

國立中央大學

資訊管理研究所

碩士論文

(初稿)

特徵選取對智慧型時間序列預測之效能研究

Research on the Effectiveness of Feature Selection for  
Intelligent Time Series Prediction

研究生：陸怡廷

指導教授：李俊賢 博士

中華民國 108 年 6 月

## 特徵選取對智慧型時間序列預測之效能研究

研究生：陸怡廷

指導教授：李俊賢博士

國立中央大學資訊管理學系碩士班

### 摘要

在機器學習的過程中為正確反應出資料的全貌，使用大量的數據是不可或缺的，其代價即為運算資源的消耗與浪費，數據中的雜訊或極端值更會影響輸出結果的精確度。本研究提出一個多目標預測模型，透過計算特徵的資訊熵（Entropy）評估該特徵當中所涵蓋的資訊量，從原始數據中挑選出真正對預測目標產生有效貢獻的特徵作為機器學習的輸入值，藉此達到特徵挑選（Feature selection）的主要目的。除此之外，本研究採用球型複數類神經模糊系統（Sphere complex neuro-fuzzy system, SCNFS）搭配由鯨群演算法（Whale Optimization Algorithm, WOA）與遞迴最小平方估計法（Recursive least square estimator, RLSE）所構成的 WOA-RLSE 複合式學習演算法進行參數訓練，相較於自適應類神經模糊推論系統，SCNFS 透過球型複數模糊集合乘載更多維度的資訊量，為多目標預測的主要核心；複合式學習演算法則是藉由將參數進行拆解之後分別以 WOA 與 RLSE 訓練之，使模型能夠更有效率地接近最佳解。綜合上述架構所組建而成的預測模型得以吸收多樣化的資料集並且同時預測兩個以上的時間序列數據，最後本文將此模型進一步應用於中國股市的預測，藉由多國金融資料預測多目標股票市場證實中國與他國股市的連動性。

**關鍵字：** 特徵選取、球型複數模糊集、球型複數模糊類神經系統、鯨群演算法、遞迴最小平方演算法、多目標預測

## **Research on the Effectiveness of Feature Selection for Intelligent Time Series Prediction**

Graduate student: Yi-Ting Lu

Advisor: Dr. Chunshien Li

Department of Information Management, National Central University, Taiwan

### **Abstract**

In the process of machine learning, in order to correctly reflect the full picture of the data, the use of a large amount of data is indispensable, which cost is the consumption and waste of computing resources, the noise or extreme values in the data will affect the accuracy of the output. This study proposes a multi-target prediction model, which estimates the amount of information held by the feature by calculating the information entropy of the feature, and selects the feature that truly contributes to the predicted target from the original data as the input value for machine learning to achieve the main purpose of feature selection. In addition, this study uses the sphere complex neuro-fuzzy system(SCNFS) with a hybrid WOA-RLSE learning algorithm consist of the Whale optimization algorithm and recursive least square estimator, in the architecture of the training model. Compared with the adaptive neuro-fuzzy inference system, SCNFS multiplies the information volume of more dimensions through the sphere complex fuzzy set, which is the main core of multi-target prediction. The hybrid learning algorithm is based on the disassembled parameters, which is trained by WOA and RLSE respectively so that the model can approach the optimal solution more efficiently. The prediction model based on the above architecture can absorb diverse data sets and predict more than two time series data at the same time. Finally, this paper further applies this model to the forecasting of the Chinese stock market, to confirm the relationship between China and the international market by collecting international financial data and multi-target stock market prediction.

## 目錄

章節	頁次
中文摘要.....	i
英文摘要.....	ii
目錄.....	iii
圖目錄.....	v
符號與專有名詞說明.....	viii
第一章 緒論.....	1
1.1 研究背景.....	1
1.2 研究動機與目的.....	1
1.3 研究方法概述.....	4
1.4 論文架構.....	4
第二章 文獻探討.....	6
2.1 複數模糊集合.....	6
2.2 自適應類神經模糊推理系統.....	6
2.3 鯨群最佳化演算法.....	7
2.3.1 探索階段.....	8
2.3.2 搜尋階段.....	8
2.4 遞迴最小平方估計法.....	10
2.5 複合式學習演算法.....	11
第三章 研究方法.....	12
3.1 特徵挑選.....	12
3.1.1 資料矩陣 (Data matrix) .....	12

3.1.2	影響資訊矩陣 (Influence information matrix) .....	13
3.1.3	多目標特徵選取 (Multi-target feature selection) .....	17
3.2	球型複數模糊集合.....	18
3.3	球型複數類神經模糊系統.....	19
3.3.1	輸入層 (Input layer) .....	20
3.3.2	球型複數模糊集合層 (Sphere complex fuzzy layer) .....	20
3.3.3	前鑑部層 (Premise layer) .....	21
3.3.4	正規化層 (Normalization layer) .....	21
3.3.5	後鑑部層 (Tagaki-Sugeno layer) .....	21
3.3.6	輸出層 (Output layer) .....	21
3.4	WOA-RLSE 複合式學習演算法 .....	22
第四章	實驗內容.....	25
4.1	特徵的擷取與影響.....	25
4.2	中國金融市場預測.....	28
4.3	中國與國際市場的相互作用 .....	31
4.3.1	多目標預測與比較.....	31
4.3.2	多目標與單目標預測之差異性.....	34
第五章	結果與討論.....	37
第六章	結論.....	39
第七章	參考文獻.....	40

## 圖目錄

圖	頁次
圖 1：鯨群最佳化演算法流程圖 .....	9
圖 2：資訊熵與事件機率 $P_i$ 關係圖 .....	14
圖 3：球型複數類神經模糊系統架構 .....	20
圖 4：WOA-RLSE 複合式學習演算法流程圖 .....	23
圖 5：2018 年 SSEC 目標與特徵 <b>f30</b> 、 <b>f76</b> 條件下的目標機率密度分布（順序選取） .....	26
圖 6：2018 年 SSEC 目標與特徵 <b>f111</b> 、 <b>f39</b> 條件下的目標機率密度分布（倒序選取） .....	26
圖 7：2018 年 Nikkei 225 目標與特徵 <b>f30</b> 、 <b>f76</b> 條件下的目標機率密度分布（順序選取） .....	26
圖 8：2018 年 Nikkei 225 目標與特徵 <b>f111</b> 、 <b>f39</b> 條件下的目標機率密度分布（倒序選取） .....	26
圖 9：2018 年 NASDAQ 目標與特徵 <b>f30</b> 、 <b>f76</b> 條件下的目標機率密度分布（順序選取） .....	27
圖 10：2018 年 NASDAQ 目標與特徵 <b>f111</b> 、 <b>f39</b> 條件下的目標機率密度分布（倒序選取） .....	27
圖 11：2018 年 HSI 目標與特徵 <b>f30</b> 、 <b>f76</b> 條件下的目標機率密度分布（順序選取） .....	27
圖 12：2018 年 HSI 目標與特徵 <b>f111</b> 、 <b>f39</b> 條件下的目標機率密度分布（倒序選取） .....	27
圖 13：2000 年 SSEC 目標與特徵 <b>f6</b> 、 <b>f23</b> 條件下的目標機率密度分布 .....	29

