

王慶豐、李俊賢 (2019), 『複數模糊類神經於趨向類別預測之研究』, 中華民國資訊管理學報, 第二十六卷, 第四期, 頁 413-450。

複數模糊類神經於趨向類別預測之研究

王慶豐

國立中央大學資訊管理學系

李俊賢*

國立中央大學資訊管理學系

摘要

面對現今的國際化環境，投資已成為許多企業與人們的獲利方式，股票為此領域相當普遍的交易模式，但股價波動所牽涉層面廣泛，固然難以估計與預測，人工智慧中的深度學習即為當今預測的最佳工具之一。本研究提出一種新形態之複數模糊類神經分類模型 (Complex Neuro-Fuzzy Classification Model; CNFC)，以減法分群演算法 (Subtractive Clustering Algorithm; SCA) 識別資料趨向類別，輔助模型進行動態式分類預測，其中採用粒子群最佳化演算法 (Particle Swarm Optimization; PSO) 與遞迴最小平方法 (Recursive Least Squares Estimator; RLSE) 為複合式最佳化演算法 (Hybrid optimization algorithm)，針對模型不同部分的參數進行優化，將有效提升模型優化效率。透過複數高斯模糊集合的特性，模糊化輸入資料的類別隸屬程度，更加精確描述類別值域，增強模型的預測及應用能力。實驗透過重複性與文獻多樣化模型比較結果，驗證 CNFC 的預測效能與 PSO-RLSE 的最佳化成效於股價時間序列資料具有較佳能力。

關鍵詞：複數模糊類神經、複數模糊集、特徵選取、粒子群最佳化演算法、遞迴最小平方法

* 本文通訊作者。電子郵件信箱：jamesli@mgt.ncu.edu.tw
2019/04/12 投稿；2019/07/22 修訂；2019/08/24 接受

Wang, C.F. and Li, C. (2019), 'Prediction of directional classification using complex neuro-fuzzy model', *Journal of Information Management*, Vol. 26, No. 4, pp. 413-450.

Prediction of Directional Classification Using Complex Neuro-Fuzzy Model

Ching-Feng Wang

Department of Information Management, National Central University

Chunshien Li*

Department of Information Management, National Central University

Abstract

Purpose—Facing the current international environment, investment has become a way of profit for many businesses and people, stocks are a common method of trading in this area, but stock price fluctuations have a wide range of influences, it is difficult to estimate and forecast. Deep learning in Artificial Intelligence (AI) is one of the best tools for current prediction.

Design/methodology/approach—This study proposes a novel Complex Neuro-Fuzzy Classification Model (CNFC), identifying the directional classification of data by Subtractive Clustering Algorithm (SCA) and assisting models for dynamic classification prediction. The model uses Particle Swarm Optimization (PSO) and Recursive Least Squares Estimator (RLSE) as the hybrid optimization algorithm for parameters optimization of different parts of the model will effectively improve the efficiency of model optimization.

Findings—The experiments to verify the predictive performance of CNFC and the optimization effect of PSO-RLSE through repetitive and literature diversification models, which has better ability in stock price time series data.

Research limitations/implications—In this study, the class degree of input data is fuzzified through the characteristics of complex Gaussian fuzzy sets, which more

* Corresponding author. Email: jamesli@mgt.ncu.edu.tw

2019/04/12 received; 2019/07/22 revised; 2019/08/24 accepted

accurately describes the class value and enhances the prediction and application ability of the model. In the future, we plan to combine the famous classifier (e.g., SVM, Softmax) with the CNFC model.

Practical implications — In this study, we provide an innovative stock price forecasting model that can be used as an auxiliary investment tool for investors, and financial practitioners can further explore the relationship between stock prices and the overall economy.

Originality/value — This study is the first attempt to optimize the parameters of the CNFC model using PSO-RLSE hybrid optimization algorithm.

Keywords: complex neuro-fuzzy, complex fuzzy set, feature selection, particle swarm optimization, recursive least squares estimator

壹、緒論

股價指數由早期發展至今，已成為多數人們用於控管風險與獲利的工具之一，其價格走勢直接影響投資者的利益關係，更大程度可以表示國家整體的經濟水平。不過影響股價波動之因素繁多，舉凡政治、物價或匯率等，故欲預測股價並協助建構投資策略的行為便難以實現。早期許多研究學者認為股價即為布朗運動 (Brownian motion)，該概念最早由法國一位數學家所提出，在數學領域稱為隨機漫步理論 (Random walk theory)，核心理念認為證券價格波動為一隨機形式，有如行走於空曠場地的人們，下一步的走向皆是未知，沒有規律存在，價格運動方向可能因為一件小事引發大幅震盪，否定其預測的可能性。然而，股價資料屬於時間序列 (Time series) 型態，表示一組由時間的先後順序所形成的數據點，每個數據點的觀察時距一般是固定的。

隨著科技不斷突破與創新，人工智慧技術亦隨之興起，其概念由 John McCarthy 於西元 1955 年所提出。直至西元 1982 年，Engle (1982) 教授提出了 ARCH 模型 (Autoregressive Conditional Heteroskedasticity Model; ARCH)，此模型在當時為非常重大的創新方法，因能夠更好地解決股價波動之特徵問題，大幅提升預測金融市場的可能性。隨後，許多人開始投入股市預測相關研究，並運用機器學習或深度學習等人工智慧方法進行預測，包括分類 (Classification)、聚類 (Clustering)、迴歸 (Regression) 與關聯規則 (Association) 等，每種方法皆具有各自發展之人工智慧模型。Bollen、Mao 與 Zeng (2011) 應用自組織模糊類神經網路 (Self-Organizing Fuzzy Neural Network; SOFNN) 及格蘭傑因果關係檢驗 (Granger causality analysis) 研究社群網路公眾情緒假設，並預測美國道瓊工業指數。Dimpfl 與 Jank (2016) 研究股價波動與散戶對於股市關注度之關聯性，採用自回歸模型 (Autoregressive model) 進行預測驗證。Patel 等 (2015) 將人工神經網路 (Artificial Neural Network; ANN)、支援向量機 (Support Vector Machine; SVM)、隨機森林 (Random forest) 與朴素貝葉斯分類器 (Naive Bayes classifier) 應用於印度股價市場之預測。

在本研究中，將先以夏農熵 (Shannon entropy) (Shannon & Weaver 1949) 作為基礎，設計一套特徵萃取算法，提取股價影響特徵資訊，挑選重要特徵以提升股價預測能力，並建構複數模糊類神經分類模型 (Complex Neuro-Fuzzy Classification Model; CNFC) 透過特徵進行股價趨勢預測。現今的股價預測研究，廣義可區分為精準預測與分類預測，本研究目標為將上述預測方式結合成為一種新穎的模型架構，由精準預測的複數模糊類神經模型 (Complex neuro-fuzzy model) 與模糊分類器 (Fuzzy classifier) 合併為一複數模糊類神經分類模型，此

模型可於精準預測後轉換為股價趨向的類別預測，不同於以往人工智慧模型的應用模式。在股價資料部分，採用了減法分群演算法 (Subtractive Clustering Algorithm; SCA) 建構股價各屬性波動之分群資料，其特性可以使用相同分群影響範圍情況下分析，將目標資料轉換為趨向類別型態，動態式預測輸入資料的股價運動方向，使得模型預測足以獲得較佳之成效。透過運用夏農熵作為計算資訊混亂程度指標，獲得每一特徵對於目標之影響程度，除了往後可以提升模型的計算效率之外，更可避免冗餘資料進入模型進而影響模型準確率的可能性。同時，模型加入機器學習之參數最佳化演算法過程以提升模型精確率，以粒子群最佳化演算法 (Particle Swarm Optimization; PSO) (Eberhart & Kennedy 1995) 結合遞迴最小平方估計演算法 (Recursive Least Squares Estimator; RLSE)，形成 PSO-RLSE 複合式最佳化演算法 (Hybrid optimization algorithm) 進行參數最佳化流程，作為主要調整複數模糊類神經分類模型參數的方法。

貳、文獻探討

本章節將依序簡介研究中所涉及的理論方法，包括針對預測資料之分群演算法；探討對於原始資料的特徵選取方法，以及運用於模糊類神經模型的模糊集隸屬程度，最後為最佳化人工智慧模型的機器學習演算法。

一、分群演算法

分群為一種將資料分類成群的技術，原始資料並無任何標記與特徵，屬於非監督式的學習方法，目的在於找尋資料中屬性相似的數據，各群之間相較於同一群體的數據較無相似性，直至今日的分群演算法有許多種類，並有著各自不同的特性。其中較傳統的分群演算法如減法分群演算法 (Subtractive Clustering Algorithm; SCA)，此為一種快速且一次性的分群方式，最初是由山峰聚類演算法 (Mountain clustering) (Yager & Filev 1994) 所延伸而來，假設每一個數據點皆為潛在的群體中心，進而計算該數據點周圍的數據點密度，依此判斷數據點成為群體中心的可能性。模糊 C 均值聚類演算法 (Fuzzy C-Means Clustering Algorithm; FCM) (Dunn 1973; Bezdek 1974; Bezdek et al. 1984) 則運用模糊理論計算數據屬於類別的程度，且以不斷優化目標函數的方式，達成自動分群之目的，隨後 Bezdek (2013) 將模糊 C 均值聚類演算法進行數據於隸屬程度計算之改進，Zheng 等 (2015) 則解決了其魯棒性與目標函數之歐幾里得距離對於離群值的敏感問題。

二、特徵選取

在機器學習領域中，特徵選取 (Feature Selection; FS) 亦可稱為變量選取或屬性選取。目的為搜尋原始資料集中重要的關鍵訊息，並移除冗餘與相關程度較低的資料，故特徵挑選的標準制定至關重要。現今的特徵選取演算法可以劃分為三種類別 (Guyon & Elisseeff 2003)，分別為包裝類、過濾類與嵌入類。

包裝類為運用多個特徵子集合不斷重複進行模型訓練，以驗證資料集計算預測結果，其缺點為計算量過於龐大，但能夠有效搜尋特定模型的最佳特徵子集合，遞歸特徵消除 (Recursive Feature Elimination; REF) 即運用機器學習模型多次訓練，每次移除多個特徵，再藉由新的特徵子集合進行下一輪訓練，由此獲得最佳特徵。Guyon 等 (2002) 提出基於遞規特徵消除之支援向量機 (Support Vector Machine; SVM) 方法消除冗餘資訊。過濾類為針對資料的每個特徵進行評分，相較於包裝類計算特徵子集合之方式，不容易獲得對於特定模型之最佳特徵子集合，卻可大幅減少運算量，此類所獲得的特徵同時亦具有較高的通用性，如卡方檢驗定 (Pearson 1900) 透過觀察值與理論值的誤差估計，進而挑選有效特徵，Shannon 與 Weaver (1949) 將熱力學的熵，引入至資訊理論當中，作為計算資訊混亂程度的指標，萃取特徵資料集中的重要影響特徵。

嵌入類表示特徵選取流程與模型訓練為同時進行之方式，如 Tibshirani (1996) 所提出的最小絕對值收斂 (Least Absolute Shrinkage and Selection Operator; LASSO)，將回歸係數加入懲罰函數之回歸分析特徵選取方法。此外，Tsai 與 Hsiao (2010) 結合主成分分析法 (Principal Components Analysis; PCA)、基因演算法 (Genetic Algorithm; GA) 及分類與回歸樹 (Classification And Regression Trees; CART) 多種特徵選取演算法進行識別更具代表性之變量，並應用於股價預測。

三、模糊集合

真實世界中，存在著許多口語化之模糊意義，舉凡車速之快或慢、物體的大或小等，皆無存在一精確標準，每個人所認知的狀態不同，但在於每個人的理解中卻可快速並明確地判別。而根據此問題的產生，即奠定了數學運算相關領域的模糊集合 (Fuzzy Set; FS) (Zadeh 1965)，亦可稱為模糊集，此一方式將上述存在之問題，轉化成數學的集合概念，並保有其原先的模糊性質，更可以反映真實世界的狀況。在數學早期的集合 (或稱為明確集合) (Cantor 1874) 中，集合中之事物或東西稱之為元素，並採取二分條件方式判斷一元素為屬於或不屬於該集合，故於數值表示上，即採取 0 與 1 之方法，0 表示不屬於，反之，屬於則為 1。假設有一個元素為 x ，若 x 屬於一個集合 A ，可以記為 $x \in A$ ，否則 $x \notin A$ 。

