

陳振東、謝政翰（2019），『應用機器學習與模糊推論於股價漲跌預測之研究』，*中華民國資訊管理學報*，第二十六卷，第二期，頁153-178。

應用機器學習與模糊推論於股價漲跌預測之研究

陳振東*

國立聯合大學資訊管理學系

謝政翰

國立聯合大學資訊管理學系

摘要

近年來，利用智慧數據分析方法以預測股價乃是金融科技（Financial Technology; FinTech）領域的重要議題。然而，有許多的技術指標以及人為主觀因素會影響股價的漲跌預測，因此必須有效掌握重要的影響指標，才能提高股價漲跌預測的正確率。為此，本研究透過技術指標的篩選程序，使用四種機器學習演算法進行股價漲跌的預測分析，進而篩選重要的技術指標。此外，由於技術指標的屬性及人為的主觀判斷具有不確定性與模糊性，因此本研究應用模糊推論方法建構模糊推論系統以進行股價漲跌的預測，並提出股價漲跌幅度區間的預測方法。最後，本研究針對三家公司的股價資料進行實證分析，研究結果顯示股價漲跌預測的正確率都達82.13%以上，三家公司股票價格的漲跌幅度區間涵蓋真實股價的平均預測正確率都高達83%以上。由此可知，本研究所提出的模糊推論預測系統不僅具有學理基礎，同時能夠有效預測股票的漲跌趨勢及漲跌幅區間，對投資人具有實務應用的價值與貢獻。

關鍵詞：金融科技、股價漲跌預測、機器學習演算法、模糊推論預測系統

* 本文通訊作者。電子郵件信箱：ctchen@nuu.edu.tw
2018/11/08 投稿；2018/12/21 修訂；2019/01/20 接受

Chen, C.T. and Sie, J.H. (2019), 'A study on application of machine learning and fuzzy inference for the prediction of stock price', *Journal of Information Management*, Vol. 26, No. 2, pp. 153-178.

A Study on Application of Machine Learning and Fuzzy Inference for the Prediction of Stock Price

Chen-Tung Chen*

Department of Information Management, National United University

Jheng-Han Sie

Department of Information Management, National United University

Abstract

Purpose—The purpose of this study is to propose a fuzzy inference forecasting system to predict the variation of stock price of each company and the stock price fluctuation range.

Design/methodology/approach—In this study, four machine learning algorithms are used to predict the stock prices and select the important technical indicators. And then, this study applies fuzzy inference to construct a fuzzy inference system to predict stock price fluctuation based on the critical technical indices.

Findings—The results of case study showed that the accuracy of the stock price fluctuation is more than 82.13% for three companies. In addition, this study also proposes a new method for predicting the stock price fluctuation range. The research results show that the prediction accuracy rate of stock price intervals are more than 83% for three companies.

Research limitations/implications—This study focused on the numeral data of technical indicators but not non-numeral data in the fuzzy inference system to predict stock price fluctuation. Future research can combine different data types to construct the prediction model of stock price.

* Corresponding author. Email: ctchen@nuu.edu.tw
2018/11/08 received; 2018/12/21 revised; 2019/01/20 accepted

Practical implications — According to the fuzzy inference forecasting system, investors can use only the five technical indices to predict the fluctuation trend and the interval of the stock price. The prediction accuracy rate of stock price intervals are good enough for investors.

Originality/value — The fuzzy inference forecasting system proposed in this study not only have the academic values, but also can effectively predict the fluctuation trend and the interval of the stock price, and the contributions of practical applications for investors.

Keywords: FinTech, stock price forecasting, machine learning, fuzzy inference forecasting system

壹、前言

隨著人工智慧與網路資訊技術的快速發展，應用互聯網技術於金融領域的金融科技（Financial Technology; FinTech）乃因應而生（曹磊 & 錢海利 2016；Shim & Shin 2016）。因此，利用機器人處理龐大的財務資料，以協助投資人進行理財投資分析已是金融科技領域的重要方向。機器學習（machine learning）即是以大量資料為基礎，透過演算法自動化的學習辨識複雜的資料樣式，進而做出智慧的決策（Han et al. 2011）。

針對股價預測問題，傳統的研究大都利用多變量迴歸（Li et al. 2015; Zhang et al. 2018），一般化自動化迴歸條件變量（Yu et al. 2018; Kim & Won 2018）等方法進行預測。隨著資訊科技的快速發展，有效應用資訊科技在短時間內考量眾多指標，進行股價預測乃是投資者關心的課題。因此，利用智慧數據分析方法以預測股價乃是金融科技領域的重要議題之一。然而，股票市場受到許多不確定性因素的影響而不易準確預測，因此股價的預測被認為是極具挑戰性的課題。為此，有許多學者利用機器學習演算法針對股價的預測進行研究。Kara 等（2011）利用類神經網路（artificial neural networks; ANN）和支援向量機（support vector machines; SVM），使用 10 個技術指標預測伊斯坦布爾證券交易所（ISE）國家 100 指數的表現。Laboissiere 等（2015）針對聖保羅證券交易所的巴西配電公司股價，利用類神經網路進行股價預測，並以平均絕對誤差（MAE）、平均絕對百分比誤差（MAPE）和均方根誤差（RMSE）進行效能評估。Patel 等（2015a）利用類神經網路、支援向量機、隨機森林和單純貝氏等四種方法，針對印度股市股價漲跌進行預測。因此，使用機器學習演算法進行股價漲跌的預測，以提供投資的決策與分析乃是重要的發展趨勢。然而，影響股價漲跌預測的技術指標及因素甚多，且由於指標的特性及投資者的主觀認知，都會影響股價漲跌預測的準確性。因此，利用模糊語意來表示投資者語意判斷的認知，將可降低股價漲跌預測過程的不確定性與模糊性，進而提供投資人更具彈性的決策分析結果（Chen & Chen 2015; Dash & Dash 2016）。因此，本研究以機器學習演算法為基礎，藉由重要技術指標的篩選並結合模糊推論方法，建構一個股價漲跌的預測系統，藉以預測股票的漲跌趨勢及漲跌幅度的區間，以利投資人進行決策與分析。

貳、文獻探討

一、機器學習演算法

(一) 支援向量機 (support vector machine; SVM)

Cortes 與 Vapnik (1995) 提出支援向量演算法，Vapnik 等 (1997) 將支援向量機應用於解決多維函數的估計問題。SVM 是在高維度空間中尋找一個分類的超平面，將不同類別的資料分開，使得不同類別點之間的間隔最大。之後，SVM 也應用在各個領域上應用包括文字分類、影像辨識、財務預測 (Kang et al. 2015)。