圖 14：2001 年 SSEC 目標與特徵 <b>f25</b> 、 <b>f22</b> 條件下的目標機率密度分布 .....	29
圖 15：2002 年 SSEC 目標與特徵 <b>f4</b> 、 <b>f16</b> 條件下的目標機率密度分布 .....	29
圖 16：2003 年 SSEC 目標與特徵 <b>f11</b> 、 <b>f23</b> 條件下的目標機率密度分布 .....	29
圖 17：2004 年 SSEC 目標與特徵 <b>f26</b> 、 <b>f1</b> 條件下的目標機率密度分布 .....	29
圖 18：2005 年 SSEC 目標與特徵 <b>f23</b> 、 <b>f25</b> 條件下的目標機率密度分布 .....	29
圖 19：2006 年 SSEC 目標與 <b>f2</b> 、 <b>f7</b> 條件下的目標機率密度分布 .....	30
圖 20：2002 年 SSEC 目標與特徵 <b>f83</b> 、 <b>f112</b> 條件下的目標機率密度分布 .....	32
圖 21：2002 年 HSI 目標與特徵 <b>f83</b> 、 <b>f112</b> 條件下的目標機率密度分布 .....	32
圖 22：2002 年 Nikkei 225 目標與特徵 <b>f83</b> 、 <b>f112</b> 條件下的目標機率密度分布 .....	32
圖 23：2002 年 DJI 目標與特徵 <b>f83</b> 、 <b>f112</b> 條件下的目標機率密度分布 .....	32
圖 24：2003 年 SSEC 目標與特徵 <b>f120</b> 、 <b>f30</b> 條件下的目標機率密度分布 .....	32
圖 25：2003 年 HSI 目標與特徵 <b>f120</b> 、 <b>f30</b> 條件下的目標機率密度分布 .....	32
圖 26：2003 年 Nikkei 225 目標與特徵 <b>f120</b> 、 <b>f30</b> 條件下的目標機率密度分布 .....	33
圖 27：2003 年 DJI 目標與特徵 <b>f120</b> 、 <b>f30</b> 條件下的目標機率密度分布 .....	33
圖 28：SSEC 股市預測線 .....	35
圖 29：SSEC 預測誤差圖 .....	35
圖 30：HSI 股市預測線 .....	35
圖 31：HSI 預測誤差圖 .....	35
圖 32：Nikkei 股市預測線 .....	35
圖 33：Nikkei 預測誤差圖 .....	35
圖 34：DJI 股市預測線 .....	35
圖 35：DJI 預測誤差圖 .....	35

## 表目錄

表	頁次
表 1：股票預測文獻彙整 .....	2
表 2：單目標預測資料矩陣 .....	12
表 3：多目標預測資料矩陣 .....	13
表 4：影響資訊矩陣 .....	17
表 5：四目標預測特徵多寡之效能比較 (RMSE) .....	28
表 6：實驗 2 單目標預測實驗設定 .....	28
表 7：實驗 2 SSEC 之效能比較 (RMSE) .....	30
表 8：實驗 3-1 四目標預測實驗設定 .....	31
表 9：實驗 3-1 四目標預測之效能比較 (RMSE) .....	33
表 10：模型平均之效能比較 (RMSE) .....	33
表 11：實驗 3-2 單目標預測實驗設定 .....	34
表 12：實驗 3-2 四目標預測實驗設定 .....	34
表 13：單目標與多目標預測之效能比較 (RMSE) .....	36



## 符號與專有名詞說明

### 專有名詞

ANFIS	：自適應類神經模糊系統 (Adaptive neuro fuzzy inference system)
ANNs	：人工類神經網路 (Artificial Neural Networks)
CFS	：複數模糊集合 (Complex fuzzy set)
CNFS	：複數類神經模糊系統 (Complex neuro-fuzzy system)
IIM	：影響資訊矩陣 (Influence information matrix)
NFS	：類神經模糊系統 (Neuro-fuzzy system)
RLSE	：遞迴最小平方估計法 (Recursive least square estimator)
SCFS	：球型複數模糊集合 (Sphere complex fuzzy set)
SCNFS	：球型複數類神經模糊系統 (Sphere complex neuro-fuzzy system)
WOA	：鯨群演算法 (Whale Optimization Algorithm)

### 符號

Rule	：表示模糊規則
$x$	：語意變數 (Linguistic variables)
$A$	：模糊集合
$h$	：數值變數 (Numerical variable)
$\hat{y}$	：模型輸出
$a$	：模糊規則參數
$\vec{A}$	：隨迭代縮小的係數向量
$\vec{C}$	：根據 $\vec{r}$ 變動的係數向量
$\vec{X}$	：鯨魚的位置向量
$t$	：第 $t$ 次迭代
$\bar{D}$	：當前鯨魚與所選鯨魚之間的距離
$b$	：定義螺旋更新位置中螺旋形狀的常數，預設值為 1
$l$	：介於[-1,1]之間的亂數
$\vec{h}$	：模型的輸入向量
$f$	：已知函數
$\theta$	：欲估計的未知參數
$e$	：為誤差值
$I$	：一單位矩陣
$s$	：為使用者設定的候選特徵的個數
$f$	：候選特徵變數
$t$	：欲進行預測的目標變數
$TS$	：目標變數的集合

- $CP$  : 候選特徵池 (Candidate feature pool), 候選特徵的集合
- $I$  : 用以表示資訊量
- $H$  : 表示資訊熵
- $P$  : 離散事件發生的機率密度函數
- $p_d$  : 連續事件發生的機率密度函數
- $g$  : 增益資訊量 (Selection gain)
- $R$  : 冗餘資訊量 (Redundant information)
- $SP$  : 已選特徵池 (Selected feature pool), 已選特徵的集合
- $\Omega$  : 所有已選特徵池中特徵變數的集合
- $\Phi$  : 為集合 $\Omega$ 中的特徵變數
- $n_{OL}$  : 每個 $\Phi$ 都被重複選取的數目
- $\omega$  : 覆蓋率 (Covering rate)
- $g_{sum}$  : 總增益資訊量
- $\rho$  : 貢獻指數 (Contribution index)
- $FP$  : 最終特徵池 (Final pool)
- $r$  : 歸屬程度之振幅值
- $\theta_1$ 、 $\theta_2$  : 歸屬程度之相位值
- $\mu$  : 歸屬程度
- $\bar{\mu}$  : 球型模糊集合的歸屬程度
- $\beta$  : 啟動強度 (firing-strengths)
- $\lambda$  : 啟動強度經過正規化後的值