模糊集合與明確集合 (Crisp set) 同樣地擁有各自的元素，不過兩者相較而言，前者可更精確表示元素屬於一個集合的程度，故不只有 0 與 1 的二元劃分，而是定義為 0 至 1 區間中的實數，其稱為隸屬程度 (Membership degree)，此隸屬程度必須藉由各模糊集合的隸屬函數 (Membership Function; MF) 決定。故相較於明確集合，模糊集合能夠更加精準判別元素的隸屬關係，真實地反映出現實世界存在之問題，現今廣泛應用於模糊相關領域之隸屬函數種類包括三角形 (Triangle)、梯形 (Trapezoid)、鐘形 (Bell) 與高斯 (Gaussian) 等隸屬函數。其中，高斯隸屬函數具備了延展性、非線性函數與可微分等特殊性質，將可以提高對於模糊隸屬程度的運算準確性。如 Nauck 與 Kruse (1999) 運用模糊類神經系統解決函數近似 (Function approximation) 相關問題。Manogaran、Varatharajan 與 Priyan (2018) 使用基於多核心學習之自適應模糊類神經推理系統應用於心臟疾病診斷。

四、機器學習最佳化演算法

機器學習領域屬於人工智慧之分支，由早期以推理為重點，逐漸轉化為知識導向，再延伸至現今的學習模式，其主要為設計與分析使電腦能夠自動學習的演算法，至今具有許多已成熟的機器學習演算法。

人工神經網路 (Artificial Neural Network; ANN) 可以簡稱為類神經網路 (Neural Network; NN) 或神經網路，早期於西元 1943 年由 McCulloch 與 Pitts (1943) 所創造，Rosenblatt (1957) 提出的感知器模型 (Perceptron) 為最早的類神經網路研究。其核心概念為模仿生物的神經系統，尤其為模擬大腦的機制，且具有多個神經元連結，部分負責接收資料，部分則負責傳遞資料，進而運用於機器學習相關領域中的模型架構或計算架構。概觀而言，類神經網路包含了三層神經元，分別為輸入層、隱藏層與輸出層。其中，隱藏層之主要目的為增加類神經網路的複雜度，越趨複雜的類神經網路相較於簡易架構更能夠解決非線性問題，同樣地，對於模型預測能力將有所提升。如 Khosravi 等 (2018) 建立多層前饋類神經網路 (Multilayer Feed-Forward Neural Network; MLFFNN) 進行風力相關預測。但需要注意當模型過於複雜，很可能會導致模型具有過度學習 (Overfitting) 的問題，即模型已高程度符合訓練資料集的特性，反而造成預測能力有所下降。

最佳化演算法領域中，仿生優化演算法的優點在於它屬全域極值解法，以及結構簡單、容易使用與快速收斂之特性，目前已有效地運用於廣泛的複雜問題上。如粒子群最佳化演算法 (Eberhart & Kennedy 1995) 為一個群體智慧 (Swarm Intelligence; SI) 協作機制之隨機搜尋演算法，最初為模擬鳥群優美而不可預測的

運動行徑，透過動物群體於社會行為之觀察，每一隻鳥之間運用資訊傳遞，使整體鳥群朝向食物源移動，由此啟發資訊共用的想法，作為此演算法的核心理念，Hasanipanah 等（2018）即以粒子群演算法結合自適應模糊類神經推理系統。Colorni 等（1992）所提出的蟻群演算法（Ant Colony Optimization; ACO）以螞蟻覓食之行為建立核心概念，運用螞蟻搬運食物路途中所留下的費洛蒙化學物質，轉換為其他螞蟻搜尋食物源之依據。此外，其他的最佳化演算法如遞迴最小平方估計演算法（Recursive Least Squares Estimator; RLSE）為最小平方估計演算法（Least Squares Estimator; LSE）之延伸，目的為根據輸入資料建立一個線性模型，而此線性模型與輸入資料的平方誤差須為最小值。

參、研究方法

本節將詳細說明研究運用之技術架構，以及由複數模糊集合（Complex Fuzzy Set; CFS）（Ramot et al. 2002）、T-S（Takagi-Sugeno）模糊模型（Takagi & Sugeno 1985）與模糊分類器（Fuzzy classifier）結合而成的複數模糊類神經分類模型（Complex Neuro-Fuzzy Classification Model; CNFC），並以 PSO-RLSE 複合式最佳化演算法調整模型參數的設計內容。圖 1 為整體研究流程圖：

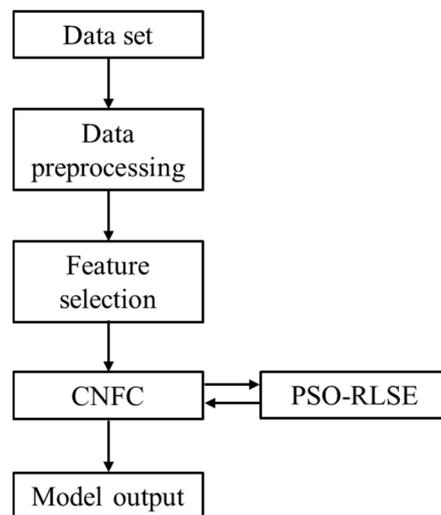


圖 1：整體研究流程圖

一、特徵選取

本篇論文對於欲處理資料的特徵選取部分，採用夏農資訊熵作為資訊混亂程度之指標，公式如下所示：

$$H(X) = -\sum_{i=1}^n P(x_i) \cdot \log P(x_i), \quad (1)$$

其中， $H(X)$ 為隨機變數 X 的資訊熵， $P(x_i)$ 為事件 x_i 的發生機率。由上式可得知，若事件 x_i 的發生機率越高，即 $P(x_i)$ ，資訊熵的計算結果將越小，若事件發生機率越趨近於平均值，則資訊熵將傾向最大值。目標於加入特徵後之資訊熵變化，表示特徵提供給予目標的資訊量，故將具備兩個隨機變數資訊，分別為原始的目標，以及在特徵條件前提下的目標。以式(2)為兩個隨機變數資訊共同發生的資訊熵計算公式：

$$H(X, Y) = -\sum_{x_i \in U_X} \sum_{y_j \in U_Y} P(i, j) \cdot \log P(i, j), \quad (2)$$

$H(X, Y)$ 為隨機變數 X 與 Y 發生共同事件情況下的資訊熵， $P(i, j)$ 表示事件 x_i 與 y_j 共同發生之機率， U_X 與 U_Y 則分別表示隨機變數 X 和 Y 的宇集合 (Universal set)。為了滿足特徵挑選的目的，我們必須了解在隨機變數 X 發生之後，目標隨機變數 Y 的資訊熵，方能與未發生隨機變數 X 的資訊熵進行比較，獲得其資訊量，故此資訊熵可稱為條件式資訊熵，亦即隨機變數 X 發生的條件之下，目標隨機變數 Y 的資訊熵，如式(3)所示：

$$H_X(Y) \equiv H(Y|X) = -\sum_{x_i \in U_X} \sum_{y_j \in U_Y} P(i, j) \log P_i(j), \quad (3)$$

其中， $p_i(j)$ 為事件 x_i 發生機率的前提之下，事件 y_j 發生之機率。此外，以上資訊熵的相關內容皆為離散型資料時所運用之計算方式，故 x_i 與 y_j 亦可稱為離散事件。不過倘若資料為連續型態時，我們將以機率密度函數 (Probability Density Function; PDF) 取代上述離散機率的方法。將上述式(1)與式(3)的資訊熵公式改寫為以下式(4)及(5)：

$$\begin{aligned} H(Y) &= -\int_y p(y) \cdot \log(p(y)) dy, \\ &= \int_y p(y) \cdot -\log(p(y)) dy, \\ &= \int_y p(y) \cdot \log\left(\frac{1}{p(y)}\right) dy, \end{aligned} \quad (4)$$

$$\begin{aligned} H_X(Y) &= -\int_X \int_Y p(x, y) \log p(y|x) dx dy, \\ &= \int_X \int_Y p(x, y) \cdot -\log p(y|x) dx dy, \end{aligned}$$

$$= \int_X \int_Y p(x, y) \cdot \log p\left(\frac{1}{p(y|x)}\right) dx dy, \quad (5)$$

$p(x, y)$ 為機率密度函數之表示， $p(y|x)$ 同樣表示為 x 事件發生前提下，事件 y 所發生的機率。經由上述關於資訊熵的解釋後，我們便可進一步地將式(4)減去式(5)，得出隨機變數 X 與 Y 的互資訊 (Mutual information)，如式(6)所示：

$$I(X, Y) = H(Y) - H_X(Y), \quad (6)$$

該互資訊中兩項隨機變數相互所提供之資訊量將為一致，即 $I(X, Y) = I(Y, X)$ 。但由於資料集數值分佈因素，造成資訊熵的值可能有負數之問題存在，根據夏農資訊熵的理論，其資訊熵之值不可為一負數值，故將於式(4)及(5)加入一參數 a 進行調整，以解決資訊熵計算可能為負數問題。 a 定義如式(7)所示：

$$a = \max\{\max[p(y)], \max[p(y|x)], 1\}. \quad (7)$$

加入參數 a 之式(4)及式(5)如式(8)與式(9)所示：

$$H(Y) = \int_Y p(y) \cdot \log\left(\frac{a}{p(y)}\right) dy, \quad (8)$$

$$H_X(Y) = \int_X \int_Y p(x, y) \cdot \log\left(\frac{a}{p(y|x)}\right) dx dy. \quad (9)$$

本研究於預測方面，主要以時間序列特徵資料之差量進行目標差量的預測。進一步地，我們更將時間序列特徵資料差量細分為正值與負值，此舉將可更加廣泛探討區域性之互資訊，故在條件式資訊熵中可分別表示為時間序列特徵資料差量 X 的正值前提條件下，對於目標差量的資訊熵，以及時間序列特徵資料差量 X 的負值前提條件下，同樣對於目標差量的資訊熵。故計算方式如以下式(10)與式(11)所示：

$$H_{X^+}(Y) = \int_X \int_Y p(x^+, y) \cdot \log\left(\frac{1}{p(y|x^+)}\right) dx dy, \quad (10)$$

$$H_{X^-}(Y) = \int_X \int_Y p(x^-, y) \cdot \log\left(\frac{1}{p(y|x^-)}\right) dx dy. \quad (11)$$

我們所需要之資訊為原始目標的資訊熵，以及在時間序列特徵資料前提條件下對於目標的資訊熵，如此即可獲得兩者差量，並計算互資訊，如式(12)與式(13)所示。計算完上述式(10)與式(11)並與原始目標資訊熵相減的兩種互資訊後，本研究進一步運用期望值概念，再分別與時間序列特徵資料差量 X 之機率密度分佈面積相乘，最終進行總和計算，此值稱為影響資訊 (Influence information)，意旨時

間序列特徵資料差量 X 對於目標差量 Y 所提供的影響訊息量，而互資訊則成為影響資訊的其中一種特例，且 $I_{X \rightarrow Y} \neq I_{Y \rightarrow X}$ 。其公式為以下式(14)所示：

$$I(X^+, Y) = H(Y) - H_{X^+}(Y), \quad (12)$$

$$I(X^-, Y) = H(Y) - H_{X^-}(Y), \quad (13)$$

$$I_{X \rightarrow Y} = I(X^+, Y) \int_0^\infty p(x)dx + I(X^-, Y) \int_{-\infty}^0 p(x)dx, \quad (14)$$

其中，機率密度函數 (Girolami & He 2003; Parzen 1962) 為根據目標與特徵資料所隱含之機率分布估計得知，描述某個取值點附近發生可能性的函數，其面積代表為資料發生於一區間之機率。在第四章的實驗中，將以常態分布 (Normal distribution) 方式計算原目標資料與特徵變數發生前提下之目標機率密度函數。

繼影響資訊計算後，我們將進一步建置如以下之影響資訊矩陣 (Influence Information Matrix; IIM)：

$$\begin{matrix}
 & f_1 & f_2 & \dots & f_n & t \\
 \begin{matrix} f_1 \\ f_2 \\ \vdots \\ f_n \\ t \end{matrix} & \left[\begin{array}{cccccc}
 I(f_1 \rightarrow f_1) & I(f_1 \rightarrow f_2) & \dots & I(f_1 \rightarrow f_n) & I(f_1 \rightarrow t) \\
 I(f_2 \rightarrow f_1) & I(f_2 \rightarrow f_2) & \dots & I(f_2 \rightarrow f_n) & I(f_2 \rightarrow t) \\
 \vdots & \vdots & \ddots & \vdots & \vdots \\
 I(f_n \rightarrow f_1) & I(f_n \rightarrow f_2) & \dots & I(f_n \rightarrow f_n) & I(f_n \rightarrow t) \\
 I(t \rightarrow f_1) & I(t \rightarrow f_2) & \dots & I(t \rightarrow f_n) & I(t \rightarrow t)
 \end{array} \right] , & (15)
 \end{matrix}$$