(二) 決策樹 (decision tree; DT)

決策樹是根據目標設定來選擇分枝變數，並以樹枝狀的層級架構呈現結果 (簡禎富 & 許嘉裕 2014)。它是利用一個樹狀結構來處理分類問題，且視覺效果非常的豐富 (Keramati et al. 2014)。決策樹的內部節點代表其測試的屬性，而每一個分枝表示可能的測試結果，葉節點則代表類別標籤。

(三) K 最近鄰居法 (K-nearest neighbor; KNN)

K 最近鄰居法是一個非參數的分類方法 (Adeniyi et al. 2016)。KNN 是利用一個測試數據與一組訓練數據的相似度計算結果，根據最近的鄰居來進行分類。若有 K 個鄰居需要考慮時，即為 K 最近鄰居法 (Han et al. 2011; Adeniyi et al. 2016)。KNN 的優點在於簡單的概念算法直觀且易於實現，但在樣本資料數量龐大時，其距離的計算次數將會過多是主要的缺點。

(四) 倒傳遞類神經網路 (back propagation neural network; BPNN)

類神經網路 (artificial neural network; ANN) 是一種非線性的系統，由多個神經元模擬生理學模型的架構 (Valdez et al. 2014; Yu et al. 2016)。一個類神經網路是由許多個人工神經元與其連結所組成，並且可以組成各種網路模式。其中，倒傳遞類神經網路 (BPNN) 是一種具有學習能力的多層網路結構，也是目前應用最為廣泛的一種類神經網路演算法 (Wang et al. 2012; Mo et al. 2016)。

二、股價預測分析

近年來，隨著金融科技的崛起，智慧理財或投資分析已成為重要的發展趨勢 (Demirbag et al. 2016)。其中，股票市場更是投資人金融理財重要的管道之一。然而，由於影響股票市場的因素甚多，且充滿許多人為與經濟環境的不確定性，使得股票市場預測被認為是極具挑戰性的議題。本研究彙整機器學習方法於股價預測的相關文獻，如表 1 所示。同時，本研究彙整 45 個影響股價預測的技術指

標，如表 2 所示。

表 1：股價預測的相關文獻

作者名稱	預測方法
Atsalakis and Valavanis (2009)	Neuro-Fuzzy
鄭健毅 (2010)	Support Vector Regression、Artificial Neural Network
Hadavandi et al. (2010)	Genetic Fuzzy、Artificial Neural Network
陳鄺貞 (2011)	Back Propagation Neural Network
Kara et al. (2011)	Artificial Neural Network、Support Vector Machine
Chang (2012)	Evolving Partially Connected Neural Networks
Lahmiri (2014)	Support Vector Machine
Laboissiere et al. (2015)	Artificial Neural Network
Patel et al. (2015a)	Artificial Neural Network、Support Vector Machine、Random Forest、Naive-Bayes
Patel et al. (2015b)	Support Vector Regression + Artificial Neural Network、Support Vector Regression + Random Forest、Support Vector Regression + Support Vector Regression
Wang et al. (2016)	Support Vector Regression + Principal Component Analysis + Beam Search Optimization
Lahmiri (2016)	Particle Swarm Optimization+ Back Propagation Neural Network
Chen and Hao (2017)	Feature Weighted Support Vector Machine、Feature Weighted K-Nearest Neighbor

資料來源：本研究彙整

表 2：技術指標

技術指標名稱 \ 文獻編號	A	B	C	D	E	F	G	H
6 日相對強弱指標 (RSI)	v			v		v		v
12 日相對強弱指標 (RSI)				v		v		v
相對強弱指標 (RSI)	v		v		v		v	
快速隨機指標 (K)	v		v	v	v	v	v	v
慢速隨機指標 (D)	v		v	v	v	v	v	v
移動平均收斂發散指標 (MACD)	v		v	v	v		v	
5 日移動平均線 (MA)				v	v	v		v
6 日移動平均線 (MA)				v		v		
10 日移動平均線 (MA)			v	v	v	v	v	v
20 日移動平均線 (MA)				v	v	v		v
10 日加權移動平均線			v				v	
開盤價		v						v
收盤價		v						v
加權收盤價 (WC)						v		
最高價		v				v		v
最低價		v				v		v
平均價						v		
典型價格 (TP)						v		
交易量					v	v		v
交易額								v
變化率						v		v
PVC						v		
漲跌額								v
動量指標 (MTM)	v		v			v	v	
威廉指標 (W%R)	v		v	v		v	v	
5 日乖離率 (BIAS)				v				v
6 日乖離率 (BIAS)								v
12 日乖離率 (BIAS)	v			v				
24 日乖離率 (BIAS)								v

DIF					v			v
技術指標差異 (Δ)				v				
能量潮指標 (OBV)	v				v	v		
A/D			v			v	v	
順勢指標 (CCI)			v				v	
心理線 (PSY)	v				v			v
停損點轉向操作系統 (SAR)	v							
趨向指標 (DMI)	v				v			
逆勢操作系統 (CDP)	v							
加速量指標 (ACC)						v		
蔡金擺動指標 (Chaikin)						v		
蔡金波動性指標 (Chaikin)						v		
負量指標 (NVI)						v		
正量指標 (PVI)						v		
布林帶指標 (BB)						v		
威廉多空力度 (WAD)						v		
A：鄭健毅 (2010)；B：Hadavandi (2010)；C：Kara et al. (2011)；D：Chang (2012)；E：廖日昇 (2012)；F：Lahmiri (2014)；G：Patel et al. (2015a)；H：Wang et al. (2016)								

資料來源：本研究整理

三、模糊推論

(一) 模糊推論的概念

事實上，人們常常不易使用二元的明確值來表達日常生活中，對於許多事物或問題的感受。利用模糊集合的邏輯觀念，可以更彈性地表達人為的主觀意見（李允中等 2008；Lincy & John 2016）。模糊推論（Fuzzy Inference）是利用隸屬函數計算輸入值隸屬於每個規則的隸屬程度，綜合所有規則的隸屬度後，可得到適當的推論結果。

一般而言，模糊推論的規則表示如下（李允中等 2008；Lincy & John 2016）：

$$IF \ x_1 \text{ is } X_1^i \text{ and } x_2 \text{ is } X_2^i \text{ and...and } x_n \text{ is } X_n^i \text{ THEN } y \text{ is } Y_i \quad (1)$$

其中， $X_1^i, X_2^i, \dots, X_n^i$ 和 $Y_i (i = 1, 2, \dots, m)$ 是模糊集合， $x_t (t = 1, 2, \dots, n)$ 為第 t 個模