# 第一章 緒論

## 1.1 研究背景

資訊爆炸的年代，大量的數據被生產並加以紀錄與分析，面對海量的可用數據眾人開始期望電腦能夠利用超越人類的運算能力從歷史資料中找尋蛛絲馬跡幫助決策。然而傳統數學模型早已不適用於處理現實世界中複雜且不精確的議題[1]，反之人工類神經網路（Artificial Neural Networks, ANNs）能夠在數據當中找到固定的模式並且從中學習以適應不斷變化的環境[2]，此特性更能夠貼合現實世界的運作方式，至今仍被廣泛運用至不同的領域當中。

Khashei and Bijari [3]統整出四項 ANNs 之所以能夠作為時間序列預測模型的主要特性：首先，類神經網路是基於數據導向的自適應模型；其二為當樣本數據中含有雜訊時，類神經網路仍然能夠推斷出隱藏的資訊；第三是類神經網路已成為通用的近似函數工具，能夠近似任何連續性函數；最後則是類神經網路的非線性特性使其能夠運算現實世界中非線性的時間序列數據，達到更有效的預測效果。在大型的預測競賽當中（如 M3 Time series competition）ANNs 亦被證明為時間序列預測的有效工具[4]，故在時間序列預測領域中類神經網路已然成為目前最常被使用的預測模型之一。

## 1.2 研究動機與目的

時間序列數據的種類多不勝數，與利益具有直接相關性的即為股市預測，由於股票預測能夠顯著影響全球經濟，分析股市波動早已成為各領域間的主要議題[5]。雖然各界學者們提出各種理論試圖預測市場趨勢，市場經濟的變化仍會受到國家政策、全球和區域經濟、投資心理、人力和其他因素的影響，導致金融市場預測往往無法達到預期的結果，但在近年來的研究當中，類神經網絡已經在股票預測領域中獲得廣泛的認可[6-9]。使用類神經網路進行股市預測的主要優勢在於減少投資決策中的主觀與盲目，做出合理的投資決策[10, 11]，因此成為全世界進

行股市分析時的主要手段之一。各界學者持續不斷地將預測模型加以精進與改良，試圖尋找任何能夠提升預測精確度的方法[12]，故與之相關的研究文獻非常豐富，在此將近幾年有關於使用類神經網路研究股票預測的文獻彙整於表 1。

表 1：股票預測文獻彙整

作者 (年)	資料集來源	輸出目標	取樣期間	方法	表現指標
Li et al. [10]	SSEC, NASDAQ	Stock price	2011-2012	SVM, EMD	RMSE, MAE, MAPE
Xi, Muzhou, Lee, Li, Wei, Hai and Wu [13]	Chongqing Iron & Steel	Stock price	01.04.2012- 10.08.2012	RBF	RMSE
Bas, Yolcu, Egrioglu and Aladag [14]	BIST, TAIFEX	Stock price	10. 01.2010- 12. 23.2010	FFANN	RMSE
Ye and Wei [15]	SSEC	Stock price	2012 -2014	WNN	RMSE, MAPE
Khuat, Le, Nguyen and Le [16]	Apple, Yahoo, Google	Stock price	2009-2013 2013-2014 2014-2015.	MLP	RMSE
Qiu and Song [7]	Nikkei 225	Stock price	2007-2013	GA-ANN	Hit ratio
Chen, Cheng, Chiu and Huang [17]	TAIEX HSI	Stock price	1998-2006	ANFIS-based	RMSE, Wilcoxon test, Profitable unit
Zhang, Zhang, Zhang, Yu and Huang [18]	SSEC, TAIEX	Stock price	2000-2006 1990-1999	Type-2 FTS	RMSE, MAPE
Wei, Lou and Lei [19]	SSEC	Stock price	2009-2014	2RS-WNN	RMSE, MAD, MAPE, DS%
Chong, Han and Park [20]	KOSPI	Stock return	2010-2014	DNN	NMSE, RMSE, MAE, MI
Liu et al. [8]	000573: Shenzhen	Stock volatility	2015-2016	RNNs	Accuracy

Chatzis, Siakoulis, Petrooulos, Stavroulakis and Vlachogiannaki s [21]	39 Countries	Stock direction	1996-2017	LogR, RF, SVMs, NNs, CART, XG- Boost, MXNET	Accuracy
Pang et al. [9]	SHASHR, TMSE, TMBA, SINOPEC	Stock price	2006 -2016	ALSTM, ELSTM	MSE, DA
Lei [22]	SSEC, All Ords, CSI 300, Nikkei 225, DJI	Stock price, Stock direction	2009-2014	BP-NN, RBF-NNAN, FIS-NN, SVM, WNN, RS-WNN, 2RS-WNN	RMSE, MAD, MAPE, DS%, CP%, CD%
Shastri, Roy and Mittal [23]	Apple	Stock price	2013 - 2016	ANN	MAPE, Accuracy

除了對單一市場進行預測之外，國際市場間的高度關聯性也必需進行考量，以 1997 年的亞洲金融危機為例，其於泰國爆發之後隨即傳播到其他亞洲國家，最終波及至遙遠的美國金融市場。由此可見不同金融市場之間的相關性會對股票預測產生影響[21, 24]，能夠吸收不同金融市場數據的多目標預測模型是必要的學術進展，可惜的是目前的研究多進行單一金融市場預測，將國際股市之間相互作用納入預測因子的文獻並不多，同時進行多目標預測的學術研究更是相當稀少。除此之外，大多數文獻均以西方證券市場為主，如那斯達克綜合指數、道瓊工業平均指數等都是極其普遍的研究對象，針對中國股票市場的預測則相對稀少[9]，故本研究將特徵挑選與球型複數類神經模糊系統應用於中國證券市場，透過運用國際間多樣化的金融市場數據進行多目標預測，驗證中國與國際股票市場之間的連動性。

### 1.3 研究方法概述

本研究取得原始數據之後並不直接使用之，而是透過特徵挑選（Feature selection）的結構將實驗數據轉化為數個特徵，計算其資訊熵（Entropy）以評估該特徵所持有的資訊量，藉此挑選出對預測目標產生有效貢獻的特徵作為機器學習的輸入值，達到降低運算資源、減少極端值和雜訊的效用。

