二項，分別以 A_1, \dots, A_{12} 表示；基金績效分為短期、中期、長期績效，並分別以次期一年報酬率、三年報酬率及五年期報酬率表示。基金屬性、說明、資料類型及參考文獻對照表如表 1 所示。

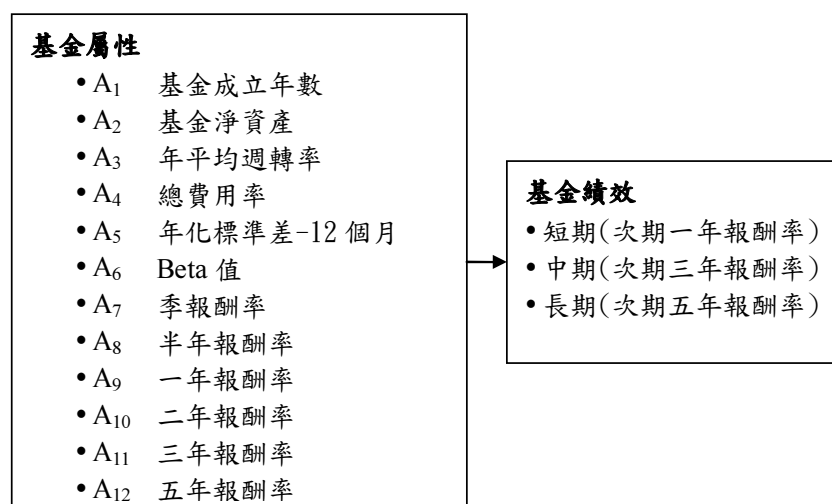


圖 2：本研究之研究模式

二、資料探勘流程

本研究之資料探勘流程分為資料蒐集、資料前置處理及資料探勘與分析三部

份，如圖 3 所示。在資料蒐集部份，在本研究中分析的變數為基金屬性，主要包含基金成立年數、基金淨資產等 12 項指標。本研究之資料來源取自台灣經濟新報資料庫 (Taiwan Economic Journal; TEJ)、基智網及中華民國證券投資信託暨顧問商業同業公會 (Securities Investment Trust and Consulting Association of R.O.C.; SITCA)，資料範圍擷取自 2007 年 1 月至 2011 年 12 月共 5 年的股票型基金資料，並以月為單位計算，總資料筆數約 33,998 筆。

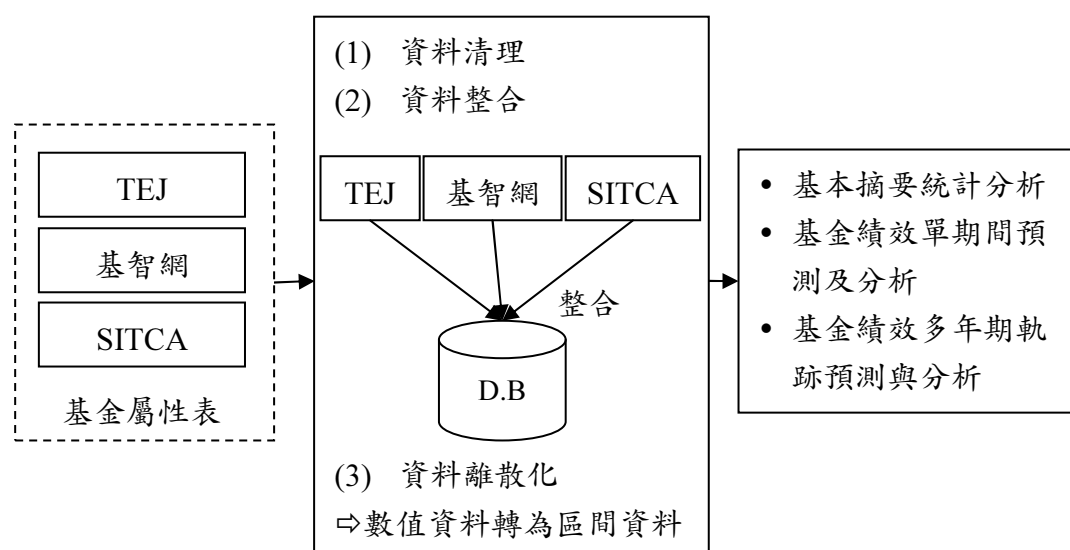


圖 3：本研究資料探勘流程圖

在資料前置處理部份，分為資料清理、資料整合及資料離散化三個步驟。首先，資料清理主要是針對蒐集到的資料進行遺漏值及雜訊資料偵測。而本研究所蒐集有缺值的基金資料中，因大多是缺少三年及五年之報酬率，經進一步檢查發現主要原因為該基金已清算，無法追蹤及分析五年趨勢，因此，本研究選擇刪除有缺值的基金資料。第二步驟為資料整合，針對不同資料來源所蒐集到的資料彙整為同一資料庫，並檢測是否有多餘性及資料不一致的問題。本研究資料來源除了變數 A_4 總費用率資料取自於基智網及 $A_{10} \sim A_{12}$ 取自於 SITCA 外，其餘皆是取自於 TEJ，每個來源皆會根據該基金名稱、代號、日期及變數公式進行資料比對及檢測是否存在資料多餘性與一致性問題。最後，資料離散化是採用等頻分割法，等頻分割法是常用的資料轉換方式，可將數值資料轉換為類別資料，先將資料依數值大小進行排序，接著將資料劃分為數個資料數量大致相同的區間，以達到資料精簡的目的。本研究採用等頻分割法主要是因為在不同區間的基金筆數相似，較不會因各區間中基金筆數差異過大產生影響，且等頻分割法可避免極端值

的影響。在國內基金篩選常用的四四三三法則中，將短期報酬率分為三等，長期報酬率分為四等，因本研究所採用變數多數以一年以內期間為計算基準，故取三區間，又考慮變數分群的一致性，故將所有變數均採三區間。

資料探勘與分析方面，先針對資料做摘要統計分析，再進行單期間基金績效預測，並與不同之模式進行比較。最後，針對預測結果最佳之群集數進行多年期基金軌跡分析。

三、研究架構

本研究結合倒傳遞類神經網路、基因演算法、自組織映射圖及案例式推理法等四種方法之優點進行資料探勘與分析，以十二個常用基金屬性，對短、中、長期基金績效，做單期間基金績效及多年期軌跡的預測與分析。本研究架構如圖 4 所示，分為單期間與多年期兩個部份，分述如下。