糊集合的輸入值， y 為輸出值變數。

(二) 定義隸屬函數方法

由於資料的前處理及離散化，將影響資料分析的效果與精確性，因此必須適當的將資料模糊化以建構隸屬函數（李允中等 2008）。然而，若是由使用者或是專家自行決定隸屬函數，則較無法真實的呈現資料散佈的情況，也無法客觀的進行資料分類（Escobar et al. 2013; Chourmouziadis & Chatzoglou 2016; Ghadimi et al. 2017）。因此，應利用系統化的方法建構資料屬性的隸屬函數。有關產生隸屬函數中心點的主要方法包括 K-means 分群法（Arora & Varshney 2016）、模糊 C-means 分群法（fuzzy C-means clustering）（Kesemen et al. 2016）、四分位數法及「平均值加減標準差」的方式（Anderson et al. 2014）。

(三) 模糊推論的運算

當推論法則並非只有一個規則時，一般會採用 Mamdani 的 min-max 方法進行推論規則的整合（李允中等 2008; Lincy & John 2016）。利用 min-max 的方法在前提部分所觸發隸屬程度的交集值為 α_i ，計算如下所示（李允中等 2008）：

$$\alpha_i = \min\{\mu_{x_1^i}(x_1), \mu_{x_2^i}(x_2), \dots, \mu_{x_n^i}(x_n)\} \quad (2)$$

其中， $\mu_{x_t^i}(x_t)$ 為輸入變數 x_t 隸屬於 X_t^i ($t = 1, 2, \dots, n; i = 1, 2, \dots, m$) 的隸屬程度， α_i 為第 i 條規則前提部分的交集值。

第 i 條規則的輸出結果計算如下：

$$B_i(y) = \min\{\alpha_i, Y_i(y)\}, y \in Y \quad (3)$$

其中， $B_i(y)$ 為第 i 條規則的輸出結果， $Y_i(y)$ 為輸出值 y 隸屬第 i 條規則後件 Y_i 的程度， Y 為語意變數值域。

整合所有規則的輸出結果，計算如下所示（李允中等 2008）：

$$B(y) = \max\{B_i(y)\}, i = 1, 2, \dots, m \quad (4)$$

其中， $B(y)$ 為推論結果的隸屬函數。

(四) 解模糊化的方法

一般而言，解模糊化的方法主要有三種方式，分別為離散式重心法、連續型重心法、及最大均值法（李允中等 2008）。本研究將推論結果的隸屬函數區分成 K 個區間，並使用離散式重心法進行解模糊化，計算如下（李允中等 2008; Lincy & John 2016）：

$$y^* = \frac{\sum_{k=1}^K y_k B(y_k)}{\sum_{k=1}^K B(y_k)} \quad (5)$$

其中， y^* 為解模糊化的數值。

四、預測效能的評估

針對類別資料的預測效能評估常用的方法是二元分類的混亂矩陣（confusion matrix）分析（Patel et al. 2015a）。然而，在預測結果有多種分類的情況下，必須使用多分類的混亂矩陣如表 3 所示（Deng et al. 2016）。

表 3：多分類的混亂矩陣

		預測值		
		A_1	A_j	A_n
真實值	A_1	N_{11}	N_{1j}	N_{1n}
	A_i	$\vdots$ N_{i1} $\vdots$	$\vdots$ N_{ij} $\vdots$	$\vdots$ N_{in} $\vdots$
	A_n	N_{n1}	N_{nj}	N_{nn}

整體預測的準確度（accuracy）的計算公式如下：

$$Accuracy = \frac{\sum_{i=1}^n N_{ii}}{\sum_{i=1}^n \sum_{j=1}^n N_{ij}} \quad (6)$$

其中， N_{ii} 代表真實值為第 i 類且預測值為第 i 類的個數， N_{ij} 代表真實值為第 i 類，而預測值為第 j 類的個數。

第 i 類的預測精確度（ P_i ）及召回率（ R_i ）的計算方式為：

$$P_i = \frac{N_{ii}}{\sum_{k=1}^n N_{ki}} \quad (7)$$

$$R_i = \frac{N_{ii}}{\sum_{k=1}^n N_{ik}} \quad (8)$$

其中， N_{ki} 代表真實值為第 k 類而預測值為第 i 類的個數， N_{ik} 代表真實值為第 i

類，而預測值為第 k 類的個數。

參、重要指標之篩選

本研究彙整文獻中出現次數較高的 18 個技術指標（分別為 A1 至 A18），如表 4 所示。「股價漲跌」是本研究的目標變數為 O19，以預測股價的「上漲」、「持平」與「下跌」。本研究利用技術指標的篩選方式及機器學習演算法，找出重要的技術指標，並選定股價變化較大的「大立光」、股價變化較一般的「台積電」以及股價變化較小的「鴻海」等三家上市公司為研究對象。資料收集的時間為 2014/10/13 至 2017/06/05，資料共有 646 筆，其中以 2016/12/30 至 2017/06/05 共 100 筆作為預測的資料。

表 4：技術指標的名稱及代號

A1. 開盤價差（元）	A2. 最高價差（元）	A3. 最低價差（元）	A4. 12 日相對強弱指標（RSI）
A5. 5 日平均線（MA）	A6. 10 日平均線（MA）	A7. 20 日平均線（MA）	A8. 12 日威廉指標（W%R）
A9. 6 日乖離率（BIAS）	A10. 12 日乖離率（BIAS）	A11. 6 日心理線（PSY）	A12. 9 日快速隨機指標（K）
A13. 9 日慢速隨機指標（D）	A14. 9 日移動平均收斂發散指標（MACD）	A15. 成交量（千股）	A16. 10 日動量指標（MTM）
A17. A/D	A18. 10 日順勢指標（CCI）		O19. 股價漲跌

資料來源：本研究整理

本研究首先計算每個技術指標與三家公司股價漲跌的皮爾森相關係數（Anderson et al. 2014），並挑選出相關係數較高的前 10 名技術指標進行股價漲跌的預測，並逐漸減少至 5 個技術指標。同時計算在不同指標個數下，不同公司在不同機器學習演算法的預測準確度，結果如表 5 至表 7 所示。針對「大立光」、「台積電」、「鴻海」的分析結果得知，利用 SVM 預測股價漲跌的準確率最高。當技術指標個數逐步減少到 5 個時，利用不同機器學習演算法進行三家公司股價漲跌預測準確率均達 80% 以上。「大立光」所選出的 5 個重要的技術指標分別為 A12（9 日快速隨機指標）、A4（12 日相對強弱指標）、A17（買賣力指標 A/D）、A3（最低價差）、A9（6 日乖離率）。「台積電」所選出的 5 個重要的技術指標分別為 A12（9 日快速隨機指標）、A3（最低價差）、A17（買賣力指標 A/D）、A2