除此之外，本研究使用由球型複數模糊集合（Sphere complex fuzzy set, SCFS）構成的球型複數類神經模糊系統（Sphere complex neuro-fuzzy system, SCNFS），球型模糊集合的特性能夠乘載多維度的資訊量，為多目標預測的關鍵所在。接著採取分而治之的方式將參數分為前鑑部與後鑑部二者分別由鯨群演算法與遞迴最小平方估計法加以訓練，參數空間的分割與縮小使模型能夠更有效率的接近最佳解，故命名為 WOA-RLSE 複合式學習演算法。

### 1.4 論文架構

本文一共分為七章，以緒論為首，介紹本研究的背景、動機並且概述本研究所使用的研究方法。第二章則概述本研究中所使用的模型架構其根基與歷史脈絡，介紹模糊集合的源起、自適應類神經模糊推理系統的核心理論，本文採用的鯨群演算法（Whale Optimization Algorithm, WOA）與遞迴最小平方估計法（Recursive least square estimator, RLSE）亦在此進行說明，最後則為複合式學習演算法的原理與功效。第三章為研究方法，該章節中會詳述特徵挑選的步驟，如何透過資訊熵評估特徵的資訊量以及對預測目標的有效貢獻，篩選的條件又該如何定義。接著介紹球型複數模糊集合的運用，說明如何構建球型複數類神經模糊系統進行多目標預測，以及使用 WOA-RLSE 複合式學習演算法訓練前鑑部與後鑑部參數的具體流程。

第四章為實驗內容，本研究透過三個實驗進行時間序列預測以驗證模型的有效性。實驗一使用上海證券交易所綜合股價指數、香港恒生指數、日經平均指

數、那斯達克綜合指數進行多目標預測，藉由特徵選取的數量與順序測試特徵挑選的效用。實驗二利用上海證券交易所綜合股價指數進行單目標預測，結合其他學術文獻的預測結果，評斷本模型的預測效能。實驗三同時預測上海證券交易所綜合股價指數、香港恒生指數、日經平均指數、道瓊工業指數四個目標，並且個別將上述四個金融市場進行單目標預測，比對本模型在多目標預測與單目標預測上的差異，同時也與過去文獻進行效能比較以檢測本研究方法的優劣。第五章則針對實驗結果討論之，判斷實驗結果是否能夠達到預期。第六章為本文結論，統整本研究模型的效用與影響。第七章為參考文獻，標示本研究中所有參考資料與來源。

## 第二章 文獻探討

### 2.1 複數模糊集合

在數學的傳統集合（Crisp set）當中，元素與一集合的歸屬程度（Membership degree）只存在 0 與 1 兩者，學者 Zadeh [25]所提出的模糊集合即允許歸屬程度為介於[0,1]區間的連續型實數，此特性更能符合人類對一件事物的描述方式，為模糊類神經網路發展的基礎。Ramot, Milo, Friedman and Kandel [26]將模糊集合加以延伸，提出複數模糊集合（Complex fuzzy set, CFS）的概念，使歸屬程度進一步擴展於複數平面的單位圓盤上（Unit circle of complex plane）以複數型態表現之[27]。

Li and Chiang [28]將複數模糊集合實際應用於類神經模糊系統（Neuro-fuzzy system, NFS），提出複數類神經模糊系統（Complex neuro-fuzzy system, CNFS）的構想，並經由實驗證實 CNFS 相較於 NFS 在近似函數上的表現更為優異。本研究中所使用的球型複數類神經模糊系統（Sphere complex neuro-fuzzy system, SCNFS）即是以複數類神經模糊系統為基礎，加以擴增與延伸的結果。

### 2.2 自適應類神經模糊推理系統

將自適應類神經模糊推理系統進行拆解，模糊推理系統可以透過容許資料及的不確定性來模擬人類決策的模式，類神經網路則能夠根據所獲得的數據進行學習以適應多變的環境，自適應類神經模糊系統（Adaptive neuro fuzzy inference system, ANFIS）即是結合此兩大概念而得[29]。ANFIS最早由Jang [1]所發明，在模糊if-then規則（Fuzzy if-then rules）與輸入輸出資料對的基礎上，使輸入值與模型輸出得以相互對應。模糊if-then規則通常以 IF A THEN B 的形式呈現，通常被用來判斷不精確的推論，這些推論的模式對於人類在不確定的環境中進行決策時經常出現，同時也是模糊推論系統的核心[1]。



ANFIS 所使用的模糊推論系統則是由日本學者 Takagi and Sugeno [30]提出的 Takagi-Sugeno 模糊模型 (T-S model) 作為基礎，由多條 T-S 模糊規則建構而成，其中每條規則的輸出為輸入變數加上常數項的線性組合，表示如下：

$$\begin{aligned} R^{(k)} &= IF \ x_1 \text{ is } A_1^{(k)}(h_1) \text{ and } x_2 \text{ is } A_2^{(k)}(h_2) \text{ and } \dots \text{ and } x_M \text{ is } A_M^{(k)}(h_M) \\ THEN \ \hat{y}^{(k)} &= a_0^{(k)} + a_1^{(k)}h_1 + \dots + a_M^{(k)}h_M \end{aligned} \quad (1)$$

其中  $R^{(i)}$  為第  $i$  條規則，且  $k = 1, 2, \dots, K$ ， $K$  為總規則數； $x_j$  代表第  $j$  個輸入的語意變數 (Linguistic variables)，且  $m = 1, 2, \dots, M$ ， $M$  為總輸入個數； $A_m^{(k)}$  為第  $k$  個條件中第  $m$  個輸入的模糊集合； $h_m$  為第  $m$  個輸入的數值變數 (Numerical variable)； $\hat{y}^{(k)}$  為第  $k$  條規則的輸出； $a_m^{(k)}$  為第  $k$  條規則中第  $m$  個輸入的參數[31]。

由於自適應類神經模糊推論系統 (ANFIS) 是在自適應網路 (Adaptive network) 框架下實踐的模糊推論系統，顧名思義是由許多節點與具有方向性的點對點連結所組成的網路結構。其中部分的節點具有自適應的特性，代表該節點輸出可以透過調整其中的參數來校正，這些參數的調整值則可藉由學習規則取得，最終使模型輸出值與目標值之間的誤差達到最小化。參數自我學習的能力使自適應類神經模糊推理系統能夠應對不精確、不確定的環境，模糊規則的架構則能夠將專家的經驗和知識以模糊規則的方式呈現之，藉此補足資料的不完整性，在張克群等人 [29] 的研究當中更進一步證明了 ANFIS 在企業財務危機的預測表現比 ANNs 更加優越。

## 2.3 鯨群最佳化演算法

鯨群演算法 (Whale Optimization Algorithm, WOA) 主要受到座頭鯨的覓食與社交行為所啟發，Mirjalili and Lewis [32] 兩人將座頭鯨的補食行為開發成兩種最佳化的方法，分別為收縮環繞機制 (Shrinking encircling mechanism) 與螺旋更新位置 (Spiral updating position)。在演算法的初始狀態中，必須事先定義兩個係數