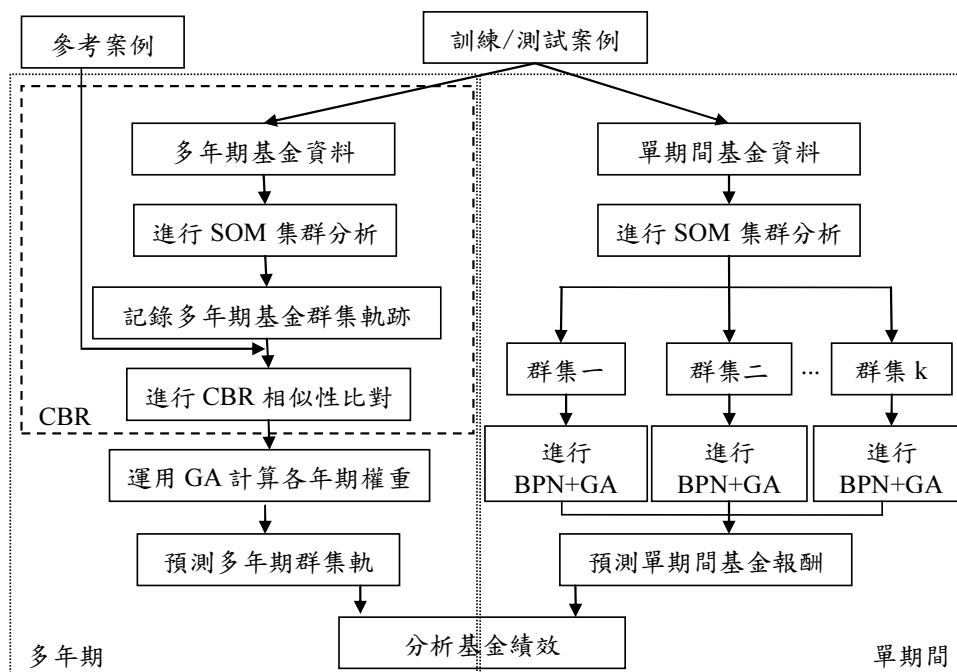


圖 4：本研究之研究架構圖

(一) 基金績效單期間預測與分析

在單期間預測方面，流程如圖 4 右半部，利用自組織映射圖將資料以相似性分群，接著，根據分群結果進行倒傳遞類神經網路結合基因演算法分析，以透過

分群加分類方式預測基金報酬率。倒傳遞類神經網路有能處理非線性問題及運算速度快等優點，但可能受隨機設定的起始值影響而落入局部最佳解；基因演算法具有多點搜尋的特性，可改善倒傳遞類神經網路的問題，故本研究結合此兩種方法，先以倒傳遞類神經網路找出十組最佳解，將其設定為基因演算法的初始解，再進行基因演算法分析，以提升預測的正確率。本研究在單期間預測部份，是依時間序列採移動視窗法（Sliding window）分析，即以前期資料預測次期結果，並依時間軸往後以視窗方式移動資料輸入與預測期間。

在自組織映射圖群集分析方面，輸入單元設定為基金屬性之 12 個變數，將 12 個維度空間向量資料映射到二維度的拓撲空間，其中 A_1, \dots, A_{12} 表示基金屬性之 12 個變數，輸出則為分群結果。此外，本研究並針對分群結果進行單變量變異數分析（ANOVA）檢定，藉此分析出不同群組間的主要差異。

在倒傳遞類神經網路部份，架構如圖 5 所示，輸入層類神經元採用基金屬性 A_1, \dots, A_{12} 等 12 個變數，輸出層有 1 個類神經元，以 O_1 表示，分別採用短期、中期、長期基金績效，隱藏層神經元數量是採用常用的輸入層神經元數加上輸出層神經元數除以二的方式計算，因此，隱藏層神經元設定為 7 個，以 H_1, \dots, H_7 表示。倒傳遞類神經網路輸入層到隱藏層之間的權重值以 W_{ij} 表示；隱藏層到輸出層之間的權重值以 W_{jk} 表示；隱藏層偏差值以 b_j ；輸出層偏差值以 b_k 表示。

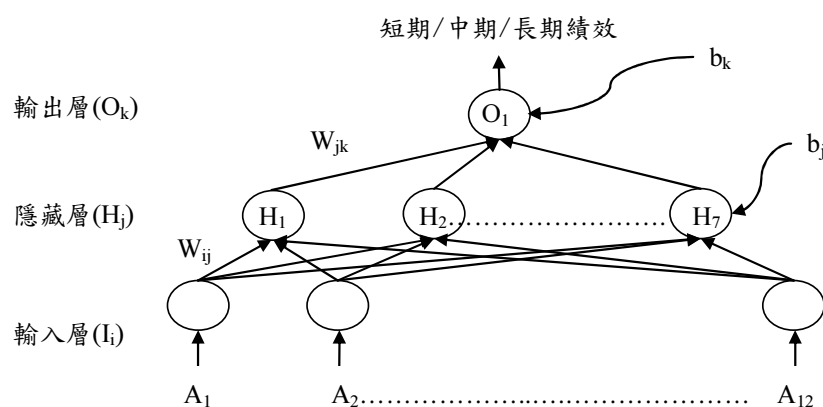


圖 5：本研究之倒傳遞類神經網路架構圖

在基因演算法方面，應用流程如圖 6 所示，共分為編碼方式、產生初始群體、計算和評估適應值、選擇與複製、交配與突變及終止條件六個步驟。第一個步驟為設定編碼方式，染色體長度設定為倒傳遞類神經網路輸入層到隱藏層之間的權重值 $84(=12*7)$ 個，加上隱藏層到輸出層之間的權重值 $7(=7*1)$ 個，再加上隱藏層與輸出層的偏差值 $8(=7+1)$ 個，一共 99 個變數組成一條染色體，為一組