(最高價差)、A4 (12 日相對強弱指標)。「鴻海」所選出的 5 個重要的技術指標分別為 A17 (買賣力指標 A/D)、A4 (12 日相對強弱指標)、A13 (9 日慢速隨機指標)、A3 (最低價差)、A2 (最高價差)。因此，本研究分別針對「大立光」、「台積電」、「鴻海」所選出的 5 個重要的技術指標，做為股價漲跌預測的指標，並依此進行預測系統的開發。

表 5：「大立光」在不同技術指標個數的準確度

方法 指標個數	SVM	KNN	BPNN	DT
10	88.45%	85.45%	85.71%	85.45%
9	87.64%	84.55%	84.79%	84.54%
8	87.18%	83.18%	84.79%	84.54%
7	86.64%	82.45%	84.43%	83.63%
6	86.36%	81.64%	83.63%	82.45%
5	84.36%	80.00%	83.51%	81.81%
4	76.36%	75.45%	78.89%	76.36%

表 6：「台積電」在不同技術指標個數的準確度

方法 指標個數	SVM	KNN	BPNN	DT
10	88.18%	82.72%	85.32%	84.54%
9	86.36%	82.72%	84.79%	83.63%
8	85.45%	81.81%	83.48%	82.72%
7	85.45%	80.90%	83.48%	82.72%
6	84.54%	80.90%	81.35%	80.90%
5	84.54%	80.00%	81.35%	80.90%
4	77.72%	73.63%	77.83%	74.54%

表 7：「鴻海」在不同技術指標個數的準確度

方法 指標個數	SVM	KNN	BPNN	DT
10	88.77%	84.69%	87.15%	84.54%
9	86.34%	82.65%	84.40%	82.72%

8	84.69%	82.65%	83.51%	82.72%
7	83.67%	81.63%	83.51%	81.81%
6	82.65%	80.90%	82.23%	81.81%
5	81.81%	80.00%	81.68%	80.27%
4	74.48%	72.44%	77.83%	75.45%

肆、模糊推論預測系統

本研究主要是藉由指標的篩選，並利用機器學習演算法進行預測效能分析，以選出重要的技術指標，進而建構一個模糊推論系統，以預測股價的漲跌程度以及漲跌幅的區間。本研究提出的預測系統的概念架構如圖 1 所示。

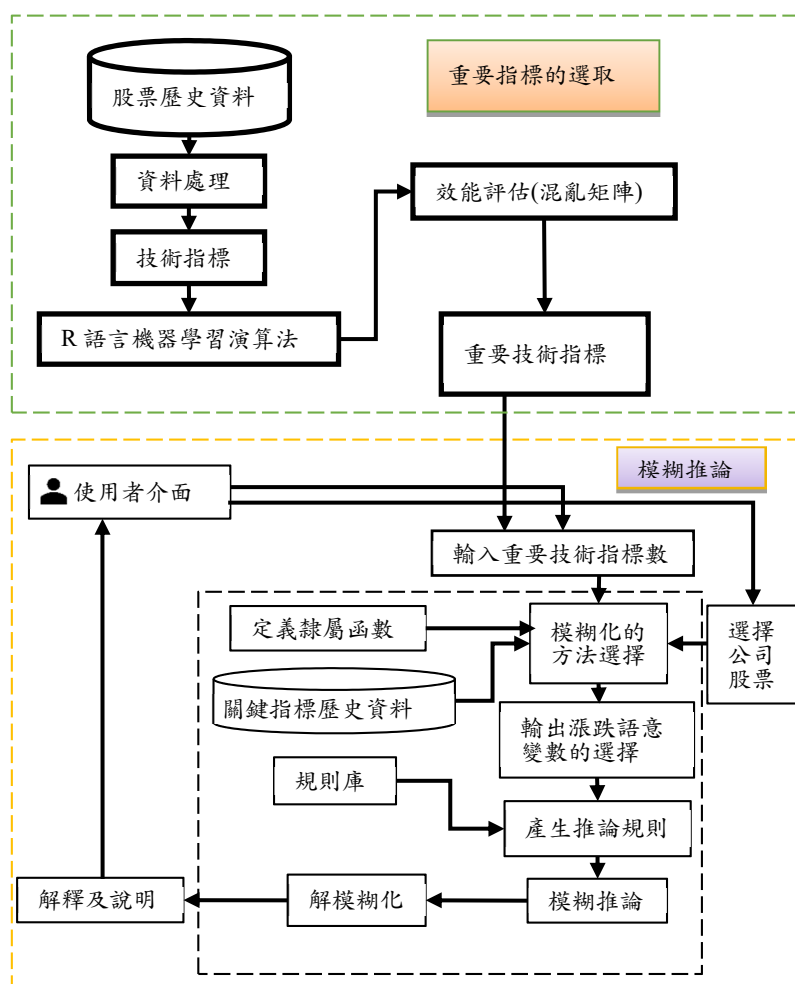


圖 1：預測系統的概念架構

一、股價漲跌的推論預測

(一) 定義技術指標的語意變數

本研究利用六種資料模糊化方法來決定語意變數之隸屬函數中心點，模糊化方法分別為四分位數、平均值加減 1 倍標準差 ($\bar{x} \pm s$)、平均值加減 0.5 倍標準差 ($\bar{x} \pm 0.5s$)、平均值加減 1.5 倍標準差 ($\bar{x} \pm 1.5s$)、K-mean 以及 Fuzzy C-mean，並將重要技術指標的資料分為「低」、「中」、「高」三個語意變數，中心點為 C_1 、 C_2 、 C_3 ，其隸屬函數表示如圖 2 所示。