向量 $\vec{A}$  和 $\vec{C}$ ，其公式如下：

$$a = 2 - t \cdot \frac{2}{t_{max}} \quad (2)$$

$$\vec{A} = 2a \cdot \vec{r} - a \quad (3)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (4)$$

其中 $a$ 會隨著迭代由 2 線性下降至 0， $t$ 指的是第  $t$  次迭代，且 $t_{max}$ 為迭代總次數，故 $\vec{A}$ 具有隨迭代縮小範圍的特性； $\vec{r}$  是介於[0,1]之間的亂數。

鯨群演算法的最佳化策略可分為兩階段：探索階段（Exploration phase）與搜尋階段（Exploitation phase），其運作方式將於以下進行詳細的說明。

### 2.3.1 探索階段

探索階段的目的是在於搜索更廣泛的區域，避免在迭代初期就陷入局部最佳解，故此階段隨機選擇一隻鯨魚作為其他鯨魚進行移動的參考對象以進行全域搜索，在條件 $|\vec{A}| > 1$  的時候啟動此機制，其計算方式如下：

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}_i(t)| \quad (5)$$

$$\vec{X}_i(t+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \quad (6)$$

其中 $\vec{X}_{rand}$ 為隨機一隻鯨魚的位置向量， $\vec{X}_i$ 代表第  $i$  隻鯨魚，故 $\vec{D}$ 表示第  $i$  隻鯨魚與所選隨機鯨魚之間的距離。

### 2.3.2 搜尋階段

此階段會採用迭代過程表現最佳的鯨魚（表示為 $\vec{X}^*$ ）作為鯨群移動的參考位置，透過介於[0,1]的均勻分布亂數  $p$  作為使用收縮環繞機制與螺旋更新位置兩種優化方式其一的憑依，當  $p < 0.5$  時則採用收縮環繞機制，其公式如下：

$$\vec{D}' = |\vec{C} \cdot \vec{X}^* - \vec{X}_i(t)| \quad (7)$$

$$\vec{X}_i(t+1) = \vec{X}^* - \vec{A} \cdot \vec{D}' \quad (8)$$

反之採用螺旋更新位置，透過螺旋的路徑往最佳鯨魚的位置移動，公式如下：

$$\vec{D}'' = |\vec{X}^* - \vec{X}_i(t)| \quad (9)$$

$$\vec{X}_i(t+1) = \vec{X}^* + \vec{D}'' \cdot e^{bl} \cdot \cos(2\pi l) \quad (10)$$

其中  $b$  為用來定義螺旋形狀的常數，預設值為 1； $l$  為介於  $[-1,1]$  之間的亂數，替螺旋路徑的旋轉角度提供隨機性。鯨群演算法的運作流程如圖 1 所示，經由 29 個基準函數與 6 個工程設計問題的測試，WOA 的表現均優於其他演算法。

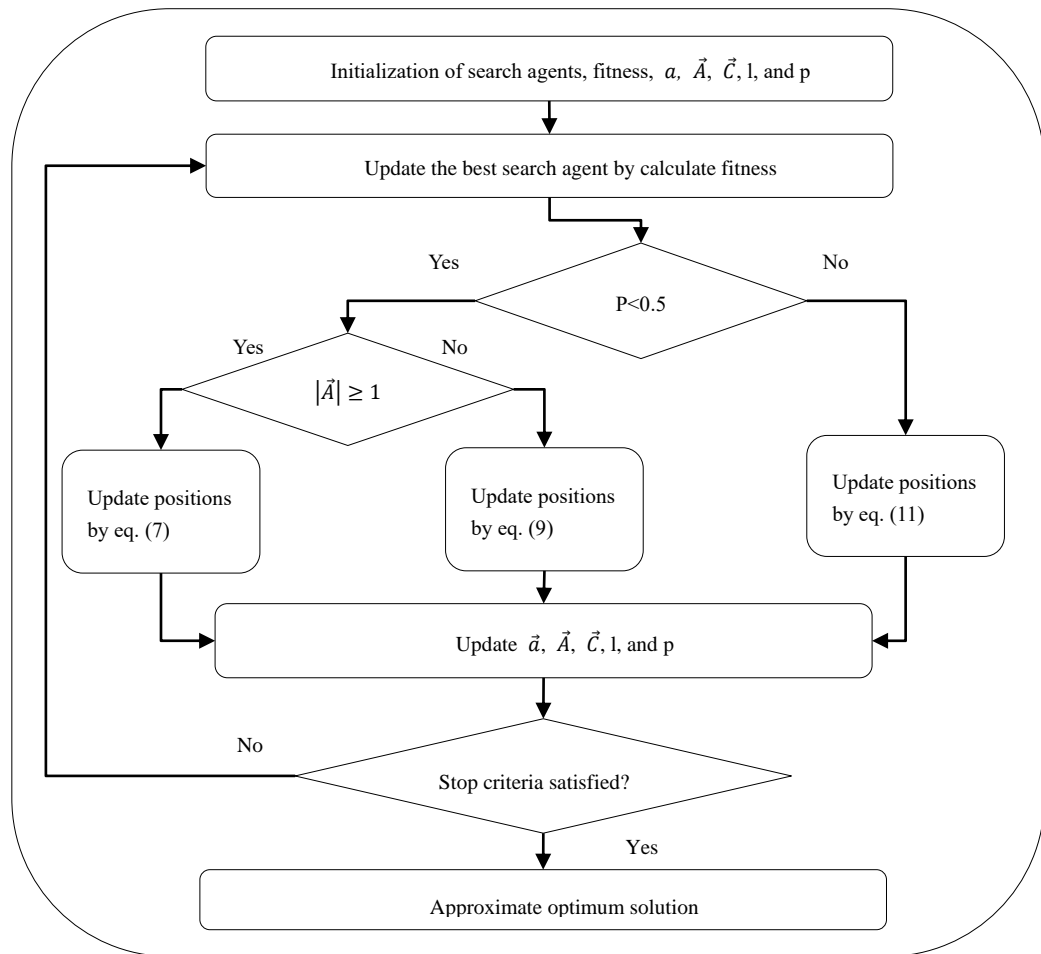


圖 1：鯨群最佳化演算法流程圖

## 2.4 遞迴最小平方估計法

遞迴最小平方估計法 (Recursive least square estimator, RLSE) 是最小平方估計法 (Least square estimator, LSE) 延伸而得的產物，根據學者 Jang et al. [2] 所述，在最小平方估計法中，線性模型的輸出  $y$  可表示如下：

$$y = \theta_1 f_1(\bar{h}) + \theta_2 f_2(\bar{h}) + \cdots \theta_n f_n(\bar{h}) + e \quad (11)$$