可行解，並採用數值進行染色體編碼。第二步驟為產生初始族群，初始族群即為第一代染色體，也就是初始解集合，本研究採用倒傳遞類神經網路產生之十組權重作為基因演算法中的十組染色體，並另隨機產生十組染色體，共二十組染色體做為多點搜尋的初始值，即初始族群。第三步驟為計算和評估適應值，本研究希望最小化實際值與目標值之差異，因而採用三種常見的誤差衡量方法評估其差異，分別為均方誤差（mean square error; MSE）、平均絕對差（mean absolute error; MAE）及均方根誤差（root mean square error; RMSE）方式計算誤差值，以均方誤差為例，設 A_i 為第 i 筆訓練資料實際值， P_i 為其預測值， r 為訓練資料總筆數，其適應函數設定如公式(1)。

$$\text{Min} \frac{\sum_{i=1}^r (P_i - A_i)^2}{r} \quad (1)$$

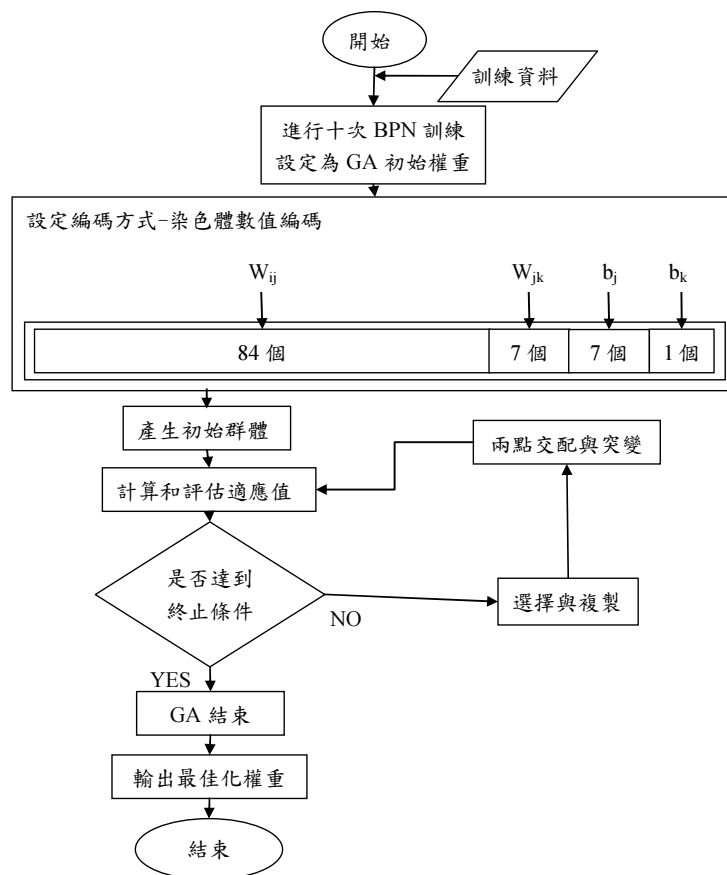


圖 6：本研究之基因演算法架構圖

第四步驟為選擇與複製，本研究染色體挑選是採用常用的輪盤法進行染色體選擇，進而將所選擇之染色體進行複製，以進行下一個世代。第五步驟為交配與突變，為了找到最適合的突變率及交配方式，進行了多組實驗測試，在突變率方面設計了十二種突變率測試，包括 0.001、0.01 及從 0.1 開始以 0.1 為間距至 1；交配方式設定為單點及兩點兩種交配方式測試。最後是終止條件，如果子代數已達到所設定之最大演化代數或適應函數已收斂，則停止運算及輸出基因演算法的最佳化權重，本研究最大演化代數分別設為 100 及 1000 代進行測試。

在自組織映射圖結合倒傳遞類神經網路方法部分，首先，在自組織映射圖群集分析中，輸入單元設定為基金屬性之 12 個變數，輸出則為分群結果。接者，針對不同群集，分別進行倒傳遞類神經網路分析，其中輸入層類神經元採用基金屬性之 12 個變數，輸出層為 1 個類神經元，分別採用短期、中期、長期基金績效，隱藏層神經元數量是採用常用的輸入層神經元數加上輸出層神經元數除以二的方式計算，因此，隱藏層神經元設定為 $7=(12+1)/2$ 個，藉由以上運算找出類神經網路最佳權重，並用以預測該群集資料。

在自組織映射圖結合基因演算法方面，先在自組織映射圖群集分析方法中，輸入單元設定為基金屬性之 12 個變數，輸出則為分群結果。接者，針對不同群集，分別進行基因演算法分析。本研究將基因演算法共分為編碼方式、產生初始群體、計算和評估適應值、選擇與複製、交配與突變及終止條件進行，相關流程及參數設定如前述基因演算法說明，如果子代數等於設定之最大子代數 1000 代或適應函數已達到收斂，則停止運算及輸出基因演算法的最佳化權重，並用以預測該群集資料。

在自組織映射圖結合倒傳遞類神經網路與基因演算法方面，先在自組織映射圖群集分析中，輸入單元設定為基金屬性之 12 個變數，輸出則為分群結果。接者，針對不同群集，分別以倒傳遞類神經網路找出十組最佳解，其中輸入層類神經元採用基金屬性之 12 個變數，輸出層為 1 個類神經元，分別採用短期、中期、長期基金績效，隱藏層神經元數量設定為 7 個。最後，以倒傳遞類神經網路找出之十組最佳解，設定為基因演算法的初始權重，基因演算法相關流程及參數設定如前述說明，找出基因演算法的最佳化權重後，並用以預測該群集資料。

在模式測試部份，本研究採用五次交叉驗證，即將資料隨機分成數量相等的五個樣本，在第一次交叉驗證中，先以第 1 個樣本作為測試資料，其餘作為訓練資料，其他樣本選取由此類推。從實驗設計五次交叉驗證中選取所需之測試資料，進行單期間混合模式之測試，藉此計算出該模式的預測正確率。

(二) 基金績效多年期軌跡預測與分析

在多年期軌跡預測與分析方面，流程如圖 4 左半部，先將資料以年度區分，

接著，進行自組織映射圖分群，透過各年度分群結果建構出多年期基金群集軌跡，且進一步利用案例式推理法進行多年期群集軌跡的相似性比對，並結合基因演算法計算各年度的權重，以追蹤群集軌跡趨勢。在模式建構部份，從實驗設計五次交叉驗證中選取所需之訓練及測試資料，其中再將訓練資料的一半設為參考資料，另一半設為訓練資料，建構多年期混合模式。

在自組織映射圖分析方面，為了分析與比較不同年度的群集變動情形，本研究以 2010 年為基準年先進行自組織映射圖群集分析，其設定方式與單期間之自組織映射圖分析相同，並採用 ANOVA 分析不同群體間的差異。接著，以基準年分析結果為基礎，推導出各年度各筆基金所屬群組，以分析出每檔基金的多年期群集軌跡，並將參考案例之多年期基金軌跡儲存到案例資料庫中，多年期軌跡推導過程概念圖如圖 7 所示。