由於目標差量同樣被視為一特徵變數，故上述矩陣中特徵 f_i 之總數量為 $n + 1$ 個， $n = 1, 2, \dots, n, t$ 。每一個特徵變數除了對於目標會產生資訊量之外，對於其他特徵變數同樣將有不同的資訊量產生，若需要選取多個特徵變數進入模型進行學習，將必須考量影響資訊矩陣內所有相關之資訊量，在此亦著重於候選特徵對於目標之影響資訊，而非反向的影響。各特徵變數與目標對於本身所提供之資訊量，對應至影響資訊矩陣內中的對角線之值，將為零。此外，若對於候選的特徵變數所給予目標之資訊量已有部分由已選取的特徵變數提供，我們稱為這些資訊量為冗餘資訊 (Redundant Information; RI)。

面對上述問題，本研究定義特徵變數最終給予目標的資訊量為選取增益 (Selection Gain; SG)，其計算方式如式(16)所示。此亦可針對每一個特徵變數對於目標的選取增益進行排序，形成增益表 (Gain table)。

$$SG = I_{f_i \rightarrow f_n} - RI_{f_i \rightarrow SF^{(n)}}, \quad (16)$$

其中，上述 $I_{f_i \rightarrow f_n}$ 為第 i 個候選時間序列特徵資料差量 f_i 對於第 n 個目標差量 f_n 所提供的影響資訊量， $i = 1, 2, \dots, n + 1$ 。 $RI_{f_i \rightarrow SF^{(n)}}$ 則為第 i 個候選時間序列特徵資料差量 f_i 對於第 n 個已選取特徵 $SF^{(n)}$ 所產生的冗餘資訊量。而冗餘資訊的詳細計算方式為候選特徵對於已選取特徵相互間之資訊量平均，公式如下所示：

$$RI_{f_i \rightarrow SF^{(n)}} = \sum_{i=1}^{n+1} \frac{I_{f_i \rightarrow f_n} + I_{f_n \rightarrow f_i}}{2}, \quad (17)$$

其中， f_i 為目前已被選取之特徵， n 表示為第 n 個已被挑選的特徵， $i = 1, 2, \dots, n + 1$ 。以下將針對特徵選取流程的步驟詳細說明：

- 步驟1. 首先確立候選特徵池（Candidate Feature Pool; CFP）及已選特徵池（Selected Feature Pool; SFP），將所有計算完資訊量之特徵歸納至候選特徵池，以便後續特徵選取流程的進行。
- 步驟2. 由於初始時還未有特徵已被挑選，故式(17)之初始化設置為0。而第一個要挑選的特徵將為對於目標具有最大資訊量提供之特徵，成為第一個進入已選特徵池的特徵，並且該特徵同時亦不存在於候選特徵池。
- 步驟3. 根據步驟2挑選完第一個特徵後，將可接續逐一計算式(17)，並以選取增益為主要指標，如上述式(16)所示，將該值為正數之特徵選取至已選特徵池，並同樣地將該特徵由候選特徵池移除。
- 步驟4. 不斷重複執行步驟3，直至候選特徵池之特徵皆完成選取增益計算，並正確選取特徵至已選特徵池，特徵選取作業才得以結束，進而根據特徵挑選順序，依序進入模型進行訓練。

以下圖2將以流程圖的方式呈現特徵選取整體流程：

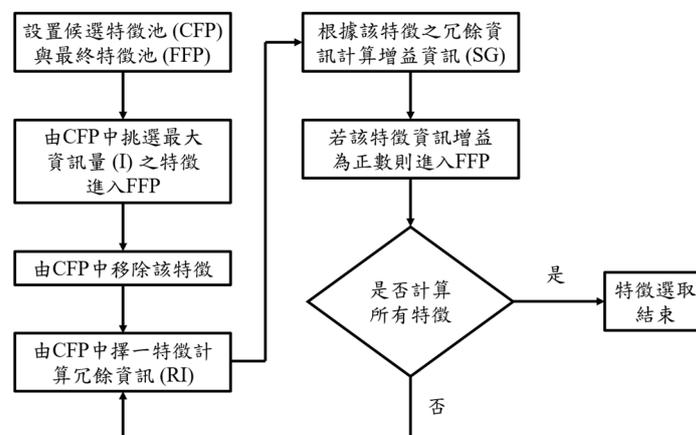


圖 2：特徵選取流程圖

二、複數模糊集合

如同前面章節所述，模糊集合可以以連續型的方式說明元素的隸屬程度，其值介於 0 至 1 之間。西元 2002 年，Ramot 等 (2002) 提出了複數模糊集合的概念，核心在於將原先模糊集合的一維隸屬程度擴展至二維之複數單位圓盤，由實部與虛部組合而成，相較於原本的模糊集合，增加了虛部值的空間。此改良將可大幅延伸隸屬程度的值域，更加精細描述一個元素的隸屬程度。同樣地，若有一個複數模糊集合 A ，則其元素 x 的隸屬程度可以表示如下：

$$\mu_A(x) = r_A(x) \cdot e^{jw_A(x)}, = r_A(x) \cdot \cos(w_A(x)) + j \cdot r_A(x) \cdot \sin(w_A(x)), \quad (18)$$

其中， j 為 $\sqrt{-1}$ ， $r_A(x)$ 稱為振幅函數， $w_A(x)$ 則為相位函數，若相位函數為 0 的情況而言，其隸屬程度之值將回歸至原本的模糊集合，亦即為一個 0 至 1 之間的實數。故由此可得知，當虛部為 0 時，複數模糊集合將還原為一維模糊集合。若此複數模糊集合 A 定義於宇集中，則如下式(19)所示：

$$A = \{(x, \mu_A(x)) | x \in U\}. \quad (19)$$

西元 2013 年，Li 與 Chiang (2013) 將高斯函數與複數模糊集合結合，形成複數高斯隸屬函數 (Complex Gaussian Membership Function; cGMF)，此函數的應用有效提升模型的預測能力，並保有前述高斯函數之特性，其公式如下所示：

$$cGMF(x, c, \sigma, \lambda) = r_A(x, c, \sigma) \cdot \exp(j\omega_A(x, c, \sigma, \lambda)), \quad (20)$$

$$r_A(x, c, \sigma) = Gaussian(x, c, \sigma), = \exp\left[-\frac{(x-c)^2}{2\sigma^2}\right], \quad (21)$$

$$w_A(x, c, \sigma, \lambda) = -\exp\left[-\frac{(x-c)^2}{2\sigma^2}\right] \cdot \frac{(x-c)}{\sigma^2} \cdot \lambda, = -r_A \cdot \frac{(x-c)}{\sigma^2} \cdot \lambda, \quad (22)$$

其中， x 為輸入資料， c 為高斯的中心參數， σ 則為其延展參數，而 λ 則為相位頻率參數。

本研究運用 T-S 模糊模型作為複數模糊類神經分類模型的模糊系統，其為 Takagi 與 Sugeno (1985) 兩位研究學者所提出的理論。其核心思想是採用多個線性系統，針對一個非線性系統進行擬合的技術，故此模型亦可稱為複合式非線性系統。該系統內部由多個 If-Then 模糊規則組合而成，具有連續函數的映射能力，而此方式非常貼近人們平時的邏輯經驗，故於理解方面相對容易許多，有關模型設計說明如下：

$$Rule^{(i)}: IF x_1 \text{ is } A_1^{(i)}(h_1) \text{ and } x_2 \text{ is } A_2^{(i)}(h_2) \text{ and } \dots \text{ and } x_M \text{ is } A_M^{(i)}(h_M), \quad (23)$$

$$\text{THEN } S^{(i)} = a_0^{(i)} + a_1^{(i)}h_1 + \dots + a_M^{(i)}h_M,$$

其中， i 為第 i 條規則， $i = 1, 2, 3, \dots, R$ ， $\{x_1, x_2, \dots, x_M\}$ 為輸入資料的語意變數， $\{A_1^{(i)}, A_2^{(i)}, \dots, A_M^{(i)}\}$ 為第 i 條規則的複數高斯模糊集合，複數高斯函數之參數為前鑑部 (If part) 參數， $\{h_1, h_2, \dots, h_M\}$ 為規則之輸入資料，而 $\{a_0^{(i)}, a_1^{(i)}, \dots, a_M^{(i)}\}$ 為第 i 條規則的後鑑部 (Then part) 參數。

三、複數模糊類神經分類模型

基於 T-S 模糊模型所設計的複數模糊類神經分類模型架構如以下圖 3 所示：

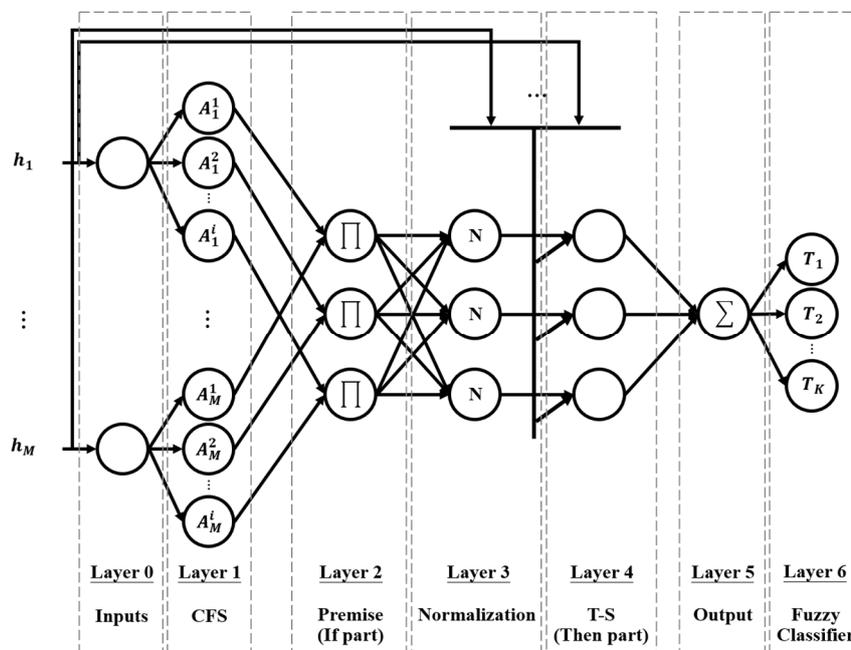


圖 3：複數模糊類神經分類模型架構

第零層 (Layer 0)：此層稱為輸入層，將欲分析之資料輸入模型並傳遞至類神經網路的下一層，即是此層的主要目的。本研究將已經過特徵挑選流程之時間序列資料輸入模型，時間 t 的輸入資料向量如下所示：

$$\bar{H} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_M \end{bmatrix}. \tag{24}$$

第一層 (Layer 1)：此層稱為複數模糊集合層，如前所敘，將其結合高斯函數型態，公式如式(20)。經由此隸屬函數可以計算輸入資料屬於複數模糊集合的隸屬程度，而此層將具有複數高斯隸屬函數之中心 (c)、延展度 (σ) 以及相位頻率 (λ) 參數，於此模型中，共同稱為前鑑部參數，其將使用 PSO 機器學習演算法進行優化。

第二層 (Layer 2)：此層稱為啟動強度層，亦可稱為前鑑部層，將計算每條規則的啟動強度，而計算方式採用複數模糊集合隸屬程度之積運算，其公式如下所示：

$$\beta^{(i)} = \prod_{k=1}^M \mu_{A_k^{(i)}} [h_k], \quad (25)$$

其中， $\mu_{A_k^{(i)}} [h_k]$ 表示對於 t 時間的輸入資料，其第 i 條規則的第 k 個複數高斯模糊集合之隸屬程度， $i = 1, 2, \dots, R$ ， R 為規則總數量。