其中  $\bar{h}$  是模型的輸入向量； $f_i$  則是  $\bar{h}$  的已知函數； $\theta_i$  為欲估計的未知參數； $e$  為誤差值。上式 (11) 可以矩陣改寫為：

$$\mathbf{A}\boldsymbol{\theta} + \mathbf{e} = \mathbf{y} \quad (12)$$

其中矩陣  $\mathbf{A}$ 、 $\boldsymbol{\theta}$ 、 $\mathbf{y}$ 、 $\mathbf{e}$  分別為：

$$\mathbf{A} = \begin{bmatrix} f_1(\bar{h}_1) & \cdots & f_n(\bar{h}_1) \\ \vdots & \ddots & \vdots \\ f_1(\bar{h}_m) & \cdots & f_n(\bar{h}_m) \end{bmatrix}$$

$$\boldsymbol{\theta} = [\theta_1, \dots, \theta_n]^T$$

$$\mathbf{y} = [y_1, \dots, y_m]^T$$

$$\mathbf{e} = [e_1, \dots, e_m]^T$$

最小平方估計法的目的即找到一組  $\boldsymbol{\theta}$ ，得以將誤差的平方和最小化。在此將  $\boldsymbol{\theta}$  加入下標  $k$  以表示資料對的數量，利用已知  $\boldsymbol{\theta}_k$  取得  $\boldsymbol{\theta}_{k+1}$  的作法則稱為遞迴最小平方估計法，其定義如下：

$$\begin{cases} \mathbf{P}_{k+1} = \mathbf{P}_k - \frac{\mathbf{P}_k \mathbf{a}_{k+1} \mathbf{a}_{k+1}^T \mathbf{P}_k}{\mathbf{I} + \mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1}} \\ \boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \mathbf{P}_{k+1} \mathbf{a}_{k+1} (y_{k+1} - \mathbf{a}_{k+1}^T \boldsymbol{\theta}_k) \end{cases} \quad (13)$$

其中  $[\mathbf{a}_k^T : y_k]$  代表矩陣  $[\mathbf{A} : \mathbf{y}]$  的第  $k$  列， $\mathbf{I}$  為一單位矩陣。在初始設定中， $\boldsymbol{\theta}_0$  為一零向量， $\mathbf{P}_0$  則定義如下，且  $\alpha$  為一極大值。：

$$\mathbf{P}_0 = \alpha \mathbf{I} \quad (14)$$

## 2.5 複合式學習演算法

本研究中採用 RLSE 作為訓練類神經模糊系統模型中後鑑部參數的演算法，類似的作法也出現於學者 Li and Wu [33]、李俊賢與江泰緯 [34]與 Tu and Li [31] 的研究之中，在各領域內均被廣泛運用，其中又以 Li and Wu [33]所發表的 PSO-RLSE 複合式學習演算法具體描述了 PSO-RLSE 與 PSO 的效能差異。

Li and Wu [33]的研究中採用 PSO 與 RLSE 分別訓練前鑑部與後鑑部的參數，其中前鑑部的參數子集合包含模糊集合的參數，後鑑部參數子集合則是近似函數的所需參數。實驗證實 PSO-RLSE 複合式學習演算法能夠有效提升模型的學習效能，顯示複合式學習演算法將系統參數進行劃分的策略，將參數空間分割為兩個較小的搜索空間，因此能夠更輕易的找到最佳解[33, 35]。

## 第三章 研究方法

### 3.1 特徵挑選

為了盡可能的汲取資料以重現數據的完整性，本研究導入 Tu and Li [36]所提出的多目標特徵挑選（Multi-target feature selection）將不同來源的多樣化數據集合轉化為多個特徵，從中篩選出對預測目標有利的特徵進入模型中進行訓練，並藉此移除數據資料中的雜訊與極端值。多目標特徵挑選可被歸類為三大階段，依序為：資料矩陣（Data matrix）、影響資訊矩陣（Influence information matrix, IIM）與多目標特徵挑選（Multi-target feature selection），以下將逐步進行詳細的說明。

#### 3.1.1 資料矩陣（Data matrix）

特徵挑選時所使用的特徵即是由資料矩陣而得，將蒐集而得的原始數據進行差分，根據設定的候選特徵（Candidate features）數量排列差分後的數據，即可取得資料矩陣如表 2 所示：

表 2：單目標預測資料矩陣

Candidate features					Target
$f_1 \quad f_2 \quad \dots \quad f_s$					$t$
Data matrix	$x_1$	$x_2$	$\dots$	$x_s$	$x_{s+1}$
	$x_2$	$x_3$	$\dots$	$x_{s+1}$	$x_{s+2}$
	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$

其中  $s$  為使用者設定的候選特徵的個數， $\mathbf{f}$  為候選特徵變數， $\mathbf{t}$  為欲進行預測的目標變數， $x$  為原始數據進行差分後的值。多目標資料矩陣由不同目標的候選特徵變數與目標變數進行合併而得，在多目標的情況下，資料矩陣的排列請參考下表 3：

表 3：多目標預測資料矩陣

Candidate features										Targets			
										$\mathbf{t}^{(1)}$	$\mathbf{t}^{(2)}$	...	$\mathbf{t}^{( TS )}$
Data matrix	$\mathbf{f}_1$	$\mathbf{f}_2$	...	$\mathbf{f}_s$	$\mathbf{f}_{s+1}$	$\mathbf{f}_{s+2}$	...	$\mathbf{f}_{s+ TS }$		$\mathbf{t}^{(1)}$	$\mathbf{t}^{(2)}$	...	$\mathbf{t}^{( TS )}$
	$x_1^{(1)}$	$x_2^{(1)}$	...	$x_s^{(1)}$	$x_1^{(2)}$	$x_2^{(2)}$	...	$x_s^{( TS )}$		$x_{s+1}^{(1)}$	$x_{s+1}^{(2)}$	...	$x_{s+1}^{( TS )}$
	$x_2^{(1)}$	$x_3^{(1)}$	...	$x_{s+1}^{(1)}$	$x_2^{(2)}$	$x_3^{(2)}$	...	$x_{s+1}^{( TS )}$		$x_{s+2}^{(1)}$	$x_{s+2}^{(2)}$	...	$x_{s+2}^{( TS )}$
	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$		$\vdots$	$\vdots$	...	$\vdots$

其中目標變數的集合以  $TS$  表示之，且  $TS = \{\mathbf{t}^{(j)}, j = 1, 2, \dots, |TS|\}$ ， $|TS|$  為目標變數的個數。這些候選特徵的集合則被稱作候選特徵池（Candidate feature pool），以  $CP$  標記之，且  $CP = \{\mathbf{f}_i, i = 1, 2, \dots, |CP|\}$ ， $|CP|$  為所有候選特徵變數的個數，特徵挑選的目的即是從候選特徵池中的挑出一個子集合，該子集合內的所有特徵變數都可以對  $TS$  中的目標變數造成顯著的影響。