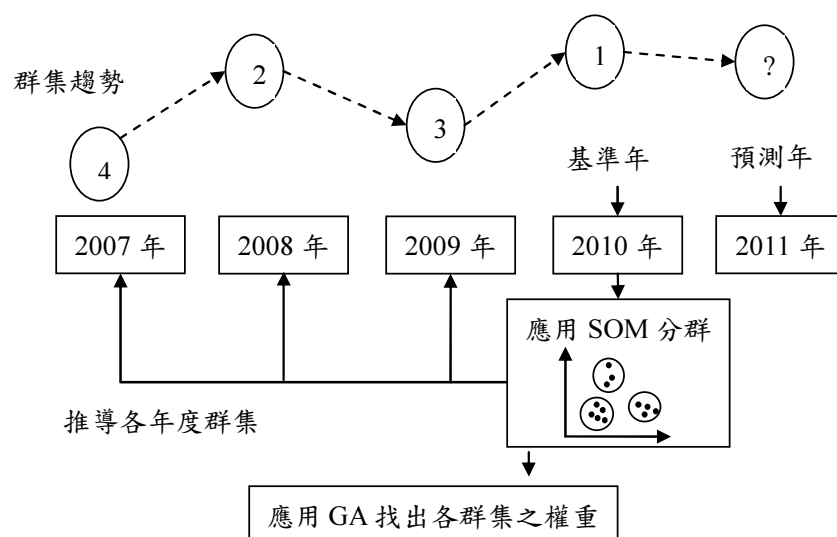


圖 7：本研究之多年期軌跡推導過程概念圖

在案例式推理方面流程，如圖 8 所示，在相似性比對方面，主要以案例距離為基礎，找出與訓練資料最小距離的案例做比較，設 w_y 為第 y 年的權重值，其中 $y=1, \dots, 4$ 分別代表 2007~2010 年四個年度，設 g_{iy} 及 g_{jy} 分別代表第 i 個訓練資料 (T_i) 與第 j 個案例資料 (R_j) 在第 y 年的群組值，案例距離表示為 $D(T_i, R_j)$ ，計算方式如式(2)所示。

$$D(T_i, R_j) = \sqrt{\sum_{y=1}^4 w_y (g_{iy} - g_{jy})^2}, y = 1, 2, 3, 4. \quad (2)$$

為了反應不同年度資料對預測年影響的大小，本研究結合案例式推理法與基因演算法，以計算各年期的權重。在基因演算法的編碼方面，染色體設定為代表 2007~2010 年在軌跡中的四個權重，並採用數值編碼。在適應函數方面，最主要的是希望最小化訓練資料與案例資料的分群差異，其餘設定如單期間基因演算法。

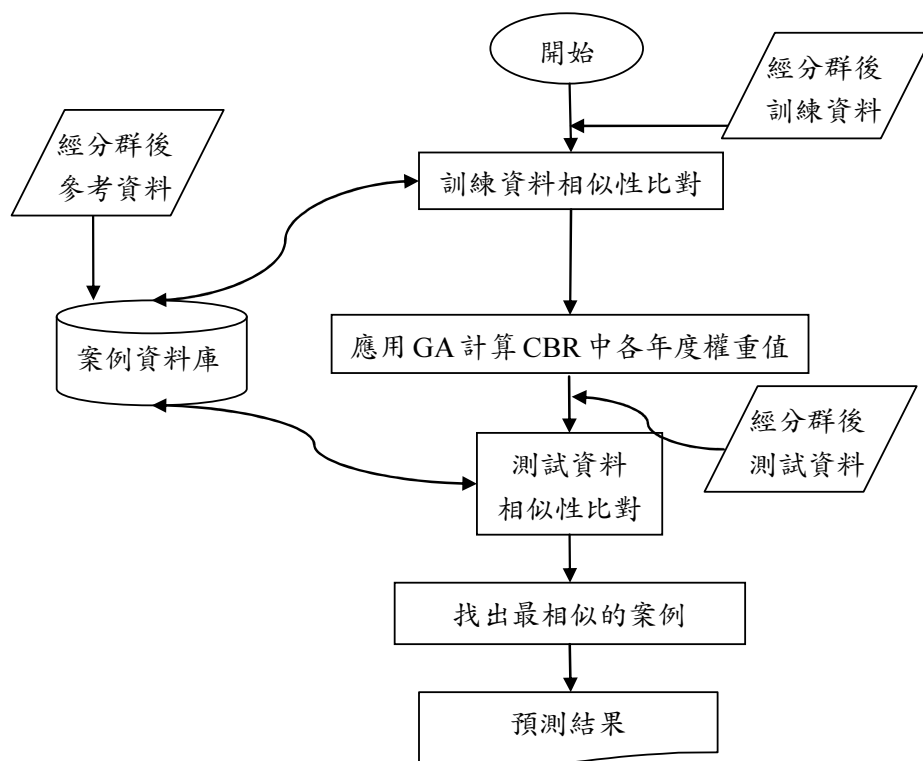


圖 8：本研究之案例式推理架構圖

肆、實證結果

本研究之研究範圍主要為國內開放型股票型基金，自 2007 年 1 月至 2011 年 12 月，共五年間之基金資料，摘要統計表如表 2 所示，其中最左邊欄為 12 項常用基金指標。第五列到十列顯示出 2008 年亦因金融海嘯關係，平均報酬率比起前一年度大幅下降，2007 至 2011 年間投資時間愈長平均報酬率愈高。在單期間基金績效預測方面，分析與比較的模式包括迴歸分析法、倒傳遞類神經網路、基因演算法、自組織映射圖結合倒傳遞類神經網路、自組織映射圖結合基因演算法、自組織映射圖結合倒傳遞類神經網路及基因演算法等六模式，進行短、中、長期基金績效比較，以找出正確率最高的預測模式，並且為避免抽樣誤差，本研究以五次交叉驗證方式，進行模式建構及測試，分析工具為 MATLAB 及 SPSS。

表 2：國內開放型股票型基金指標五年平均摘要統計表

年份	2007	2008	2009	2010	2011
基金成立年數 (年)	11.38	12.38	13.38	14.38	15.38
年化標準差	17.84	30.83	31.33	24.19	15.84
Beta 值	0.86	1.01	0.81	0.98	0.74
季報酬率 (%)	6.25	-13.43	12.42	1.95	-4.27
半年報酬率 (%)	17.25	-22.02	18.13	3.96	-4.15
一年報酬率 (%)	30.44	-22.62	3.04	21.41	1.60
二年報酬率 (%)	68.00	-1.22	-27.19	19.70	24.63
三年報酬率 (%)	88.63	29.02	-5.48	-12.66	17.81
五年報酬率 (%)	110.34	70.34	36.87	45.14	14.17
淨資產 (千元)	36128.68	28249.32	25835.52	26724.36	23807.4
年週轉率 (%)	285.63	267.81	277.23	260.31	247.41
總費用率 (%)	0.14	0.15	0.14	0.14	0.14