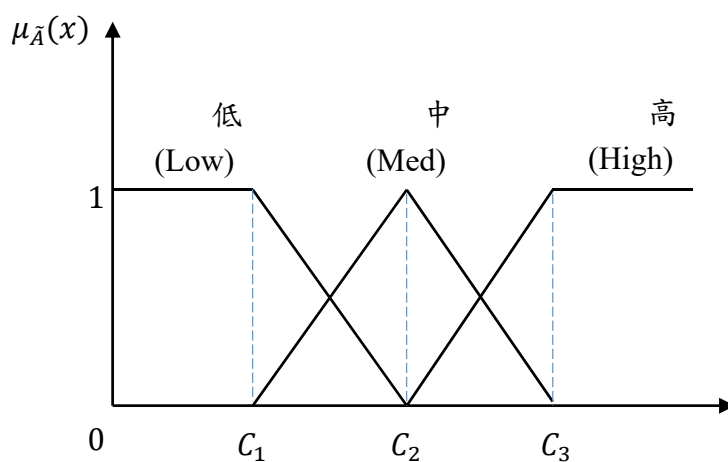


圖 2：語意變數的隸屬函數

(二) 定義輸出結果的語意變數

本研究使用兩種不同類型的語意變數來表達股價上漲的可能性，分別為「五點語意變數」及「七點語意變數」，表達方式如表 8 所示。

表 8：兩種類型之語意變數

類型	股價上漲的可能性	隸屬函數
五點	極低 (C_1)，低 (C_2)，不漲不跌 (C_3)，高 (C_4)，極高 (C_5)	圖 3
七點	極低 (C_1)，低 (C_2)，稍低 (C_3)，不漲不跌 (C_4)，稍高 (C_5)，高 (C_6)，極高 (C_7)	圖 4

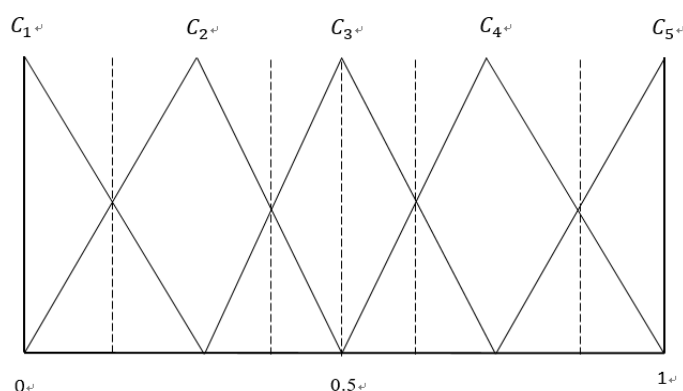


圖 3：五點語意變數的隸屬函數

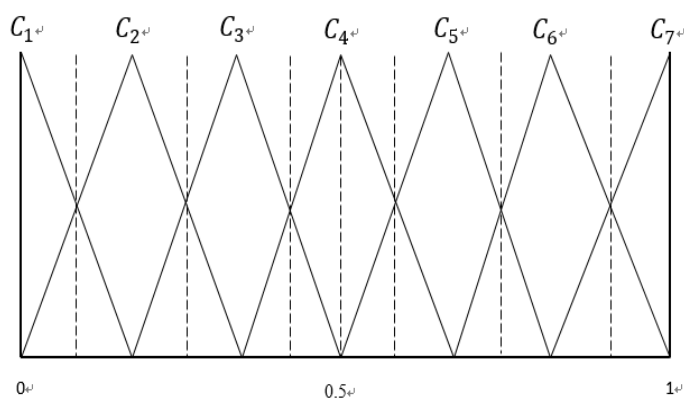


圖 4：七點語意變數的隸屬函數

（三）定義規則庫

本研究利用技術指標隸屬的語意變數「低」、「中」、「高」的個數來建立規則庫，定義的五點與七點語意的規則庫如表 9 與表 10 所示。

（四）推論的結果

本研究利用 min-max 的方法進行規則整合及的推論計算，並使用離散式重心法將推論結果的隸屬函數加以解模糊化（李允中等 2008；Lincy & John 2016）。若是，解模糊化的數值（ y^* ）大於 0.5，則預測股價漲跌的狀態為「漲」。反之，則預測股價漲跌的狀態為「跌」。

表 9：五點語意規則庫

語意變數 規則	低（個數）	中（個數）	高（個數）	輸出上漲的可 可能性程度
規則 1	0	0	5	極高
規則 2	1	0	4	極高
規則 3	0	1	4	極高
規則 4	2	0	3	高
規則 5	1	1	3	高
規則 6	2	1	2	不漲不跌
規則 7	0	4	1	不漲不跌
規則 8	1	2	2	不漲不跌
規則 9	0	5	0	不漲不跌
規則 10	1	4	0	低
規則 11	2	3	0	低
規則 12	3	2	0	極低
規則 13	4	1	0	極低
規則 14	5	0	0	極低

表 10：七點語意規則庫

語意變數 規則	低（個數）	中（個數）	高（個數）	輸出上漲的可 可能性程度
規則 1	0	0	5	極高
規則 2	1	0	4	高
規則 3	0	1	4	高
規則 4	2	0	3	稍高
規則 5	1	1	3	稍高
規則 6	0	3	2	不漲不跌
規則 7	2	1	2	不漲不跌
規則 8	0	4	1	不漲不跌
規則 9	0	5	0	不漲不跌
規則 10	1	4	0	稍低
規則 11	2	3	0	稍低
規則 12	3	2	0	低
規則 13	4	1	0	低
規則 14	5	0	0	極低

二、股價漲跌幅度區間的預測

當輸出結果為「五點語意變數」，且推論上漲可能性的解模糊化值大於 0.5 時，漲幅區間的預測為 $y \in [\hat{y}_1, \hat{y}_2]$ ，其中， $\hat{y}_1 = \min\{y_1, y_2\}$ ， $\hat{y}_2 = \max\{y_1, y_2\}$ 。 y_1 及 y_2 的計算公式如下：

$$y_1 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{高}}(x) * \text{昨日收盤價} \quad (9)$$

$$y_2 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{極高}}(x) * \text{昨日收盤價} \quad (10)$$

其中， x 為推論結果解模化的數值。

當解模糊化值小於 0.5 時，則跌幅區間為 $y \in [\hat{y}_1, \hat{y}_2]$ ，其中， $\hat{y}_1 = \min\{y_1, y_2\}$ ， $\hat{y}_2 = \max\{y_1, y_2\}$ 。 y_1 及 y_2 的計算公式如下：

$$y_1 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{低}}(x) * \text{昨日收盤價} \quad (11)$$

$$y_2 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{極低}}(x) * \text{昨日收盤價} \quad (12)$$

其中， x 為推論結果解模化的數值。

當輸出結果為「七點語意變數」，推論上漲可能性的解模糊化值大於 0.5 時，漲幅區間的預測為 $y \in [\hat{y}_1, \hat{y}_2]$ ，其中， $\hat{y}_1 = \min\{y_1, y_2\}$ ， $\hat{y}_2 = \max\{y_1, y_2\}$ 。 y_1 及 y_2 的計算如下：