第三層 (Layer 3)：此層稱為正規化層，目的為轉換每條規則於模型中之所佔比例，將每條規則得到的啟動強度進行正規化計算，其計算方式如下：

$$\lambda^{(i)} = \frac{\beta^{(i)}}{\sum_{i=1}^R \beta^{(i)}}. \quad (26)$$

第四層 (Layer 4)：此層稱為後鑑部層，即運用前一層所計算的正規化啟動強度運算每一條規則於後鑑部之反應值，故可得知各規則轉換為所佔比例後的輸出值，計算方式如式(27)所示，而此層的後鑑部參數 $\{a_0^{(i)}, a_1^{(i)}, \dots, a_M^{(i)}\}$ 將由 RLSE 演算法進行優化。

$$\hat{y}^{(i)} = \lambda^{(i)} \cdot S^{(i)}, = \lambda^{(i)} \cdot \left(a_0^{(i)} + a_1^{(i)} h_1 + \dots + a_M^{(i)} h_M \right). \quad (27)$$

第五層 (Layer 5)：此層稱為輸出層，即將前一層所計算每一條規則之輸出值進行加總，此值成為複數模糊類神經模型精準預測的總輸出：

$$\hat{Y} = \sum_{i=1}^R \hat{y}^{(i)}, \quad (28)$$

因複數高斯模糊集合之隸屬程度為複數型態，故上式中， \hat{Y} 同樣為複數型態，可以下方式表示：

$$\hat{Y} = \text{Re}(\hat{Y}) + j \cdot \text{Im}(\hat{Y}), \quad (29)$$

其中， j 為 $\sqrt{-1}$ ， $\text{Re}(\hat{Y})$ 代表精準預測總輸出之實數部，而 $\text{Im}(\hat{Y})$ 為精準預測

總輸出之虛數部數值。根據上述計算完精準預測之總輸出，將以此值與目標資料值進行誤差計算，並採用本研究所設計的成本函數進行適應值估計。假設輸入資料為以下所示：

$$TD = \{(\bar{x}^{t\{t\}}, y^{t\{t\}}), t = 1, 2, \dots, n\}, \quad (30)$$

其中，上式表示為第 t 對之輸入資料向量 \bar{x} 與目標資料 y ，兩者的誤差值計算如式(31)所示：

$$e^{t\{t\}} = y^{t\{t\}} - \hat{Y}^{t\{t\}}. \quad (31)$$

第六層 (Layer 6)：此層為模糊分類器，亦為整體模型的最後一層，目的為將前一層所計算之精準預測進行複數高斯模糊集合的模糊分類，估計其類別之隸屬程度，公式如下所示：

$$\hat{C}_q = \mu_{T_q}(\hat{Y}), \quad (32)$$

其中， $q = 1, 2, \dots, K$ ，表示為類別數量。而本研究設計之成本函數由均方根誤差 (Root-Mean-Squares Error; RMSE) 以及分類錯誤率所組成，做為複數模糊類神經分類模型之效能指標，其 RMSE 公式如下所示：

$$RMSE = \sqrt{\frac{\sum_{t=1}^n e^{t\{t\}*} \cdot e^{t\{t\}}}{n}}, \quad (33)$$

上述 RMSE 公式中， $e^{t\{t\}*}$ 為 $e^{t\{t\}}$ 的共軛轉置矩陣，其 $\{e^{t\{t\}}, t = 1, 2, \dots, n\}$ ， n 表示資料總筆數。對於成本函數的設計，考量在於若只採用 RMSE，可能導致機器學習演算法優化誤差值時，並無法有效使分類準確率提升，故同時評估精準預測與分類兩種指標，將可使模型於優化時，具有整合性的誤差估計，如式(34)所示。在成本函數後項的分類錯誤率中，將運用混淆矩陣 (Confusion matrix) (Stehman 1997) 之偽陽性 (False Positive; FP) 與偽陰性 (False Negative; FN) 進行說明，如圖 4。另外，此矩陣中亦包括了真陽性 (True Positive; TP) 與真陰性 (True Negative; TN)，特別於機器學習的監督式學習中使用，矩陣中的每一列表示類別預測，每一行代表實際類別，作為檢測分類器的指標之一。

		Real condition	
	Total population	Positive	Negative
Predicted outcome	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

圖 4：混淆矩陣

$$\text{Cost Function} = W_1 \cdot \text{RSME} + W_2 \cdot \left[\frac{n_{\text{FP}} + n_{\text{FN}}}{n} \right], \tag{34}$$

其中， W_1 為前項精準預測誤差權重， W_2 為後項之分類錯誤率權重， n_{FP} 與 n_{FN} 表示類別預測錯誤之數量， n 則為資料總筆數。若上述權重為相同，即令最佳化演算法能夠較平衡地調整兩項誤差，在本研究的實驗中，亦將針對分類錯誤率加入不同權重因子進行實驗，使其可以較大幅度地對於類別預測進行優化訓練。此外，實驗階段將以分類準確率作為效能比較指標，公式定義如下：

$$\text{Hit Ratio} = \left[\frac{n_{\text{TP}} + n_{\text{TN}}}{n} \right] \times 100\%, \tag{35}$$

其中， n 為資料總筆數。根據混淆矩陣呈現意涵，若區分為正與負類別， n_{TP} 表示正確預測正分類的資料總筆數， n_{TN} 代表正確預測負分類的資料總筆數。Hit Ratio 則為正確預測分類資料佔總筆數的比例。

四、PSO-RLSE 複合式最佳化演算法

根據前述所建構之複數模糊類神經分類模型，可得精準預測之誤差，以及分類錯誤率數據，進一步地，即運用本研究所設計的成本函數進行適應值評估。在模型參數的調整部分，區分為前鑑部和後鑑部參數，以單一演算法而言，當面對需要最佳化參數量（或稱為維度）越多時，計算將更趨複雜，導致演算法優化時程變長，最佳解亦相對較難以尋求。面對上述問題，現今於機器學習領域常用降維方式解決，以更少的數據維度呈現原數據之特性，如主成分分析法（Principal Component Analysis; PCA）等。另一方面，則為使用多演算法方式共同進行模型參數優化，減少個別演算法的最佳化維度，本研究即採用此方法，透過粒子群演算法（Particle Swarm Optimization; PSO）與遞迴最小平方法（Recursive Least

Squares Estimator; RLSE) 調整模型不同部分之參數 (Tu & Li 2019; Li & Wu 2011; 李俊賢 & 江泰緯 2013), 除了增進演算法效率外, 更提升於最佳解的搜索成效。

PSO 中的每一顆粒子 (Particle) 為一個問題的解, 其進入優化函數 (或稱為目標函數) 後皆會獲得一個適應值 (Fitness), 根據此適應值, 進而判斷該顆粒子當前位置之優劣。在粒子的移動過程中, 還有一個移動速度 (Velocity) 因子決定粒子下一次移動的行徑方向和距離, 故此演算法經由不斷的迭代以優化粒子位置搜尋最佳解。每顆粒子在移動位置時, 將受到兩個因素影響, 分別為粒子個體歷次之最佳適應值位置 ($Pbest$), 以及粒子群體歷次最佳適應值位置 ($Gbest$)。因此, 每顆粒子皆會透過這兩項因素調整搜尋方向, 粒子速度更新公式如下所示:

$$\begin{aligned} \vec{V}_i(k+1) = & w \cdot \vec{V}_i(k) + C_1 \cdot \xi_1 \cdot (Pbest_i(k) - \vec{L}_i(k)) \\ & + C_2 \cdot \xi_2 \cdot (Gbest(k) - \vec{L}_i(k)), \end{aligned} \quad (36)$$

其中, k 表示第 k 次迭代, C_1 與 C_2 為學習常數, 而 ξ_1 和 ξ_2 則為隨機亂數, 其值介於 0 至 1 之間。 $\vec{V}_i(k) = [\vec{v}_{i1}(k), \vec{v}_{i2}(k), \dots, \vec{v}_{iD}(k)]$ 代表第 i 顆粒子在第 k 次迭代之速度, 且有 D 個維度, $\vec{L}_i(k) = [\vec{l}_{i1}(k), \vec{l}_{i2}(k), \dots, \vec{l}_{iD}(k)]$ 則為第 i 顆粒子在第 k 次迭代之位置, 而 $Pbest_i(k)$ 為第 i 顆粒子在第 k 次迭代之歷次個體最佳適應值位置, $Gbest(k)$ 為粒子群體在第 k 次迭代的歷次最佳適應值位置, w 表示慣性權重, 使粒子於速度更新時可以保持慣性, 進而拓展搜尋空間。另外, 由實驗結果得知, 此值若設置於 0.4 至 0.9 之間可以獲得較佳的優化效果, 接續上述粒子的速度更新公式後, 進行每顆粒子之位置更新, 如以下式(37)所示:

$$\vec{L}_i(k+1) = \vec{L}_i(k) + \vec{V}_i(k+1). \quad (37)$$

優化模型後鑑部參數之遞迴最小平方估計演算法 (Recursive Least Squares Estimator; RLSE) 為最小平方估計演算法 (Least Squares Estimator; LSE) 之延伸, 目的為根據輸入資料建立一個線性模型, 而此線性模型與輸入資料的平方誤差須為最小值。其模型之輸出 y 由線性參數所表示:

$$y = \sum_{i=1}^n \theta_i f_i(\vec{u}) + e, \quad (38)$$

其中, \vec{u} 為模型輸入向量, $\{f_i, i = 1, 2, \dots, n\}$ 為輸入資料函數, $\{\theta_i, i = 1, 2, \dots, n\}$ 為需要搜尋之參數, 而由於模型可能不足以完全適應目標系統, 以及資料所包含的雜訊, 故於線性模型中加入誤差值 e 。假設共有 m 筆輸入資料, 可將輸

入資料表示為 $\{u_i, i = 1, 2, \dots, m\}$ ，如此，即可得到 m 個線性方程式如以下式(39)所示：

$$\begin{aligned} y_1 &= f_1(u_1)\theta_1 + f_2(u_1)\theta_2 + \dots + f_n(u_1)\theta_n + e_1 \\ y_2 &= f_1(u_2)\theta_1 + f_2(u_2)\theta_2 + \dots + f_n(u_2)\theta_n + e_2 \\ &\vdots \\ y_m &= f_1(u_m)\theta_1 + f_2(u_m)\theta_2 + \dots + f_n(u_m)\theta_n + e_m, \end{aligned} \quad (39)$$

進一步地，將上述的線性方程式運用矩陣方式表示如下：

$$Y = A\theta + E, \quad (40)$$

Y 為模型輸出向量， A 為設計矩陣 (Design matrix)， θ 為需搜尋之參數向量，而 E 則為誤差向量。可分別如下表示：

$$Y = [y_1 \ y_2 \ \dots \ y_m]^T, \quad (41)$$

$$A = \begin{bmatrix} f_1(u_1) & f_2(u_1) & \dots & f_n(u_1) \\ f_1(u_2) & f_2(u_2) & \dots & f_n(u_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(u_m) & f_2(u_m) & \dots & f_n(u_m) \end{bmatrix}, \quad (42)$$

$$\theta = [\theta_1 \ \theta_2 \ \dots \ \theta_m]^T, \quad (43)$$

$$E = [e_1 \ e_2 \ \dots \ e_m]^T. \quad (44)$$

為使線性模型與輸入資料的平方誤差為最小值，運用最小平方估計法可得 θ 為以下所示：

$$\theta = (A^T A)^{-1} A^T Y. \quad (45)$$

遞迴最小平方估計法之核心概念，在於以循序漸進的方式計算，提升計算效率。在計算完第 k 筆輸入資料之 θ_k 後，將以此為基礎，運用最小的計算資源進行 θ_{k+1} 的估計，其方法定義如下：

$$P_{k+1} = P_k - \frac{P_k a_{k+1} a_{k+1}^T P_k}{1 + a_{k+1}^T P_k a_{k+1}}, \quad (46)$$

$$\theta_{k+1} = \theta_k + P_{k+1} a_{k+1} (y_{k+1} - a_{k+1}^T \theta_k), \quad (47)$$

其中， $\{k = 1, 2, \dots, m - 1\}$ ， a_k 表示為設計矩陣 A 的第 m 組數據， y_k 代表輸出向量 Y 之第 m 項數值， $P_{k+1} a_{k+1}$ 稱為自適應增益向量 (Adaptation gain vector)， $(y_{k+1} - a_{k+1}^T \theta_k)$ 為前項估計器 (Old estimator)。而在此方程式開始計算前，必須

先將 P_0 與 θ_0 進行初始化設置，其 P_0 初始化如下所示：

$$P_0 = aI, \quad (48)$$

其中， a 為一個極大整數，而 I 為單位矩陣。至於 θ_0 之初始化設置為需調整參數數量的零向量。關於 PSO-RLSE 複合式最佳化演算法之流程圖如圖 5 所示，以下將詳細說明此演算法的流程：