在倒傳遞類神經網路部分，為了找到最適合國內開放型股票型基金績效分析的學習率，本研究將學習率設定為 0.1 至 1，以 0.1 為間距進行十次實驗。結果顯示倒傳遞類神經網路學習率設定為 0.5 時，MSE 為 14.76%，MAE 為 26.76%，RMSE 為 38.67%，錯誤率達到最低，如表 3 所示。因此，本研究將倒傳遞類神經網路學習率設定為 0.5。

表 3：應用不同學習率於倒傳遞類神經網路之結果

學習率	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
MSE	15.76%	15.56%	15.76%	15.62%	14.76%	19.89%	15.25%	16.34%	15.44%	15.41%
MAE	28.44%	27.70%	28.87%	28.35%	26.76%	30.63%	27.89%	29.43%	27.07%	28.32%
RMSE	39.70%	39.45%	39.70%	39.52%	38.67%	44.60%	39.05%	40.42%	39.30%	39.25%

在基因演算法方面，為了找到最適合的交配方式、突變率及演化代數，本研究設計了十二種突變率測試，包括 0.001、0.01 及自 0.1 開始以 0.1 為間距測試至 1；交配方式設定為單點及兩點兩種交配方式進行實驗；最大演化代數設定為常見的 100 及 1000 代測試，結果顯示演化至 1000 代時，實驗結果已趨於穩定，整體而言以兩點交配突變率設為 0.1，且最大演化代數設定為 1000 代時，錯誤率 (MSE) 達到最低 22.66%，如表 4 及表 5 所示。因此，本研究將基因演算法交配方式設為兩點交配，突變率設為 0.1，最大演化代數設定為 1000 代。

表 4：基因演算法演化代數設為 100 代時不同交配方式、突變率之測試結果

突變率	單點交配			兩點交配		
	MSE	MAE	RMSE	MSE	MAE	RMSE
0.001	46.30%	49.78%	66.90%	40.16%	47.07%	61.43%
0.01	36.82%	43.32%	58.40%	35.78%	42.56%	57.46%
0.1	29.43%	42.66%	53.21%	30.63%	39.78%	53.61%
0.2	35.31%	42.46%	57.51%	30.36%	44.60%	53.64%
0.3	34.32%	43.56%	56.81%	35.78%	45.34%	57.65%
0.4	36.20%	43.73%	57.86%	33.59%	43.59%	55.88%
0.5	39.06%	48.77%	60.50%	33.59%	42.68%	56.33%
0.6	38.96%	47.29%	60.51%	40.89%	47.81%	62.13%
0.7	37.97%	47.06%	59.36%	34.90%	44.92%	56.95%
0.8	39.84%	49.53%	61.19%	35.05%	44.69%	57.95%
0.9	36.98%	51.88%	61.25%	37.50%	46.22%	59.09%
1	40.52%	52.68%	62.15%	40.89%	47.26%	62.08%

表 5：基因演算法演化代數設為 1000 代時不同交配方式、突變率之測試結果

突變率	單點交配			兩點交配		
	MSE	MAE	RMSE	MSE	MAE	RMSE
0.001	42.03%	47.49%	63.06%	38.88%	47.16%	60.29%
0.01	23.45%	35.02%	21.93%	23.45%	34.59%	48.17%
0.1	23.47%	41.32%	48.45%	22.66%	40.52%	47.60%
0.2	27.73%	41.49%	48.71%	24.56%	42.22%	49.55%
0.3	25.17%	43.05%	50.17%	23.21%	41.06%	48.18%
0.4	26.78%	44.45%	51.75%	26.78%	44.62%	51.75%
0.5	27.21%	45.06%	52.16%	25.09%	42.83%	50.09%
0.6	32.02%	46.69%	56.59%	25.91%	44.06%	50.90%
0.7	28.85%	46.42%	53.71%	25.69%	43.89%	50.69%
0.8	28.64%	46.77%	53.52%	29.93%	47.62%	54.71%
0.9	33.16%	50.36%	57.59%	28.29%	46%	53.19%
1	35.63%	51.91%	59.69%	29.57%	47.73%	54.38%

在自組織映射圖結合倒傳遞類神經網路及基因演算法方面，本研究為了找尋最適合評估基金績效的分群數，分別從三群測試到十群，因分群數太多時，較難分析多年期群體軌跡，故本研究最高測試分群數定為十群，網格設定為 $n*1$ ，其 n 代表分群數。其中當分群數設定為四群時，預測結果最佳，實驗結果如表 6（以 MSE 評估結果為例）。因此，本研究在自組織映射圖相關模式中，皆以自組織映射圖群集數設定為 4 群，進行模式分析與比較。

表 6：各分群數預測結果 (MSE)

分群數	三	四	五	六	七	八	九	十
短期績效	15.33%	13.24%	15.96%	17.88%	17.41%	16.84%	16.04%	19.21%
中期績效	13.74%	13.70%	14.00%	15.36%	16.57%	15.66%	16.89%	19.19%
長期績效	14.04%	13.21%	14.71%	15.54%	16.23%	15.55%	15.22%	15.50%

以六種模式預測基金短、中及長期績效之結果比較表，如表 7 所示，前三種模式迴歸分析法 (RA)、倒傳遞類神經網路 (BPN) 及基因演算法 (GA) 是執行單一方法的實證結果，為了與混合模式比較，GA 單一方法部份是以隨機方式產生二十組染色體作為初始族群。結果顯示基金績效無論在短、中及長期績效預測結果，皆是以本研究所提出結合自組織映射圖、倒傳遞類神經網路及基因演算法之混合模式表現最好，短、中及長期績效預測錯誤率分別為 13.24%、13.70% 及 13.21%。若以單一演算法比較，短期績效以迴歸分析法的預測錯誤率最低；中及長期績效以倒傳遞類神經網路表現最好，而基因演算法預測結果最不理想。最後，三種混合型模式的預測結果均比原單一演算法模式好。

表 7：單期間基金績效預測六種模式比較表 (MSE)