1. 當解模糊化值隸屬於「稍高」及「高」，則計算如下：

$$y_1 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{稍高}}(x) * \text{昨日收盤價} \quad (13)$$

$$y_2 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{高}}(x) * \text{昨日收盤價} \quad (14)$$

2. 當解模糊化值隸屬於「高」及「極高」，則計算如下：

$$y_1 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{高}}(x) * \text{昨日收盤價} \quad (15)$$

$$y_2 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{極高}}(x) * \text{昨日收盤價} \quad (16)$$

若是輸出結果為「七點語意變數」，且解模糊化值小於 0.5 時，跌幅區間的預測為 $y \in [\hat{y}_1, \hat{y}_2]$ ，其中， $\hat{y}_1 = \min\{y_1, y_2\}$ ， $\hat{y}_2 = \max\{y_1, y_2\}$ 。 y_1 及 y_2 的計算如下：

1. 當解模糊化值隸屬於「稍低」及「低」，則計算如下：

$$y_1 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{稍低}}(x) * \text{昨日收盤價} \quad (17)$$

$$y_2 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{低}}(x) * \text{昨日收盤價} \quad (18)$$

2. 當解模糊化值隸屬於「低」及「極低」，則計算如下：

$$y_1 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{低}}(x) * \text{昨日收盤價} \quad (19)$$

$$y_2 = \frac{(x-0.5)}{0.5} * 7\% * \mu_{\text{極低}}(x) * \text{昨日收盤價} \quad (20)$$

三、預測系統的步驟

本研究所提出的模糊推論預測系統的運作步驟如下：(1)選擇公司股票系統將會顯示該公司重要技術指標的說明，如圖 5 所示。(2)選擇模糊化的方法以及選擇輸出結果的語意變數，如圖 6 所示。(3)顯示每個技術指標的隸屬函數中心點，並且輸入重要的技術指標數值進行模糊推論分析，如圖 7 所示。(4)顯示隸屬程度值，以及產生模糊規則，並進行解模糊化，如圖 8 所示。(5)進行解模糊化提供不同語意變數的結果，提供漲跌可能性的值，如圖 9 所示。

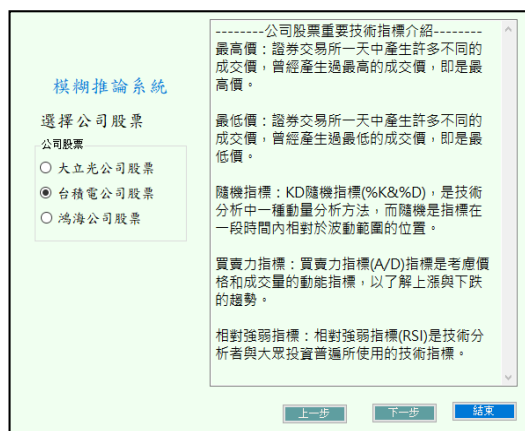


圖 5：選擇公司股票畫面

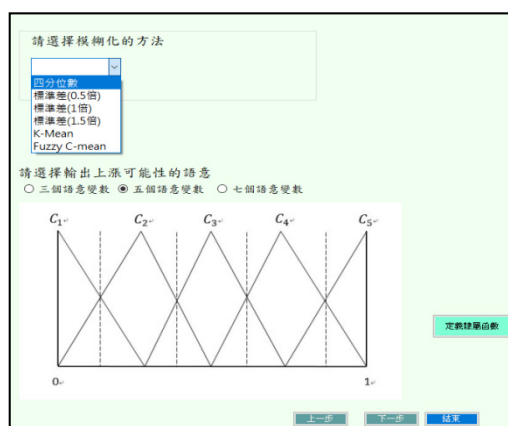


圖 6：選擇模糊化方法以及語意變數

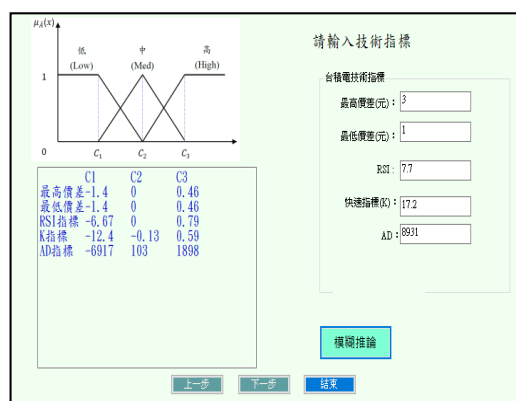


圖 7：輸入技術指標之畫面



圖 8：產生模糊規則

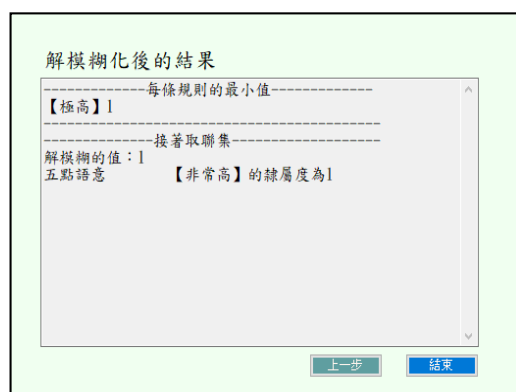


圖 9：解模糊化之畫面

伍、實證分析

本研究針對「大立光」、「台積電」、「鴻海」三家公司的股價資料進行實證分析。資料收集的時間為 2014/10/13 至 2017/06/05，資料共有 646 筆。其中，同時以 2016/12/30 至 2017/06/05 從中隨機抽 80 筆作為預測的資料。

一、分析流程

本研究利用六種資料模糊化方法定義重要技術指標的隸屬函數中心點，並使用兩種輸出上漲可能性的語意變數，進行模糊推論並且以解模糊化的數值，計算股價漲跌預測的準確率。其次，利用解模糊化的數值進行股票價格的漲跌幅度區間的預測。

二、資料分析結果

利用六種模糊化方法以及兩種輸出結果的語意變數共 12 種不同組合進行模糊推論，以「大立光」、「台積電」、「鴻海」三家公司的股票資料（2016/12/30 至 2017/06/05），從中隨機抽 80 筆進行 10 次預測並計算平均準確率，結果如表 11 至表 13 所示。由此結果可知，在「平均值加減 0.5 倍標準差」($\bar{x} \pm 0.5s$) 的模糊化方法以及上漲可能性為「七點語意變數」的組合下，三家公司都有最高的預測準確率。因此，本研究利用「平均值加減 0.5 倍標準差」($\bar{x} \pm 0.5s$) 的模糊化方法以及上漲可能性的「七點語意變數」，隨機抽 10 筆資料，進行 10 次的漲跌幅區間預測，三家公司的平均的準確率均達 83% 以上，如表 14 所示。