### 3.1.2 影響資訊矩陣（Influence information matrix）

在熱力學中，熵（Entropy）常被用來判斷系統混亂的程度，近代資訊之父 Shannon [37] 首度將熵的概念融入資訊理論之中，提出資訊熵的概念。如同在熱力學中熵用來測量能量差異一般，資訊理論中將資料整理前後兩狀態資訊熵的差異視為兩狀態中所增減的資訊量，在此將資訊熵以  $H$  表示之，其公式如下：

$$H = \sum_{i=1}^n P_i \log \left( \frac{1}{P_i} \right) \quad (15)$$

其中 $P_i$ 為離散事件變數（Event variable） $x_i$ 發生的機率，可根據機率密度估計法計算而得。資訊熵與 $P_i$ 之間的關係可用下圖 2 表示之，當事件發生的機率趨近於 1 則  $H$  則趨近於 0。

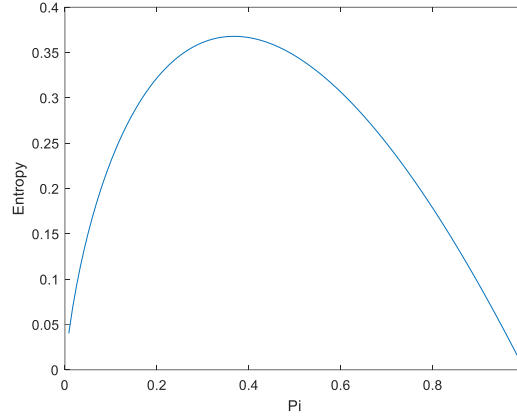


圖 2：資訊熵與事件機率 $P_i$ 關係圖

將資訊理論中的資訊量以符號  $I$  為記， $I(X,Y)$  則用來表示事件變數  $x$  的隨機變數  $X$ （Random variable）提供給事件變數  $y$  的隨機變數  $Y$  的資訊量，並且以下式表示之：

$$I(X, Y) = H(Y) - H_X(Y) \quad (16)$$

其中  $H(Y)$  為  $Y$  的熵， $H_X(Y)$  為  $Y$  受到  $X$  所提供的資訊影響後的熵，而  $H_X(Y)$  的計算公式如下：

$$H_X(Y) = - \sum_{X_i} \sum_{Y_j} P(i, j) \log P_i(j) \quad (17)$$

其中  $P(i, j)$  為離散事件中同時發生  $X_i$  和  $Y_j$  的機率密度函數， $P_i(j)$  為發生  $X_i$

的條件下， $Y_j$  發生的機率密度函數。若  $H(X, Y)$  為同時發生  $X_i$  和  $Y_j$  的亂度，定義為：

$$H(X, Y) = - \sum_{X_i} \sum_{Y_j} P(i, j) \log P(i, j) \quad (18)$$

則可根據條件機率推導如下：

$$\begin{aligned} H(X, Y) &= - \sum_{X_i} \sum_{Y_j} P(i, j) \log P(i, j) \\ &= - \sum_{X_i} \sum_{Y_j} P(i, j) \log P(i) P_i(j) \end{aligned}$$



$$\begin{aligned}
&= - \sum_{X_i} \sum_{Y_j} P(i, j) \log P(i) - \sum_{X_i} \sum_{Y_j} P(i, j) \log P_i(j) \\
&= - \sum_{X_i} P(i) \log P(i) - \sum_{X_i} \sum_{Y_j} P(i, j) \log P_i(j) \\
&= H(X) + H_X(Y)
\end{aligned}$$

得：

$$H_X(Y) = H(X, Y) - H(X) \quad (19)$$

Shannon (1948) 所提出的互資訊 (Mutual Information) 即是指稱隨機變數  $X$  與  $Y$  彼此提供的資訊量相等，即為  $I(X, Y) = I(Y, X)$ 。若為連續事件變數，則  $H(X)$  的計算公式可表示如下：

$$H(X) = \int p_d(X) \log \frac{1}{p_d(X)} dX \quad (20)$$

其中  $p_d(X)$  為連續事件中隨機變數  $X$  的機率密度函數，而公式 (17) (18) 改寫如下：

$$H_X(Y) = - \int_X \int_Y p_d(X, Y) \log \frac{p_d(X, Y)}{p_d(Y)} dXdY \quad (21)$$

$$H(X, Y) = - \int_X \int_Y p_d(X, Y) \log p_d(X, Y) dXdY \quad (22)$$

將式 (20) 近似之後可改寫為下列公式：

$$H(X) = \sum p_d(X) \log \left( \frac{\varphi}{p_d(X)} \right) \Delta X \quad (23)$$

其中  $\log \left( \frac{\varphi}{p_d(X)} \right)$  為混亂密度 (Disorder density factor)，由於  $p_d(X)$  可能為負值，故  $\varphi$  為  $p_d(X)$  的最大值加上  $\varepsilon$ ， $\varepsilon$  為一非常小的常數用以確保累加的每個數必為正，在此定義為  $10^{-10}$ 。

由於互資訊的概念忽略了正負值可能造成的影響，Tu and Li [36] 認為  $I(X, Y) = I(Y, X)$  應屬於一種極端的現象，故將正負號進行考量之後，另稱之為影響資訊 (Influence Information)，將第  $i$  個特徵對第  $j$  個特徵提供的資訊量以下式計算之：

$$I(\mathbf{f}_i \rightarrow \mathbf{f}_j) = I(\mathbf{f}_i^-, \mathbf{f}_j) \int_{-\infty}^0 p_d(\mathbf{f}_i) d\mathbf{f}_i + I(\mathbf{f}_i^+, \mathbf{f}_j) \int_0^{\infty} p_d(\mathbf{f}_i) d\mathbf{f}_i \quad (24)$$

其中 $I(\mathbf{f}_i \rightarrow \mathbf{f}_j)$ 表示為第 $i$ 個特徵對第 $j$ 個特徵的影響資訊， $I(\mathbf{f}_i^-, \mathbf{f}_j)$ 、 $I(\mathbf{f}_i^+, \mathbf{f}_j)$ 分別為 $\mathbf{f}_i$ 的負值與正值提供給 $\mathbf{f}_j$ 的資訊量，與事件 $\mathbf{f}_i$ 的發生機率相乘後再進行加總，故 $I(\mathbf{f}_i \rightarrow \mathbf{f}_j)$ 可視為互資訊的期望值，當 $I(X, Y) = I(Y, X)$ 時即回歸於 Shannon 的互資訊理論。納入正負號的差異使影響資訊的考量範圍廣於 Shannon 的互資訊，資訊量帶有方向性之後即可透過特徵對目標的影響資訊 $I(\mathbf{f}_i \rightarrow \mathbf{t}^{(l)})$ 將對目標具有影響力的特徵篩選而出，此特性即為影響資訊的核心價值所在。