模型	短期績效(MSE)	中期績效(MSE)	長期績效(MSE)
RA	18.12%	16.98%	16.52%
BPN	19.38%	14.17%	15.66%
GA	22.69%	26.77%	23.71%
SOM+BPN	16.28%	14.01%	14.96%
SOM+GA	19.35%	19.87%	19.28%
SOM+BPN+GA	13.24%	13.7%	13.21%

在多年期軌跡分析方面，本研究先針對基準年進行自組織映射圖分析，群組

數設定為四群，並針對基準年分群結果進行 ANOVA 檢定，藉此分析出不同群組間的主要差異，ANOVA 檢定結果如表 8 所示，本研究設定顯著水準為 0.05，由表中可知，分群結果四群組在年化標準差、Beta 值、二年報酬率、三年報酬率及年週轉率五個變數上呈現顯著性差異（以*號表示）。第一群的基金數量為 341 筆，第二群為 492 筆，第三群為 352 筆，第四群為 351 筆。各群之間以代表風險高低的 Beta 值與年化標準差群集代表性最為明顯，第一群代表風險最高的，其 Beta 值與年化標準差為最大；其餘依次為第二、三、四群。在二年報酬率及三年報酬率方面，不同群組間變動較大，並未發現明顯的區隔，但存在著一些特殊現象，例如，第二群的二年報酬率最低，卻在三年報酬率反轉成最高；第三群的兩年報酬率最高，三年報酬率卻反轉成最低；另外，值得注意的是風險高的第一群與第二群，呈現出兩種不同的情況，第一群的風險是最高的，在二年與三年報酬率為次高，呈現較穩定的情況；反之，第二群的風險是次高的，在二年報酬率雖是最低，卻在三年報酬率反轉為最高，報酬率波動較大。

表 8：基準年 ANOVA 分析結果

變數名稱	群	平均數	標準差	標準誤	變數名稱	群	平均數	標準差	標準誤
基金成立年	1	0.66	0.372	0.0201	二年報酬率*	1	0.626	0.3518	0.0191
	2	0.577	0.3949	0.0178		2	0.2	0.2554	0.0115
	3	0.359	0.3592	0.0191		3	0.901	0.2202	0.0117
	4	0.463	0.4376	0.0234		4	0.416	0.3901	0.0208
年化標準差*	1	0.705	0.2493	0.0135	三年報酬率*	1	0.566	0.2934	0.0159
	2	0.641	0.2852	0.0149		2	0.935	0.1743	0.0079
	3	0.598	0.2023	0.0108		3	0.142	0.2258	0.012
	4	0.171	0.2405	0.0128		4	0.093	0.2121	0.0113
Beta值*	1	0.933	0.1711	0.0093	五年報酬率	1	0.543	0.3723	0.0202
	2	0.827	0.2987	0.0165		2	0.417	0.3737	0.0168
	3	0.783	0.2511	0.0134		3	0.625	0.3561	0.019
	4	0.161	0.2339	0.0125		4	0.588	0.4384	0.0234
季報酬率	1	0.556	0.3676	0.0199	淨資產	1	0.713	0.2648	0.0143
	2	0.461	0.337	0.0152		2	0.655	0.3218	0.0145
	3	0.597	0.3212	0.0171		3	0.689	0.2625	0.014
	4	0.61	0.2953	0.0158		4	0.711	0.2586	0.0138
半年報酬率	1	0.551	0.2593	0.014	年週轉率*	1	0.117	0.2122	0.0115
	2	0.487	0.2714	0.0122		2	0.074	0.1779	0.008
	3	0.57	0.2671	0.0142		3	0.175	0.2669	0.0142
	4	0.594	0.264	0.0141		4	0.236	0.2719	0.0145
一年報酬率	1	0.575	0.3508	0.019	總費用率	1	0.513	0.293	0.0159
	2	0.431	0.3657	0.0165		2	0.325	0.306	0.0138
	3	0.631	0.329	0.0175		3	0.668	0.3359	0.0179
	4	0.54	0.3643	0.0194		4	0.452	0.3691	0.0197

註：符號 "*" 表示達 0.05 顯著水準

為了讓各年度的分群基礎俱一致性（如第一群代表風險最高的），再以基準年自組織映射圖分群結果，推算出前三年度的分群情形，並結合案例式推理法進行相似性比對，及利用基因演算法計算各年期之權重，以預測群集軌跡。本研究以 2010 年為基準年 2011 年為預測年，以 2007 至 2010 年四年期群集資料，預測 2011 年群組趨勢，利用基因演算法計算出 2007 至 2010 年度權重分別為 0.421、0.377、0.371 及 0.264，再由此權重及案例式推理法找出的最相似群集趨勢，計算出預測年基金所屬群集。接著，再將多年期群集預測結果，進行 ANOVA 分析，分析結果如表 9 所示，分群結果在 Beta 值、二年報酬率及三年報酬率三個變數上呈現顯著差異。在風險特徵部份，在 Beta 值部份預測結果維持了相同的群集特徵，即第一群均代表風險最高的，其他依次為二三四群。唯在年化標準差部份，群集特徵不顯著。

表 9：多年期軌跡預測年群集資料之 ANOVA 結果

變數名稱	基準年	單期間	多年期軌跡	變數名稱	基準年	單期間	多年期軌跡
基金成立年				二年報酬率	*		*
年化標準差	*			三年報酬率	*	*	*
Beta值	*		* v	五年報酬率			
季報酬率				淨資產			
半年報酬率				年週轉率	*		
一年報酬率				總費用率			

註：符號 "*" 表示達 0.05 顯著水準；符號 "v" 表示具有相同對應群組

為了能夠比較多年期軌跡與單期間群集預測的差異，本研究進一步針對多年期及單期間群集預測結果進行比較分析，在單期間群集預測結果部份，ANOVA 分析結果如表 9 所示，分群結果僅在三年報酬率變數上呈現顯著性差異，完全呈現不出風險特徵差異，故無法作為風險趨勢預測模式。由上述結果分析，可以明顯看出多年期軌跡預測在風險趨勢預測效果優於單期間預測分析結果。在報酬率部份，雖然基準年、單期間與多年期軌跡分析結果在三年報酬率上皆呈現顯著，但因無法維持相同的群集特徵（例如，基準年第二群代表平均報酬率最高，但在單期間分析結果第四群代表平均報酬率最高），故無法預測報酬率的群集趨勢。