表 11：「大立光」在各方法及語意預測股價漲跌之平均準確率

語意變數 \ 方法	四分位數	標準差 (0.5 倍)	標準差 (1 倍)	標準差 (1.5 倍)	K-mean	Fuzzy C-mean
五點語意變數	77.50%	80.63%	66.13%	58.63%	75.63%	78.00%
七點語氣變數	78.00%	82.75%	69.50%	60.00%	77.13%	78.88%

表 12：「台積電」在各方法及語意預測股價漲跌之平均準確率

語意變數 \ 方法	四分位數	標準差 (0.5 倍)	標準差 (1 倍)	標準差 (1.5 倍)	K-mean	Fuzzy C-mean
五點語意變數	72.13%	77.50%	64.88%	48.00%	71.38%	76.38%
七點語氣變數	82.88%	85.38%	68.88%	53.00%	80.63%	84.63%

表 13：「鴻海」在各方法及語意預測股價漲跌之平均準確率

語意變數 \ 方法	四分位數	標準差 (0.5 倍)	標準差 (1 倍)	標準差 (1.5 倍)	K-mean	Fuzzy C-mean
五點語意變數	77.88%	79.88%	70.88%	55.75%	70.75%	70.95%
七點語氣變數	79.75%	82.13%	69.75%	58.25%	73.88%	79.63%

表 14：三家公司股價漲跌幅區間預測的平均準確度

	次數										
公司	1	2	3	4	5	6	7	8	9	10	平均
大立光	80%	80%	80%	90%	80%	90%	90%	80%	90%	70%	83%
台積電	80%	90%	90%	90%	80%	90%	90%	80%	90%	70%	85%
鴻海	90%	90%	80%	90%	80%	70%	90%	80%	80%	80%	83%

三、分析結果說明

藉由「資料模糊化方法」以及「上漲可能性輸出結果語意變數」的 12 種組合分析結果發現，使用「平均值加減 0.5 倍標準差」的資料模糊化方法，以及上漲可能性輸出結果為「七點語意變數」時，三家公司股價漲跌預測的準確率都達 82.13%以上。此外，本研究所提出的股價漲跌幅度區間之預測方法，判斷股價漲跌區間範圍涵蓋真實股價的平均準確率也都達 83%以上。因此，本研究所提出的股價漲跌模糊推論預測系統，以及漲跌幅度區間的預測方法具有實務價值與貢獻，能夠有效協助投資者進行股價預測及分析。

陸、結論

近年來，隨著人工智慧與資訊科技的快速發展，藉由機器學習方法的應用，以進行股價的預測更是目前金融科技領域重要的議題之一。為此，本研究針對三家公司股票分別利用機器學習演算法，逐步刪減技術指標的個數，進行準確率分析，以篩選出重要的技術指標。此外，為了考量資料的屬性及人為主觀判斷的模糊性，本研究以重要技術指標為基礎，應用模糊推論方法開發一個模糊推論預測系統，以進行股價漲跌的預測。同時，本研究亦提出股價漲跌幅區間的預測方法。本研究使用了六種資料模糊化的方法，將每個技術指標分成三個語意變數群集，且利用兩種不同的語意變數來表達推論輸出的結果。經由實證分析發現，本研究提出的「股價漲跌」及「漲跌幅度區間」的預測準確度均相當高且具有相當程度的穩定性。綜言之，本研究建構的模糊推論預測系統利用模糊語意表達股價漲跌的可能性，讓投資者可以更容易明瞭，且符合現實的情境。此外，利用股價漲跌幅度區間的預測，能夠精準預測股價的變化區間，藉此提供投資者具體的參考價值。整體而言，本研究所提出的分析方法與模糊推論預測系統，不僅具有學理基礎，同時具有實務的參考價值。

本研究目前僅針對三家公司的股價的數值型指標資料進行比較分析，未來可考慮多家公司或產業，並整合非數值型的指標資料，例如市場口碑、新聞事件觀

感、公司經營者形象等等，以進行股價漲跌變化及漲跌區間的預測，進而提高預測的準確率。此外，也可增加不同的機器學習演算法，例如基因演算法、粒子群演算法、螞蟻演算法、深度學習演算法等等，並將模糊推論系統建構成一個即時動態更新資料的系統，以便提供投資者即時的投資決策資訊，強化實務上的應用。

參考文獻

- 李允中、王小潘、蘇木春 (2008)，*模糊理論及其應用 (修訂版)*，全華圖書股份有限公司，新北市。
- 曹磊、錢海利 (2016)，*FinTech 金融科技革命*，商周出版家庭傳媒城邦分公司，台北。
- 陳鄺貞 (2011)，『以財務指標及技術指標建構股價預測模型－類神經網路模型之應用』，出版碩士論文，國立台北大學國際財務金融研究所，台北。
- 鄭健毅 (2010)，『應用 SVR 支援向量迴歸模式來進行電子產業股價預測』，出版碩士論文，明新科技大學工業工程與管理研究所，新竹縣。
- 廖日昇 (2012)，*我的第一本圖解技術分析*，創智文化有限公司，台北市。
- 簡禎富、許嘉裕 (2014)，*資料挖礦與大數據分析*，前程文化事業有限公司，新北市。
- Atsalakis, G.S. and Valavanis, K.P. (2009), 'Forecasting stock market short-term trends using a neuro-fuzzy based methodology', *Expert Systems with Applications*, Vol. 36, No. 7, pp. 10696-10707.
- Adeniyi, D.A., Wei, Z. and Yongquan, Y. (2016), 'Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method', *Applied Computing and Informatics*, Vol. 12, No. 1, pp. 90-108.
- Arora, P. and Varshney, S. (2016), 'Analysis of K-means and K-medoids algorithm for big data', *Procedia Computer Science*, Vol. 78, pp. 507-512.
- Anderson, D.R., Sweeney, D.J., Williams, T.A., Camm, J.D. and Cochran, J.J. (2014), *Statistics for Business and Economics 12e*, Annotated Education.
- Chen, M.Y. and Chen, B.T. (2015), 'A hybrid fuzzy time series model based on granular computing for stock price forecasting', *Information Sciences*, Vol. 294, pp. 227-241.
- Chang, P.C. (2012), 'A novel model by evolving partially connected neural network for stock price trend forecasting', *Expert Systems with Applications*, Vol. 39, No.1, pp. 611-620.
- Chen, Y. and Hao, Y. (2017), 'A feature weighted support vector machine and K-nearest