- 步驟 1. 將粒子群最佳化演算法中的粒子位置與粒子速度進行初始化設置，而此演算法運用於優化前鑑部之複數高斯模糊集合的中心 (c)、延展度 (σ) 以及相位頻率 (λ) 參數。
- 步驟 2. 以輸入於模型的訓練資料計算每條規則在正規化後之啟動強度， $\{\lambda_i, i = 1, 2, \dots, R\}$ 。
- 步驟 3. 將遞迴最小平方法之 P_0 與 θ_0 進行初始化設置，如式(48)，進而運用此演算法更新模型後鑑部參數。式(46)中的 a 定義如下：

$$a_{k+1} = [aa_1(k+1) \quad aa_2(k+1) \quad \dots \quad aa_R(k+1)], \quad (49)$$

$$aa_R(k+1) = [\lambda_R \quad \lambda_R h_1(k+1) \quad \dots \quad \lambda_R h_M(k+1)]. \quad (50)$$

依據上述公式繼而計算式(47)即可求得模型後鑑部參數向量 θ 。

- 步驟 4. 計算複數模糊類神經分類模型的精準預測輸出，並與目標資料估計誤差值，如式(31)所示，利用此值接續計算均方根誤差 (Root-Mean-Square Error; RMSE)，如式(33)，即為本研究成本函數之前項。
- 步驟 5. 將模型之精準預測輸出進一步藉由複數高斯模糊集合計算類別隸屬程度，依據此計算分類錯誤率，使其結合前一步驟所得的均方根誤差，成為模型之效能指標。
- 步驟 6. 返回至步驟 2，直至完成所有粒子數量的計算。
- 步驟 7. 根據所有粒子所得出之適應值，進行更新粒子群最佳化演算法中的粒子個體歷次之最佳適應值位置 ($Pbest$)，以及粒子群體歷次最佳適應值位置 ($Gbest$)，並進行各粒子速度與位置的更新。
- 步驟 8. 判別最佳化演算法是否達到停止條件，若以達成，則以粒子群體之 $Gbest$ 作為前鑑部最終優化參數，否則返回至步驟 2。

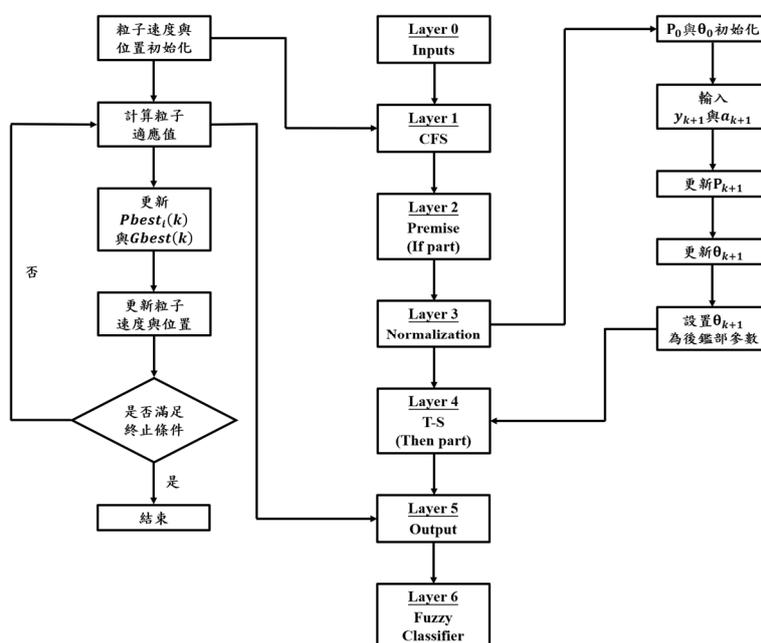


圖 5：PSO-RLSE 複合式最佳化演算法流程圖

肆、實驗

本實驗將進行單目標之時間序列型態資料預測，實驗目標為建立複數模糊類神經分類模型，並運用 PSO-RLSE 複合式最佳化演算法針對模型之前鑑部與後鑑部進行參數訓練，成本函數則採用本研究所設計之 RMSE 以及分類錯誤率作為模型的效能指標，並區分不同的 RMSE 與分類錯誤率之比例實驗，亦表示改變兩項指標於成本函數中的影響程度，包含了原本的 1 比 1 與 1 比 20 之比例。最後以模型輸出預測股價與分類。實驗方面增加高斯分布鯨魚最佳化演算法 (Gaussian Distribution Based Whale Optimization Algorithm; GD-WOA) (王伯倫 & 李俊賢 2019) 進行與本研究方法之效能比較。該演算法與粒子群最佳化演算法同樣隸屬一最佳化問題的解決方案，其核心概念為模擬座頭鯨 (Humpback whale) 捕獵水中生物進食的一套方法，該方式稱為氣幕捕獵 (Bubble-net feeding)。座頭鯨喜歡於水面附近捕殺磷蝦或小魚，其方式將先下潛約十二公尺，接續於獵物下方以螺旋狀不斷向上移動並產生氣泡，形成由氣泡組成之圓柱形氣幕牆，魚群將被困於該氣幕牆中，進而使座頭鯨可以一次捕獵魚群中大多數的個體。在數學模型方面，每隻鯨魚視為一組參數解，迭代過程中的鯨魚位置更新方向將朝向當前最佳鯨魚位置周圍之超立方體 (Hyper-cubes) 空間移動，並應用縮小環繞與螺旋更新位置兩種機制，模擬座頭鯨的氣幕捕獵行為。其中，螺旋更新位置公式為由候選

鯨魚位置與當前最佳鯨魚位置間的直線距離所創建，而縮小環繞機制除了根據當前最佳鯨魚位置進行最佳化之外，亦加入座頭鯨隨機搜索特性，隨機選取任一候選鯨魚位置作為最佳解搜尋範圍，此舉可提升演算法的全域性質。在鯨魚位置優化過程中，則以 50% 之機率條件選擇以縮小環繞或以螺旋方式進行位置更新。

一、實驗一：日經平均指數時間序列分類預測

日經平均指數又稱為日經 225 (Nikkei 225)，其為日本經濟新聞推出東京證券交易所的 225 品種股價指數。本次實驗擷取西元 2017 年 1 月至西元 2018 年 12 月作為資料集，在進入模型前，將先進行資料前處理步驟，進而計算特徵選取作業，候選特徵則選擇前 30 個交易日的股價指數漲跌變化，其中包括了開盤價、收盤價、最高價與最低價。經由特徵選取後， $\{f_{27}, f_{47}, f_{23}, f_{20}\}$ 為本實驗選入模型之特徵變數，分別為第 27 個、第 47 個、第 23 個與第 20 個特徵，預測模型的輸入資料為 $\{(\bar{x}^{(i)}, y^{(i)}), i = 1, 2, \dots\}$ ， i 表示資料對序， $\bar{x}^{(i)}$ 為特徵資料輸入向量， $y^{(i)}$ 則代表模型將預測之日經平均指數的股價指數目標資料，其 $\bar{x}^{(i)}$ 之定義如下所示：

$$\bar{x}^{(i)} = [f_{27}, f_{47}, f_{23}, f_{20}] \quad (51)$$

關於複數模糊類神經分類模型的設定如表 1 所示。進而藉由 PSO-RLSE 複合式最佳化演算法進行調整，其參數設定如表 2 與表 3。輸入特徵與目標資料相互之機率密度函數為圖 6，此外，為了驗證模型與複合式最佳化演算法的魯棒性 (Robustness)，將進行十次重複性實驗，結果呈現如表 9，最佳學習曲線如圖 7 所示，最後本實驗與文獻之模型效能比較如表 4。

表 1：實驗一：複數模糊類神經分類模型設定

複數模糊類神經分類模型設定	
輸入特徵	$x(f_{27}), x(f_{47}), x(f_{23}), x(f_{20})$
訓練次數	50
前鑑部數量	5
前鑑部參數量	24
後鑑部規則數量	5
後鑑部參數量	25
訓練資料筆數	403
測試資料筆數	60

表 2：實驗一 PSO 最佳化演算法參數設定

PSO 設定	
粒子數量	50
粒子維度	30
迭代次數	50
慣性權重	0.85
學習常數	2
粒子初始位置	The value of the target in the SC
粒子初始速度	Random in $[-1, 1]^{30}$

表 3：實驗一 RLSE 最佳化演算法參數設定

RLSE 設定	
α	10^9
l	25*25 單位矩陣
P_0	αl
θ_0	25*1 零矩陣

表 4：實驗一效能比較：日經平均指數

Method	Hit Ratio (%)
RW (Huang et al. 2005)	50
LDA (Huang et al. 2005)	55
Logit (Leung et al. 2000)	63
Probit (Leung et al. 2000)	63
Probabilistic neural network (Leung et al. 2000)	63
Discriminant analysis (Leung et al. 2000)	68
QDA (Huang et al. 2005)	69
EBNN (Huang et al. 2005)	69
SVM (Huang et al. 2005)	73
Combining model (Huang et al. 2005)	75
CNFC (GD-WOA-RLSE, Cost 1:1)	74.15

CNFC (GD-WOA-RLSE, Cost 1:20)	74.48
CNFC (PSO-RLSE, Cost 1:20)	74.65
CNFC (PSO-RLSE, Cost 1:1) (Proposed)	75.31