報酬率與風險常是投資決策的最主要考量因素，投資者可依其投資時間偏好，利用本研究提出的單期間分析方法，選擇預測某基金的短期、中期、長期報酬率，並可利用多年期群集軌跡分析，預測該基金的多年期風險趨勢，以作為投資決策的參考。

伍、結論與建議

本研究以自組織映射圖結合倒傳遞類神經網路及基因演算法協助不同投資時間長短偏好的投資者，找出單期間較佳短期、中期、長期績效的基金，並以自組織映射圖結合案例式推理法及基因演算法進行多年期軌跡的預測分析，作為投資決策的參考。研究結果顯示無論在短、中及長期績效預測結果，皆是當自組織映射圖設定為四群，並結合倒傳遞類神經網路及基因演算法之模式預測結果最好。若以單一方法比較，短期績效以迴歸分析法的預測錯誤率最低；中及長期績效以倒傳遞類神經網路表現最好。而三種混合模式的預測結果均比原單一模式好，特別是以本研究所提出之自組織映射圖結合倒傳遞類神經網路及基因演算法的模式，預測錯誤率最低。在多年期軌跡的預測分析方面，亦顯示出本研究提出的多年期軌跡預測方法，在風險趨勢預測效果上，優於單期間預測分析結果。投資者可依其投資時間偏好，利用本研究提出的單期間分析方法，選擇預測某基金的短期、中期、長期報酬率，並可利用多年期群集軌跡分析，預測該基金的多年期風險趨勢，以作為投資決策的參考。此外，在本研究中，分為四群之共同基金，雖然在報酬率上未顯示出顯著的分群差異，但在 Beta 值及年化標準差等風險指標上，顯示出顯著的區隔，第一群代表風險最高的，其 Beta 值與年化標準差為最大；其餘依次為第二、三、四群，因此，依風險指標來看，分群效果是明顯而穩定的，分群結果可供對風險敏感度高的使用者參考。

本研究之限制在於對於新發行的基金只能預測其短、中及長期績效，無法看出多年期軌跡的風險趨勢。此外，本研究擷取資料範圍雖為 2007 年至 2011 年共五年股票型基金，但因參考變數中已納入五年報酬率，因此，實際資料採用期間為 2002 年至 2011 年十年間資料，但因超過十年間的基金資料，常因多檔基金中途的新進或退出市場，導致資料完整性較不足，故本研究僅採用五年內之股票型基金資料，因此可能有測試期間過短問題，以致研究架構的長期穩定性較難以驗證。

在未來研究建議部份，第一，目前本研究資料來源只限於國內開放型股票型基金，建議未來可針對其他類型基金資料進行分析，並可進一步結合支持向量機、粒子群演算法等其他人工智慧方法，找出短、中及長期基金績效的最佳預測方法；第二，本研究資料範圍為 2007 年 1 月至 2011 年 12 月 5 年內的股票型基金資料，並未排除期間中清算合併的基金，是否會有存續者偏誤的問題，亦可進一步探討；第三，本研究涵蓋的資訊量仍不足，未來可搭配基金新進和退場的處理研究，蒐集較長時間的資料量，並搭配 Sliding Window 進行多年期軌跡預測與分析；第四，本研究所採用變數均採三區間離散化，針對其他不同區間數所造成的影響，亦可進一步進行探討；最後，投資問題在預測的錯誤率上，往往是單邊

的，也就是常重視預測誤差帶來的虧損，而較忽略誤差帶來的獲利，因此，在未來研究建議中，可加入單邊誤差的計算與比較，以增加誤差衡量方式在問題領域上之適切性。

誌謝

本研究受科技部計畫 NSC 101-2410-H-239-005-MY2 補助，特此致謝。

參考文獻

- 尹相志 (2009), *SQL Server 2008 Data Mining 資料採礦*, 悅知文化, 台北。
- 邱永和、陳玉涓、陳素緞、陳凱夫 (2008), 『國內共同基金之績效評估』, *Soochow Journal of Accounting 會計學報*, 第一卷, 第一期, 頁 29-52。
- 李顯儀, 李欣微、李亮君 (2011), 『共同基金投資集中度與績效關聯性之研究』, *管理科學研究*, 第七卷, 第二期, 頁 49-62。
- 馬麗菁、李欣頤、邱惠君、葉曉薇、周家鈺 (2011), 『結合基因演算法與類神經網路於國內共同基金績效預測之研究』, *第十七屆資訊管理暨實務研討會論文集*。
- 蕭文龍 (2009), *多變量分析最佳入門實用書—SPSS+LISREL*, 第二版, 基峰資訊, 台北。
- Akarachai, A. and Daricha, S. (2007), 'Avoiding Local Minima in Feedforward Neural Networks by Simultaneous Learning', *Australian Conference on Artificial Intelligence*, Springer, Berlin Heidelberg, Vol. 4830, pp. 100-109.
- Antonio, A.R., Victor, M.G. and Jose, L.M. (2011), 'Optimizing stochastic production-inventory systems: A heuristic based on simulation and regression analysis', *European Journal of Operational Research*, Vol. 213, No. 1, 2011, pp. 107-118.
- Babalos, V., Philippas, N., Doumpos, N. and Zopounidis, C. (2012), 'Mutual funds performance appraisal using stochastic multicriteria acceptability analysis', *Applied Mathematics and Computation*, Vol. 218, No. 9, pp. 5693-5703.
- Badrinath, S.G. and Gubellini, S. (2011), 'On the characteristics and performance of long-short, market-neutral and bear mutual funds', *Journal of Banking & Finance*, Vol. 35, No. 7, pp. 1762-1776.
- Bauer, R., Koedijk, K. and Otten, R. (2005), 'International evidence on ethical mutual fund performance and investment style', *Journal of Banking & Finance*, Vol. 29, No. 7, pp. 1751-1767.
- Berardi, V.L., Patuwo, B.E and Hu, M.Y. (2004), 'A principled approach for building