- neighbor algorithm for stock market indices prediction', *Expert Systems with Applications*, Vol. 80, pp. 340-355.
- Chourmouziadis, K. and Chatzoglou, P.D. (2016), 'An intelligent short term stock trading fuzzy system for assisting investors in portfolio management', *Expert Systems with Applications*, Vol. 43, pp. 298-311.
- Cortes, C. and Vapnik, V. (1995), 'Support-vector networks'. *Machine Learning*, Vol. 20, No.3, pp. 273-297.
- Dash, R. and Dash, P. (2016), 'Efficient stock price prediction using a self evolving recurrent neuro-fuzzy inference system optimized through a modified technique', *Expert Systems with Applications*, Vol. 52, pp. 75-90.
- Demirbag, M., McGuinness, M., Akin, A., Bayyurt, N. and Basti, E. (2016), 'The professional service firm (PSF) in a globalised economy: A study of the efficiency of securities firms in an emerging market', *International Business Review*, Vol. 25, No. 5, pp. 1089-1102.
- Deng, X., Liu, Q., Deng, Y. and Mahadevan, S. (2016), 'An improved method to construct basic probability assignment based on the confusion matrix for classification problem', *Information Sciences*, Vol. 340-341, 250-261.
- Escobar, A., Moreno, J. and Múnera, S. (2013), 'A technical analysis indicator based on fuzzy logic', *Electronic Notes in Theoretical Computer Science*, Vol. 292, pp. 27-37.
- Ghadimi, P., Dargi, A. and Heavey, C. (2017), 'Sustainable supplier performance scoring using audition check-list based fuzzy inference system: A case application in automotive spare part industry', *Computers & Industrial Engineering*, Vol. 105, pp. 12-27.
- Han, J., Kamber, M. and Pei, J. (2011), *Data Mining: Concepts and Techniques*. Elsevier.
- Hadavandi, E., Shavandi, H. and Ghanbari, A. (2010), 'Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting', *Knowledge-Based Systems*, Vol. 23, No. 8, pp. 800-808.
- Kara, Y., Boyacioglu, M.A. and Baykan, Ö.K. (2011), 'Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange', *Expert Systems with Applications*, Vol. 38, No. 5, pp. 5311-5319.
- Kang, S., Kang, P., Ko, T., Cho, S., Rhee, S.J. and Yu, K.S. (2015), 'An efficient and effective ensemble of support vector machines for anti-diabetic drug failure prediction', *Expert Systems with Applications*, Vol. 42, No. 9, pp. 4265-4273.
- Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M. and

- Abbasi, U. (2014), 'Improved churn prediction in telecommunication industry using data mining techniques', *Applied Soft Computing*, Vol. 24, pp. 994-1012.
- Kesemen, O., Tezel, Ö. and Özkul, E. (2016), 'Fuzzy c-means clustering algorithm for directional data (FCM4DD)', *Expert Systems with Applications*, Vol. 58, pp. 76-82.
- Kim, H.Y., and Won, C.H. (2018), 'Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCJ-type models', *Expert Systems with Applications*, Vol. 103, pp. 25-37.
- Laboissiere, L.A., Fernandes, R.A. and Lage, G.G. (2015), 'Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks', *Applied Soft Computing*, Vol. 35, pp. 66-74.
- Lahmiri, S. (2014), 'Entropy-based technical analysis indicators selection for international stock markets fluctuations prediction using support vector machines', *Fluctuation and Noise Letters*, Vol. 13, No. 2, pp. 1-16.
- Lahmiri, S. (2016), 'Intraday stock price forecasting based on variational mode decomposition', *Journal of Computational Science*, Vol. 12, pp. 23-27.
- Li, Johnny S.H., Ng, Andrew. W. and Chan, W.S. (2015), 'Managing financial risk in Chinese stock markets: Option pricing and modeling under a multivariate threshold autoregression', *International Review of Economics and Finance*, Vol. 40, pp. 217-230.
- Lincy, G.R.M. and John, C.J. (2016), 'A multiple fuzzy inference systems framework for daily stock trading with application to NASDAQ stock exchange', *Expert Systems with Applications: An International Journal*, Vol. 44, pp. 13-21.
- Mo, H., Wang, J. and Niu, H. (2016), 'Exponent back propagation neural network forecasting for financial cross-correlation relationship', *Expert Systems with Applications*, Vol. 53, pp. 106-116.
- Patel, J., Shah, S., Thakkar, P. and Kotecha, K. (2015a), '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.
- Patel, J., Shah, S., Thakkar, P. and Kotecha, K. (2015b), 'Predicting stock market index using fusion of machine learning techniques', *Expert Systems with Applications*, Vol. 42, No. 4, pp. 2162-2172.
- Shim, Y. and Shin, D.H. (2016), 'Analyzing China's Fintech Industry from the Perspective of Actor-Network Theory', *Telecommunications Policy*, Vol. 40, No. 2-3, pp. 168-181.
- Vapnik, V., Golowich, S.E. and Smola, A. (1997), 'Support vector method for function

- approximation, regression estimation, and signal processing', *Advances in Neural Information Processing Systems*, pp. 281-287.
- Valdez, F., Melin, P. and Castillo, O. (2014), 'Modular neural networks architecture optimization with a new nature inspired method using a fuzzy combination of particle swarm optimization and genetic algorithms', *Information Sciences*, Vol. 270, pp. 143-153.
- Wang, J., Hou, R., Wang, C. and Shen, L. (2016), 'Improved v-Support vector regression model based on variable selection and brain storm optimization for stock price forecasting', *Applied Soft Computing*, Vol. 49, pp. 164-178.
- Wang, J., Pan, H. and Liu, F. (2012), 'Forecasting crude oil price and stock price by jump stochastic time effective neural network model', *Journal of Applied Mathematics*, Vol. 2012, pp. 1-15.
- Yu, X., Ye, C. and Xiang, L. (2016), 'Application of artificial neural network in the diagnostic system of osteoporosis', *Neurocomputing*, Vol. 214, No. 1, pp. 376-381.
- Yu, H.H., Fang, L.I. and Sun, W.C. (2018), 'Forecasting performance of global economic policy uncertainty for volatility of Chinese stock market', *Physica A*, Vol. 505, No. 1, pp. 931-940.
- Zhang, Y., Zeng, Q., Ma, F. and Shi, B. (2018), 'Forecasting stock returns: Do less powerful predictors help', *Economic Modelling*, pp. 1-8 (in press).