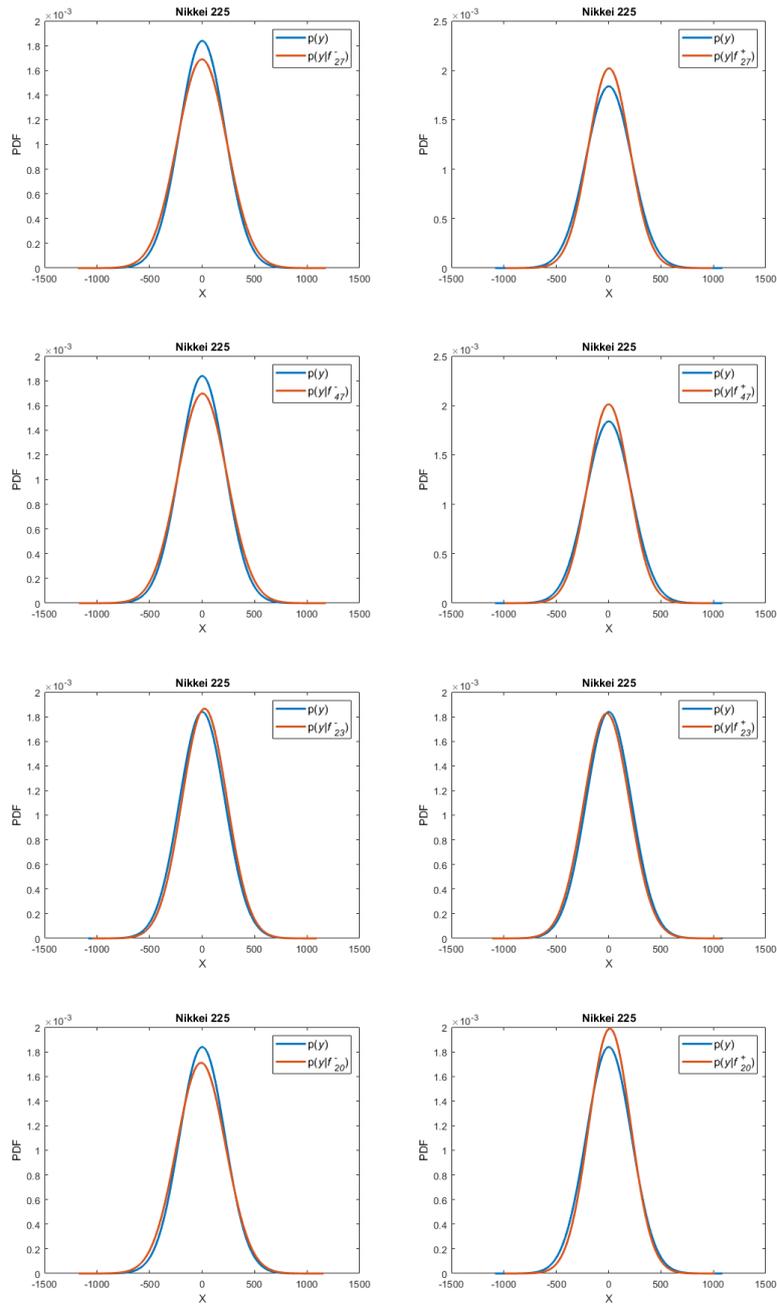


圖 6：實驗一輸入特徵與目標相互之機率密度函數

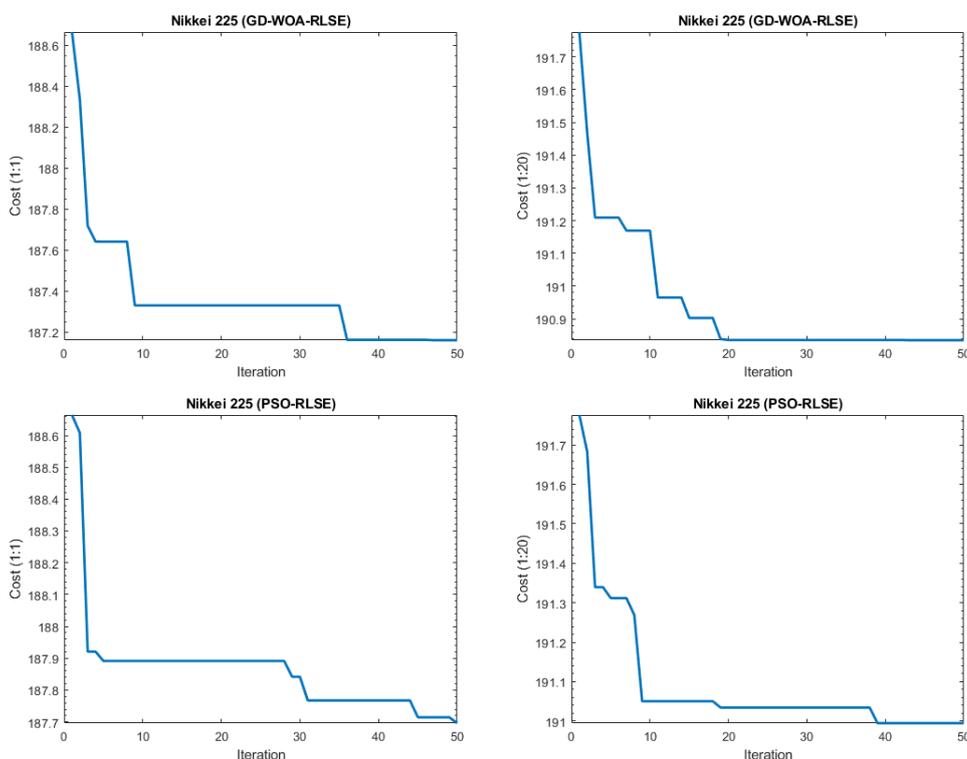


圖 7：實驗一複合式最佳化演算法學習曲線

經由實驗一對於日經平均股價指數的預測，在圖 7 中可得知複合式最佳化演算法於學習收斂之成效，更於表 4 中明顯判別本研究所提出的 CNFC 模型結合 PSO-RLSE 與文獻所提出之方法比較，凸顯其具有優越的預測能力。在第五節中，將額外加入重複性實驗結果進行綜合討論。

二、實驗二：韓國綜合股價指數時間序列分類預測

韓國綜合股價指數 (Korea Composite Stock Price Index; KOSPI) 為韓國交易所的股價指數。此次實驗擷取西元 2017 年 1 月至西元 2018 年 12 月作為資料集， $\{f_{18}, f_{27}, f_4, f_{71}\}$ 為本實驗選入模型之特徵變數，而預測模型的輸入資料為 $\{(\vec{x}^{(i)}, y^{(i)}), i = 1, 2, \dots\}$ ， i 表示資料對序， $\vec{x}^{(i)}$ 為特徵資料之輸入向量， $y^{(i)}$ 則代表模型將預測之韓國綜合股價指數的目標資料，其 $\vec{x}^{(i)}$ 之定義如下所示：

$$\vec{x}^{(i)} = [f_{18}, f_{27}, f_4, f_{71}]. \tag{52}$$

關於複數模糊類神經分類模型設定如表 5，PSO-RLSE 複合式最佳化演算法的參數設定同實驗一之表 2 與表 3 所示。輸入特徵與目標資料相互之機率密度函數

為圖 8，為了驗證模型與複合式最佳化演算法的魯棒性 (Robustness)，將進行十次重複性實驗，結果呈現如表 10 所示，最佳學習曲線如圖 9，本實驗與文獻之模型效能比較為顯示於表 6。

表 5：實驗二：複數模糊類神經分類模型設定

複數模糊類神經分類模型設定	
輸入特徵	$x(f_{18}), x(f_{27}), x(f_4), x(f_{71})$
訓練次數	50
前鑑部數量	5
前鑑部參數量	24
後鑑部規則數量	5
後鑑部參數量	25
訓練資料筆數	392
測試資料筆數	60

表 6：實驗二效能比較：韓國綜合股價指數

Method	Hit Ratio (%)
GALT (Kim & Han 2000)	50.60
BPLT (Kim & Han 2000)	51.81
RW (Wang 2014)	55
SVM (Wang 2014)	59
ANN (Wang 2014)	60
PCA-ANN (Wang 2014)	60
PCA-SVM (Wang 2014)	61
GAFD (Kim & Han 2000)	61.70
CNFC (GD-WOA-RLSE, Cost 1:1)	64.15
CNFC (GD-WOA-RLSE, Cost 1:20)	64.32
CNFC (PSO-RLSE, Cost 1:20)	64.32
CNFC (PSO-RLSE, Cost 1:1) (Proposed)	64.49

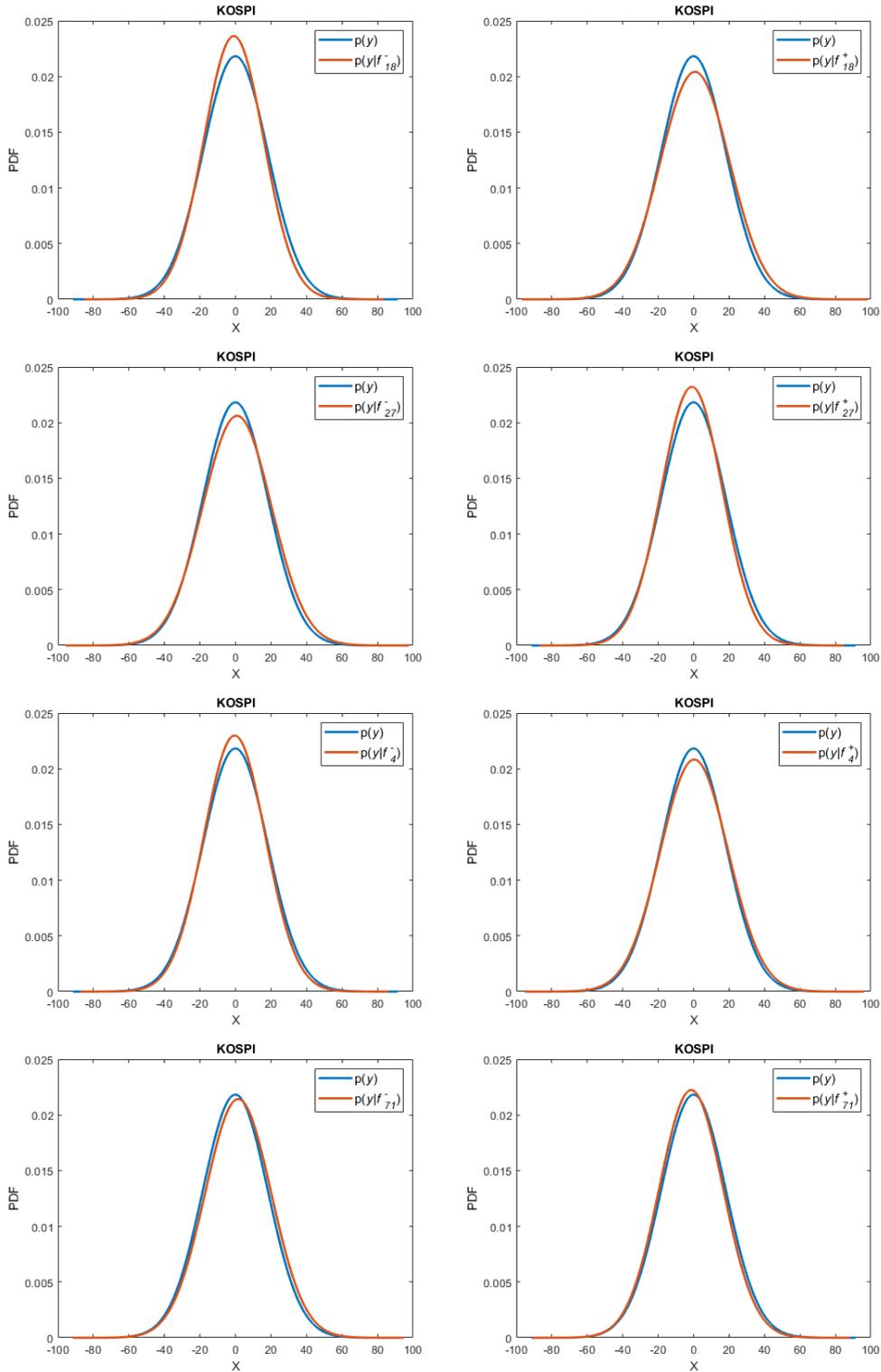


圖 8：實驗二輸入特徵與目標相互之機率密度函數

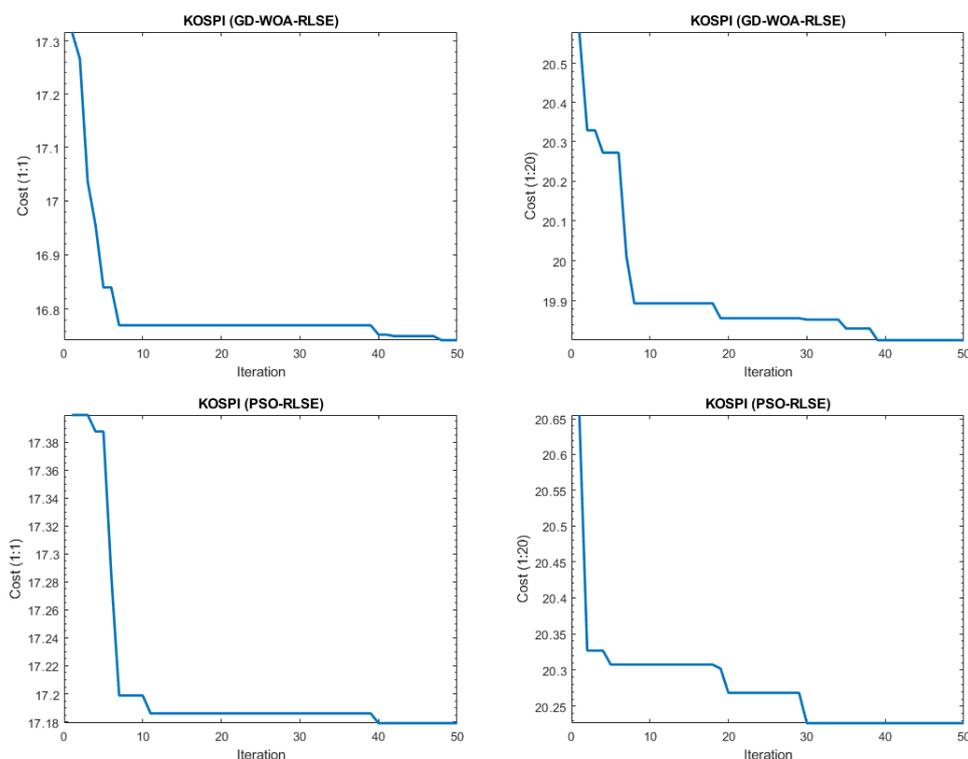


圖 9：實驗二複合式最佳化演算法學習曲線

本研究在實驗二中，驗證 CNFC 模型於韓國綜合股價指數之預測穩定性，因相較於日經平均指數，此股價指數之價差幅度較小，故於股價趨向的預測能力要求相對提升。在表 6 中的多樣化文獻模型比較中，CNFC 保持於最佳的預測效能。