- and evaluating neural network classification models', *Decision Support Systems*, Vol. 38, No. 2, pp. 233-246.
- Bianco, V., Manca, O. and Nardini, S. (2009), 'Electricity consumption forecasting in Italy using linear regression models', *Energy*, Vol. 34, No. 9, pp. 1413-1421.
- Carhart, M. (1997), 'On persistence in mutual fund performance', *Journal of Finance*, Vol. 52, pp. 57-82.
- Carlos, S.C. (1996), 'Self organizing neural networks for financial diagnosis', *Decision Support Systems*, Vol. 17, No. 3, pp. 227-238.
- Chang, P.C., Lai, C.Y. and Robert, K.L. (2006), 'A hybrid system by evolving case-based reasoning with genetic algorithm in wholesaler's returning book forecasting', *Decision Support Systems*, Vol. 42, No. 3, pp. 1715-1729.
- Chan, K. and Covrig, V. (2012), 'What determines mutual fund trading in foreign stocks?', *Journal of International Money and Finance*, Vol. 31, No. 4, pp. 793-817.
- Chen, C.S., Chen, B.P., Chou, F.N. and Yang, C.C. (2010), 'Development and application of a decision group Back-Propagation Neural Network for flood forecasting', *Journal of Hydrology*, Vol. 385, No. 1, pp. 173-182.
- Chiu, C. (2002), 'A case-based customer classification approach for direct marketing', *Expert Systems with Applications*, Vol. 22, No. 2, pp. 163-168.
- Gan, Q. (2012), 'Exponential synchronization of stochastic Cohen-Grossberg neural networks with mixed time-varying delays and reaction-diffusion via periodically intermittent control', *Neural Networks*, Vol. 31, pp. 12-21.
- Herrmann, U. and Scholz, H. (2013), 'Short-term persistence in hybrid mutual fund performance: The role of style-shifting abilities', *Journal of Banking & Finance*, Vol. 37, No. 7, pp. 2314-2328.
- Holland, J. (1975), *Adaptation in Natural and Artificial System*, University of Michigan Press, Ann Arbor, MI.
- Hsu, C.M. (2011), 'A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming', *Expert Systems with Applications*, Vol. 38, No. 11, pp. 14026-14036.
- Indro, D.C., Jiang, C.X., Patuwo, B.E. and Zhang, G.P. (1999), 'Predicting mutual fund performance using artificial neural networks', *Omega*, Vol. 27, No. 3, pp. 373-380.
- Jain, A. and Avadhnam, M.K. (2007), 'Hybrid neural network models for hydrologic time series forecasting', *Applied Soft Computing*, Vol. 7, No. 2, pp. 585-592.
- Jardin, P.D. and Severin, E. (2011), 'Predicting corporate bankruptcy using a self-organizing map: an empirical study to improve the forecasting horizon of a

- financial failure model', *Decision Support Systems*, Vol. 51, No. 3, pp. 701-711.
- Javier, V.G. (2013), 'The persistence of European mutual fund performance', *Research in International Business and Finance*, Vol. 28, pp. 45-67.
- Kohonen, T. (1990), 'The self-organizing map', *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1481-1480.
- Kuo, R.J. and Lin, L.M. (2010), 'Application of a hybrid of genetic algorithm and particle swarm optimization algorithm for order clustering', *Decision Support Systems*, Vol. 49, No. 4, pp. 451-462.
- Lai, L.L. and Lau, S.H. (2010), 'Evaluating mutual fund performance in an emerging Asian economy: The Malaysian experience', *Journal of Asian Economics*, Vol. 21, No. 4, pp. 378-390.
- Law, R. (2000), 'Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting', *Tourism Management*, Vol. 21, No. 4, pp. 331-340.
- Leigh, W., Purvis, R. and Ragusa, J.M. (2002), 'Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support', *Decision Support Systems*, Vol. 32, No. 4, pp. 361-377.
- Li, S.H. and Kuo, S.C. (2008), 'Knowledge discovery in financial investment for forecasting and trading strategy through wavelet-based SOM networks', *Expert Systems with Applications*, Vol. 34, No. 2, pp. 935-951.
- Ng, S.T., Skitmore, M. and Wong, K.F. (2008), 'Using genetic algorithms and linear regression analysis for private housing demand forecast', *Building and Environment*, Vol. 43, No. 6, pp. 1171-1184.
- Prather, L., Bertin, W.J. and Henker, T. (2004), 'Mutual fund characteristics, managerial attributes, and fund performance', *Review of Financial Economics*, Vol. 13, No. 4, pp. 305-326.
- Renner, G. and Ekart, A. (2003), 'Genetic algorithms in computer aided design', *Computer-Aided Design*, Vol. 35, No. 8, pp. 709-726.
- Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986), 'Learning representations by back propagating errors', *Nature*, pp. 533-536.
- Sexton, R.S. and Dorsey, R.E. (2000), 'Reliable classification using neural networks: a genetic algorithm and backpropagation comparison', *Decision Support Systems*, Vol. 30, No. 1, pp. 11-22.
- Shu, P.G., Yeh, Y.H. and Yamada, T. (2002), 'The behavior of Taiwan mutual fund

- investors-performance and fund flows', *Pacific-Basin Finance Journal*, Vol. 10, No.5, pp. 583-600.
- Tan, P.N., Steinbach, M. and Kumar, V. (2005), *Introduction to Data Mining*, Addison-Wesley, Boston, MA, USA.
- Thomassey, S. and Fiordaliso, A. (2006), 'A hybrid sales forecasting system based on clustering and decision trees', *Decision Support Systems*, Vol. 42, No. 1, pp. 408-421.
- Tsai, C.F. and Chen, M.L. (2010), 'Credit rating by hybrid machine learning techniques', *Applied Soft Computing*, Vol.10, No. 2, pp. 374-380.
- Tsai, T.J., Yang, C.B. and Peng, Y.H. (2011), 'Genetic algorithms for the investment of the mutual fund with global trend indicator', *Expert Systems with Applications*, Vol. 38, No. 3, pp. 1697-1701.
- Tsang, P.M., Kwok, P., Choy, S.O., Kwan, R., Ng, S.C., Mak, J., Tsang, J., Koong, K. and Wong, T.L. (2007), 'Design and implementation of NN5 for Hong Kong stock price forecasting', *Engineering Applications of Artificial Intelligence*, Vol. 20, No. 4, pp. 453-461.
- Wang, K. and Huang, S. (2010), 'Using fast adaptive neural network classifier for mutual fund performance evaluation', *Expert Systems with Applications*, Vol. 37, No. 8, pp. 6007-6011.
- Yan, X. (2008), 'Liquidity, Investment Style and the Relation between Fund Size and Fund Performance', *Journal of Financial and Quantitative Analysis*, Vol. 43, No. 3, pp. 741-768.
- Zhuang, Z.Y., Wilkin, C.L. and Ceglowski, A. (2013), 'A framework for an intelligent decision support system: A case in pathology test ordering', *Decision Support Systems*, Vol. 55, No. 2, pp. 476-487.