三、實驗三：香港恆生指數時間序列分類預測

香港恆生指數 (Hang Seng Index; HSI) 為香港非常重要的股價指數，亦為反映香港股市行情的重要指標。本次實驗擷取西元 2017 年 1 月至西元 2018 年 12 月作為資料集，並設定 70 筆資料為測試資料集，其餘為訓練資料集。經由對於資料集進行特徵選取後， $\{f_{75}, f_{22}, f_{10}, f_1\}$ 為本實驗選入模型之特徵變數，而預測模型的輸入資料為 $\{(\tilde{x}^{(i)}, y^{(i)}), i = 1, 2, \dots\}$ ， i 表示資料對序， $\tilde{x}^{(i)}$ 為特徵資料之輸入向量， $y^{(i)}$ 則代表模型將預測之香港恆生指數的目標資料，其 $\tilde{x}^{(i)}$ 之定義如下所示：

$$\tilde{x}^{(i)} = [f_{75}, f_{22}, f_{10}, f_1]. \quad (53)$$

關於複數模糊類神經分類模型如表 7 所示，PSO-RLSE 複合式最佳化演算法

的參數設定同實驗一之表 2 與表 3。輸入特徵與目標資料相互之機率密度函數為圖 10，為了驗證模型與複合式最佳化演算法的魯棒性 (Robustness)，將進行十次重複性實驗，結果呈現如表 11 所示，最佳學習曲線如圖 11，本實驗與文獻之模型效能比較如表 8。

表 7：實驗三：複數模糊類神經分類模型設定

複數模糊類神經分類模型設定	
輸入特徵	$x(f_{75}), x(f_{22}), x(f_{10}), x(f_1)$
訓練次數	50
前鑑部數量	5
前鑑部參數量	24
後鑑部規則數量	5
後鑑部參數量	25
訓練資料筆數	391
測試資料筆數	70

表 8：實驗三效能比較：香港恆生指數

Method	Hit Ratio (%)
RW (Wang 2014)	52
SVM (Wang 2014)	58
ANN (Wang 2014)	58
PCA-ANN (Wang 2014)	62
PCA-SVM (Wang 2014)	62
SSPQN (Phua et al. 2003)	63.35
TRNN (Phua et al. 2003)	64.39
CNFC (GD-WOA-RLSE, Cost 1:1)	70.26
CNFC (GD-WOA-RLSE, Cost 1:20)	70.27
CNFC (PSO-RLSE, Cost 1:20)	70.84
CNFC (PSO-RLSE, Cost 1:1) (Proposed)	71.32

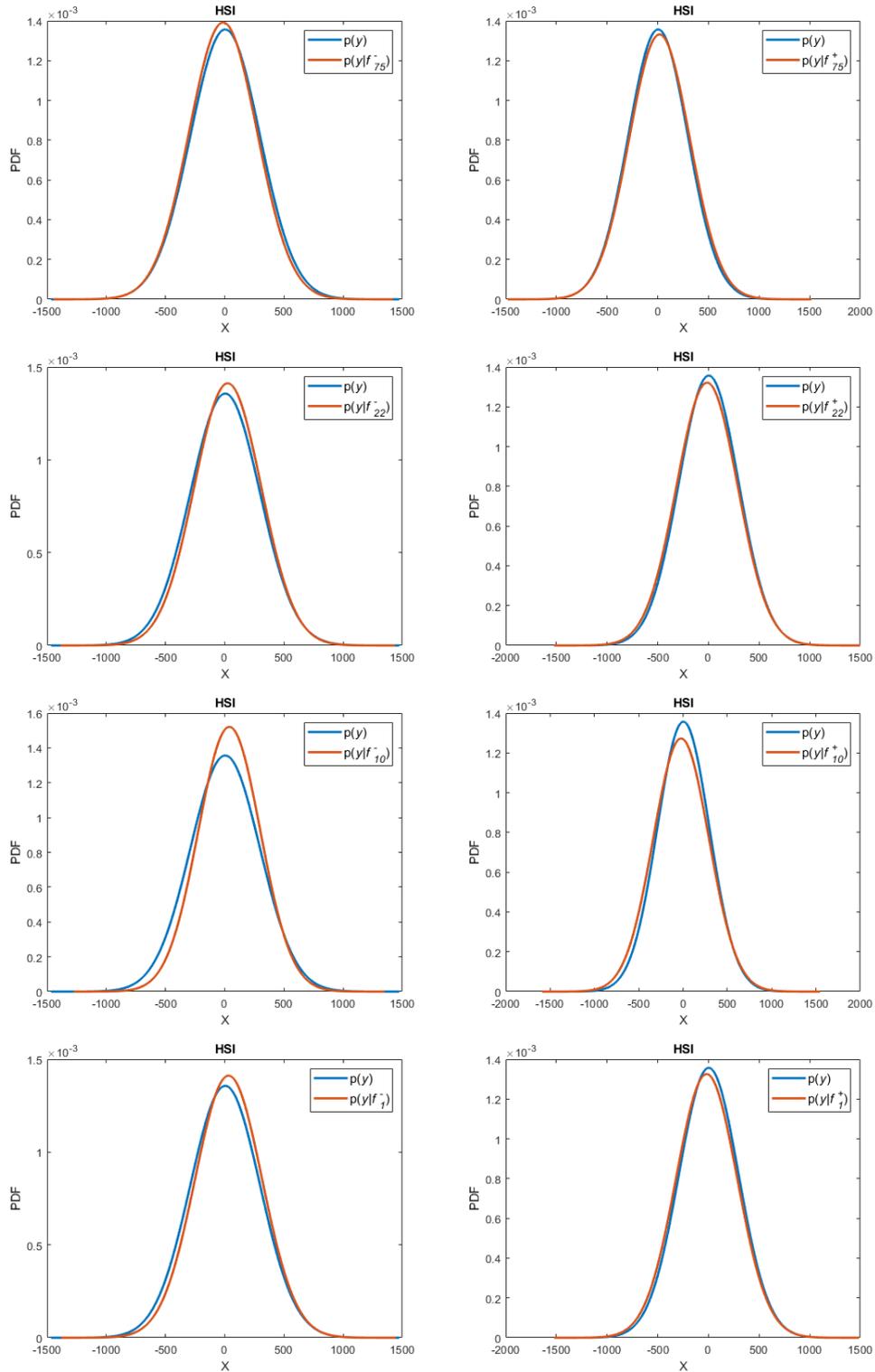


圖 10：實驗三輸入特徵與目標相互之機率密度函數

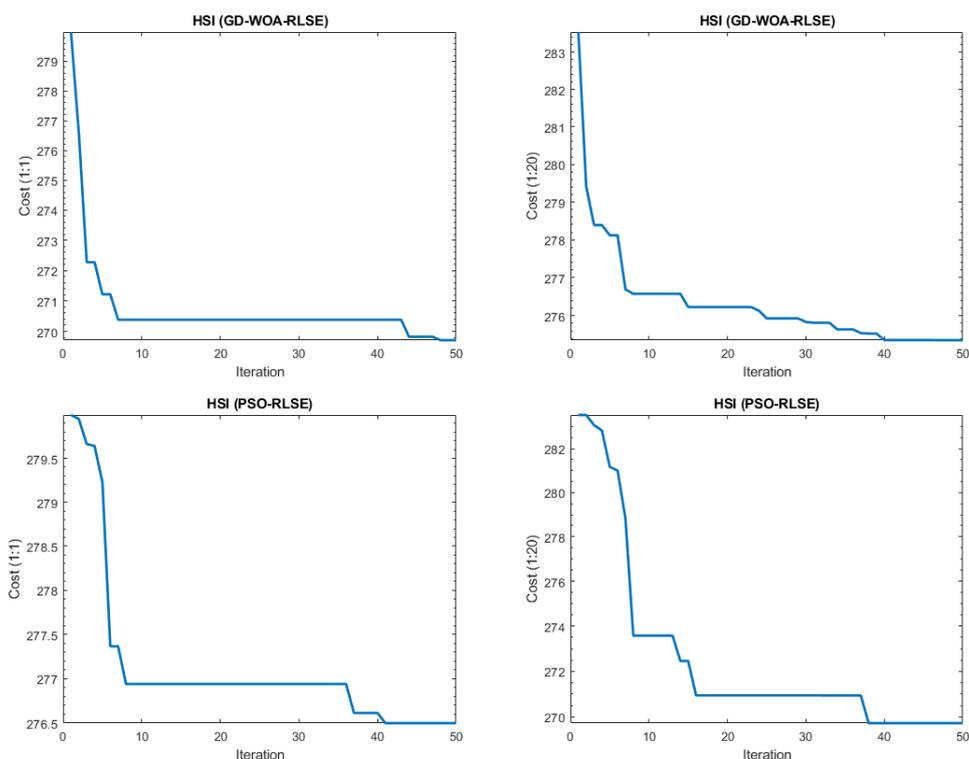


圖 11：實驗三複合式最佳化演算法學習曲線

香港恆生指數為全球非常著名的股市，故於相關研究較為廣泛，在表 8 中仍以 CNFC 搭配 PSO-RLSE 複合式最佳化演算法優於比較文獻，實驗的學習曲線同時亦凸顯複數高斯模糊集合拓展參數值域後之最佳化能力。

伍、討論

以下將對於上一章節的各實驗以及模型重複性實驗數據進行探討：

表 9：實驗一之 CNFC 模型之重複實驗

	GD-WOA-RLSE				PSO-RLSE			
	Cost 1:1		Cost 1:20		Cost 1:1		Cost 1:20	
No.	Hit (%)	Cost	Hit (%)	Cost	Hit (%)	Cost	Hit (%)	Cost
1.	75	352.364	73.3	359.103	75	352.241	75	356.991
2.	75	352.241	71.6	371.833	75	352.241	76.6	344.818
3.	73.3	361.629	75	356.991	76.6	343.807	73.3	359.557

4.	75	352.242	76.6	354.193	75	352.241	75	357.423
5.	75	352.241	75	356.991	75	352.241	75	357.142
6.	71.6	377.092	75	356.991	76.6	350.49	75	357.905
7.	73.3	355.376	73.3	360.602	73.3	352.241	73.3	357.420
8.	73.3	363.825	75	355.016	75	354.078	75	359.118
9.	75	352.241	75	356.991	75	352.241	75	356.991
10.	75	358.132	75	355.016	76.6	343.807	73.3	359.557
AVG	74.15	357.738	74.48	358.373	75.31	350.563	74.65	356.692
Max	75	377.092	76.6	371.833	76.6	354.078	76.6	359.557
Min	71.6	352.241	71.6	354.193	73.3	343.807	73.3	344.818

表 10：實驗二之 CNFC 模型之重複實驗

	GD-WOA-RLSE				PSO-RLSE			
	Cost 1:1		Cost 1:20		Cost 1:1		Cost 1:20	
No.	Hit (%)	Cost	Hit (%)	Cost	Hit (%)	Cost	Hit (%)	Cost
1.	65	25.037	65	31.66	65	25.03	65	31.6
2.	65	25.356	65	31.68	65	25.08	63.3	31.68
3.	61.6	25.512	61.6	32.34	65	24.98	63.3	32.39
4.	65	25.155	65	31.51	65	25.03	63.3	32.46
5.	61.6	25.566	65	31.66	63.3	25.34	65	31.68
6.	65	25.037	65	31.68	65	25	65	31.67
7.	65	25.003	63.3	32.16	63.3	25.24	65	31.7
8.	63.3	25.356	65	31.72	65	25.03	65	31.64
9.	65	25.037	63.3	32.34	65	25.08	63.3	32.46
10.	65	25.037	65	31.68	63.3	25.24	65	31.64
AVG	64.15	25.21	64.32	31.84	64.49	25.11	64.32	31.89
Max	65	25.566	65	32.34	65	25.34	65	32.46
Min	61.6	25.003	63.3	31.51	63.3	24.98	63.3	31.6

表 11：實驗三之 CNFC 模型之重複實驗

No.	GD-WOA-RLSE				PSO-RLSE			
	Cost 1:1		Cost 1:20		Cost 1:1		Cost 1:20	
	Hit (%)	Cost	Hit (%)	Cost	Hit (%)	Cost	Hit (%)	Cost
1.	71.4	362.26	70	377.56	73	360.83	70	375.5
2.	71.4	369.96	70	377.43	71.4	362.26	71.4	367.6
3.	70	372.04	71.4	367.68	70	387.39	70	379.1
4.	68.5	372.48	71.4	375.26	73	361.98	71.4	367.7
5.	70	372.13	71.4	375.27	70	368.37	70	368.8
6.	70	371.89	68.5	384.13	71.4	362.26	71.4	367.6
7.	68.5	374.55	70	377.56	73	361.34	70	379.2
8.	71.4	371.6	70	379.05	70	363.23	71.4	367.6
9.	71.4	370.72	70	377.57	71.4	362.26	71.4	367.6
10.	70	372.13	70	377.56	70	362.26	71.4	367.7
AVG	70.26	370.98	70.27	376.91	71.32	365.22	70.84	370.84
Max	71.4	374.55	71.4	384.13	73	387.39	71.4	379.2
Min	68.5	362.26	68.5	367.68	70	360.83	70	367.6

本研究所提出全新的複數模糊類神經分類模型，以減法分群演算法所建置之目標資料趨向類別為基礎，根據資料所呈現之特性進行動態預測。複數高斯模糊集合擴展了隸屬程度的範圍限制，使其具有實部與虛部結合而成之單位圓盤值域，增強模型優化效率。藉由 PSO-RLSE 複合式最佳化演算法，分別針對模型前鑑部之複數高斯模糊集合與後鑑部的 T-S 模糊模型參數進行最佳化，相較於單一演算法而言，將助於降低模型學習時間。基於夏農資訊熵理論，開發一套特徵選取演算方法，以影響資訊量與冗餘資訊量作為評價指標，計算每個特徵對於目標的重要程度，進而減少過多干擾特徵進入模型訓練階段，影響模型預測效能。

本實驗共進行三種實驗以驗證模型預測能力，分別為日經平均指數、韓國綜合股價指數與香港恆生指數，股票價格指數相較於成分股具有較大的股價波動幅度，故於預測上相對難以實踐。經由實驗一的表 9 進行不同複合式最佳化演算法的訓練結果，本研究主要採用的 PSO-RLSE 於相同成本函數設計比較中，重複性實驗平均之成本與分類準確率皆優於 GD-WOA-RLSE，顯示模型於 PSO-RLSE 複合式最佳化演算法中具有較佳的優化效能。此外，表 4 內所進行的多樣化模型比

較凸顯了 CFNC 於分類中擁有較高的預測水準。

實驗二於表 10 的重複性實驗中，明顯得知相較於其他股價指數之分類準確率具有些微下降情形，其因素為韓國綜合股價指數之股價波動幅度較小，故於分類預測之精確程度需求相對提升。在此情況中，PSO-RLSE 同樣保持較良好的優化能力，表 6 中驗證 CNFC 除了擁有較佳的分類準確率外，同時展現模型的魯棒性。實驗三於表 11 的 GD-WOA-RLSE 並結合以上兩種實驗結果，明顯於分類錯誤率比例增加之成本函數設計下，皆助於改變其複合式最佳化演算法之參數調整方向，助於分類準確率，表 8 亦顯示 CNFC 具有良好的分類預測效能。以上總和多樣化實驗證明複數模糊集合有效拓展模糊集合隸屬程度之值域，使複合式最佳化演算法能夠更快速進行模型參數調整，大幅提升模型的預測表現。整體而言，粒子群最佳化演算法相較於鯨魚最佳化演算法簡易且快速，雖然鯨魚演算法增加隨機機制以避免容易陷入區域最佳解問題，提升全域搜索機會，不過其運用機率條件於每次迭代中選擇不同優化方式，均化全域搜索與優化方向性質，反觀粒子群於優化時固定參照粒子個體與群體最佳位置之特性，可較穩定尋求最佳適應值 (Fitness)。在成本函數變化下，鯨魚演算法因具有較佳的全域搜索能力，故明顯於類別預測成本之權重較重時，擁有提升趨勢預測成效。

陸、結論

本研究所提出新型態的複數模糊類神經分類模型，其透過實驗結果驗證成本函數將影響複合式最佳化演算法的參數優化方向，間接提升分類準確率，對於文獻的多樣化比較，亦展現本研究模型的優越預測能力。此外，本篇論文之貢獻為藉由粒子群最佳化演算法調整複數模糊類神經分類模型之前鑑部參數，以及遞迴最小平方法優化模型的後鑑部參數，將有別於一般最佳化演算法的最佳化速度與效能，大幅提升模型的參數優化能力。以複數模糊集合擴展參數值域為實部與虛部，將助於提升複合式最佳化演算法的學習效果，增強模型之最佳化水平。在原始資料方面，根據夏農資訊熵的方法理論，開發一套特徵選取演算方式，此舉將使原始資料進行特徵篩選，透過夏農資訊熵與機率密度函數估計，搜尋助於模型預測目標之特徵，以及避免冗餘資料干擾模型預測效能，更加能夠提升模型運算效率。另外，提出新型態之複數模糊類神經分類模型，除了成本函數合併精準與分類預測效能指標，亦運用分群演算法轉換目標資料為趨向類別，具有動態類別預測能力。整體研究對於投資者角度而言，可初步提供輔助選擇方案以較高機率獲得較高投資報酬，非一取代性選項。至於在金融從業者方面，如證券分析師等職業，預判股價變化狀況將可進一步探討其與當前整體經濟之關聯性，對於未來投資與經營層面即可具有更加完善的規劃。

誌謝

這項研究工作得到了科技部經費支持 (MOST 105-2221-E-008-091 與 MOST 104-2221-E-008-116, TAIWAN), 表示感謝之意。

參考文獻

- 王伯倫、李俊賢 (2019), 『高斯分布鯨群演算法於最佳化問題之研究』, (to be submitted for publication)
- 李俊賢、江泰緯 (2013), 『混合複數類神經模糊與自動回歸差分平均移動方法之智慧型時間序列預測模型』, 電子商務學報, 第十五卷, 第一期, 頁 137-158。
- Bezdek, J.C. (1974), 'Cluster validity with fuzzy sets', *Journal of Cybernetics*, Vol. 3, No. 3, pp. 58-72.
- Bezdek, J.C. (2013), 'Pattern recognition with fuzzy objective function algorithms', Springer Science & Business Media.
- Bezdek, J.C., Ehrlich, R. and Full, W. (1984), 'FCM: The fuzzy c-means clustering algorithm', *Computers & Geosciences*, Vol. 10, No. 2-3, pp. 191-203.
- Bollen, J., Mao, H. and Zeng, X. (2011), 'Twitter mood predicts the stock market', *Journal of Computational Science*, Vol. 2, No. 1, pp. 1-8.
- Cantor, G. (1874), 'Ueber eine Eigenschaft des Inbegriffs aller reellen algebraischen Zahlen', *Journal für die reine und angewandte Mathematik*, Vol.77, pp. 252-268.
- Coloni, A., Dorigo, M. and Maniezzo, V. (1992), 'Distributed optimization by ant colonies', *Proceedings of the 1st European Conference on Artificial Life*, Vol. 142, pp. 134-142, Paris.
- Dimpfl, T. and Jank, S. (2016), 'Can internet search queries help to predict stock market volatility?', *European Financial Management*, Vol. 22, No. 2, pp. 171-192.
- Dunn, J.C. (1973), 'A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters', *Journal of Cybernetics*, Vol. 3, No. 3, pp. 32-57.
- Eberhart, R. and Kennedy, J. (1995), 'A new optimizer using particle swarm theory', *Proceedings IEEE International Symposium on Micro Machine and Human Science (Nagoya, Japan)*, pp. 39-43.
- Eberhart, R. and Kennedy, J. (1995), 'Particle swarm optimization', *Proceedings IEEE International Conference on Neural Networks (Perth, Australia)*, Vol. 4, pp. 1942-1948.

- Engle, R.F. (1982), 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *ECONOMETRICA*, Vol. 50, No. 4, pp. 987-1007.
- Girolami, M. and He, C. (2003), 'Probability density estimation from optimally condensed data samples', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 10, pp. 1253-1264.
- Guyon, I. and Elisseeff, A. (2003), 'An introduction to variable and feature selection', *Journal of Machine Learning Research*, pp. 1157-1182.
- Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002), 'Gene selection for cancer classification using support vector machines', *Machine Learning*, Vol. 46, No. 1-3, pp. 389-422.
- Hasanipanah, M., Amnieh, H.B., Arab, H. and Zamzam, M.S. (2018), 'Feasibility of PSO-ANFIS model to estimate rock fragmentation produced by mine blasting', *Neural Computing and Applications*, Vol. 30, No. 4, pp. 1015-1024.
- Huang, W., Nakamori, Y. and Wang, S.Y. (2005), 'Forecasting stock market movement direction with support vector machine', *Computer & Operations Research.*, Vol. 32, No. 10, pp. 2513-2522.
- Khosravi, A., Koury, R.N.N., Machado, L. and Pabon, J.J.G. (2018), 'Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system', *Sustainable Energy Technologies and Assessments*, Vol. 25, No. 4, pp. 146-160.
- Kim, K.J. and Han, I. (2000), 'Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index', *Expert Systems with Applications*, Vol. 19, No. 2, pp. 125-132.
- Leung, M.T., Daouk, H. and Chen, A.S. (2000), 'Forecasting stock indices: A comparison of classification and level estimation models', *International Journal of Forecasting*, Vol. 16, No. 2, pp. 173-190.
- Li, C. and Chiang, T.W. (2013), 'Complex Neurofuzzy ARIMA Forecasting-A New Approach Using Complex Fuzzy Sets', *IEEE Transactions on Fuzzy Systems*, Vol. 21, No. 3, pp. 567-584.
- Li, C. and Wu, T. (2011), 'Adaptive fuzzy approach to function approximation with PSO and RLSE', *Expert Systems with Applications*, Vol. 38, No. 10, pp. 13266-13273.
- Manogaran, G., Varatharajan, R. and Priyan, M.K. (2018), 'Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system', *Multimedia Tools and Applications*, Vol. 77, No. 4,

- pp. 4379-4399.
- McCulloch, W.S. and Pitts, W. (1943), 'A logical calculus of the ideas immanent in nervous activity', *The Bulletin of Mathematical Biophysics*, Vol. 5, No. 4, pp. 115-133.
- Nauck, D. and Kruse, R. (1999), 'Neuro-fuzzy systems for function approximation', *Fuzzy Sets and Systems*, Vol. 101, No. 2, pp. 261-271.
- Parzen, E. (1962), 'On Estimation of a Probability Density Function and Mode', *The Annals of Mathematical Statistics*, Vol. 33, No. 3, pp. 1065-1076.
- Patel, J., Shah, S., Thakkar, P. and Kotecha, K. (2015), 'Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques', *Expert Systems with Applications*, Vol. 42, No. 1, pp. 259-268.
- Pearson, K. (1900), 'X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling', *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Vol. 50, No. 302, pp. 157-175.
- Phua, P.K.H., Zhu, X. and Koh, C.H. (2003), 'Forecasting Stock Index Increments Using Neural Networks with Trust Region Methods', *Proceedings of the International Joint Conference on Neural Networks, IEEE*, Vol. 1, pp. 260-265.
- Ramot, D., Milo, R., Friedman, M. and Kandel, A. (2002), 'Complex fuzzy sets', *IEEE Transactions on Fuzzy Systems*, Vol. 10, No. 2, pp. 171-186.
- Rosenblatt, F. (1957), 'The Perceptron-A Perceiving and Recognizing Automaton', *Cornell Aeronautical Laboratory*.
- Shannon, C.E. and Weaver, W. (1949), 'The Mathematical Theory of Communication', Univ. of Illinois Press.
- Stehman, S.V. (1997), 'Selecting and Interpreting Measures of Thematic Classification Accuracy', *Remote Sensing of Environment*, Vol. 62, No. 1, pp. 77-89.
- Takagi, T. and Sugeno, M. (1985), 'Fuzzy identification of systems and its applications to modeling and control', *IEEE Systems, Man, and Cybernetics Society*, Vol. 15, No. 1, pp. 116-132.
- Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 58, No. 1, pp. 267-288.
- Tsai, C.F. and Hsiao, Y.C. (2010), 'Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches', *Decision Support Systems*, Vol. 50, No. 1, pp. 258-269.

- Tu, C.H. and Li, C. (2019), 'Multitarget prediction-A new approach using sphere complex fuzzy sets', *Engineering Applications of Artificial Intelligence*, Vol. 79, pp. 45-57.
- Wang, Y. (2014), 'Stock price direction prediction by directly using prices data: An empirical study on the KOSPI and HIS', *International Journal of Business Intelligence and Data Mining*, Vol. 9, No. 2, pp. 145-160.
- Yager, R.R. and Filev, D.P. (1994), 'Generation of Fuzzy Rules by Mountain Clustering', *Journal of Intelligent & Fuzzy Systems*, Vol. 2, No. 3, pp. 209-219.
- Zadeh, L.A. (1965), 'Fuzzy Sets', *Information and Control*, Vol. 8, No. 3, pp. 338-353.
- Zheng, Y., Jeon, B., Xu, D., Wu, Q.M. and Zhang, H. (2015), 'Image segmentation by generalized hierarchical fuzzy C-means algorithm', *Journal of Intelligent & Fuzzy Systems*, Vol. 28, No. 2, pp. 961-973.