然而在特徵數據中所取得的資訊有可能是不完整的，其實際機率密度未知，故透過數據進行機率密度函數估計，將數據的分布視為該事件的完整特徵[38]，在影響資訊當中即使用機率密度函數的積分將特徵完整化。若第 $i$ 個特徵對目標的影響資訊為負，即代表在 $\mathbf{f}_i$ 條件下的亂度增加且對目標產生負面影響，可事先去除之，在此也有過濾雜訊與極端值的作用。接著根據影響資訊建構影響資訊矩陣（Influence information matrix, IIM）如表 4 所示，將多樣化的數據來源對每個預測目標的影響資訊清楚呈現之，在下一階段中進行特徵的選取。

表 4：影響資訊矩陣

	$\mathbf{f}_1$	$\mathbf{f}_2$	...	$\mathbf{f}_{ CP }$	$\mathbf{t}^{(j)}$
$\mathbf{f}_1$	$0$	$I_{\mathbf{f}_1 \rightarrow \mathbf{f}_2}$	...	$I_{\mathbf{f}_1 \rightarrow \mathbf{f}_{ CP }}$	$I_{\mathbf{f}_1 \rightarrow \mathbf{t}^{(j)}}$
$\mathbf{f}_2$	$I_{\mathbf{f}_2 \rightarrow \mathbf{f}_1}$	$0$	...	$I_{\mathbf{f}_2 \rightarrow \mathbf{f}_{ CP }}$	$I_{\mathbf{f}_2 \rightarrow \mathbf{t}^{(j)}}$
$\vdots$	$\vdots$	$\vdots$	$0$	$\vdots$	$\vdots$
$\mathbf{f}_{ CP }$	$I_{\mathbf{f}_{ CP } \rightarrow \mathbf{f}_1}$	$I_{\mathbf{f}_{ CP } \rightarrow \mathbf{f}_2}$	...	$0$	$I_{\mathbf{f}_{ CP } \rightarrow \mathbf{t}^{(j)}}$
$\mathbf{t}^{(j)}$	$I_{\mathbf{t}^{(j)} \rightarrow \mathbf{f}_1}$	$I_{\mathbf{t}^{(j)} \rightarrow \mathbf{f}_2}$	...	$I_{\mathbf{t}^{(j)} \rightarrow \mathbf{f}_{ CP }}$	$0$

其中 $\mathbf{t}^{(j)}$ 為第  $j$  個目標變數，且  $j=1,2,\dots,|TS|$ 。

### 3.1.3 多目標特徵選取 (Multi-target feature selection)

計算出影響資訊矩陣階段之後即可取得 $|TS|$ 個 IIM，記作 $IIM^{(j)}$ 且  $j = 1, 2, \dots, |TS|$ 。接著必須計算第  $i$  個候選特徵對第  $j$  個目標的增益資訊量 (Selection gain)，以符號  $g$  表示，計算方式如下：

$$g(\mathbf{f}_i \rightarrow \mathbf{t}^{(j)}) = I_{\mathbf{f}_i \rightarrow \mathbf{t}^{(j)}} - R_{\mathbf{f}_i \rightarrow SP^{(j)}} \quad (25)$$

其中 $I_{\mathbf{f}_i \rightarrow \mathbf{t}_j}$ 為第  $i$  個特徵對第  $j$  個目標的影響資訊， $R_{\mathbf{f}_i \rightarrow SP^{(j)}}$ 為第  $i$  個特徵對 $SP^{(j)}$ 的冗餘資訊量 (Redundant information)，其目的在於去除不同的已選特徵池中 (Selected feature pool, SP) 重複汲取到的資訊量，公式如下：

$$R_{\mathbf{f}_i \rightarrow SP^{(j)}} = \frac{1}{2 \times |SP^{(j)}|} \left\{ \sum_{q=1}^{|SP^{(j)}|} I_{\mathbf{f}_i \rightarrow SP_q^{(j)}} + I_{SP_q^{(j)} \rightarrow \mathbf{f}_i} \right\} \quad (26)$$

若 $g(\mathbf{f}_i \rightarrow \mathbf{t}_j)$ 為正，將該候選特徵放入已選特徵池，以 $SP^{(j)}$ 表示第  $j$  個已選特徵池，此步驟結束後共能取得 $|TS|$ 個已選特徵池。

將所有已選特徵池中的特徵變數取出形成集合 $\Omega = \{\Phi_k, k = 1, 2, \dots, |\Omega|\}$ ，其中 $\Phi_k$ 為集合 $\Omega$ 中的第 $k$ 個特徵變數，且 $k=1, 2, \dots, |\Omega|$ 。由於已選特徵池中可能會選擇到重複的特徵變數，故對於每個 $\Phi_k$ 都必須計算它被重複選取的數目，標記為 $n_{OL}(\Phi_k)$ ，並且計算其覆蓋率（Covering rate）記作 $\omega$ ，則每個 $\Phi_k$ 對於所有目標變數的覆蓋率 $\omega$ 計算公式如下：

$$\omega(\Phi_k) = \frac{n_{OL}(\Phi_k)}{|TS|} \quad (27)$$

接著將每個 $\Phi_k$ 對目標變數的增益資訊量進行累加，得到 $\Phi_k$ 的總增益資訊量 $g_{sum}(\Phi_k)$ ：

$$g_{sum}(\Phi_k) = \sum_{j=1}^{|TS|} g(\Phi_k \rightarrow t^{(j)}) \quad (28)$$

透過 $\Phi_k$ 的覆蓋率 $\omega(\Phi_k)$ 與總增益資訊量 $g_{sum}(\Phi_k)$ 則可以計算出特徵變數 $\Phi_k$ 的貢獻指數（Contribution index）記作 $\rho$ ：

$$\rho(\Phi_k) = \omega(\Phi_k) \cdot g_{sum}(\Phi_k) \quad (29)$$

$\rho(\Phi_k)$ 可用來解釋 $\Phi_k$ 對於所有目標變數的有效貢獻，若 $\rho(\Phi_k)$ 的值大於所設定的門檻 $\rho_{th}$ ，則可說 $\Phi_k$ 對於目標變數具有顯著的影響力。預設 $\rho_{th} = \bar{\omega} \cdot \bar{g}_{sum}$ ，其中 $\bar{\omega}$ 為所有特徵變數覆蓋率的平均值， $\bar{g}_{sum}$ 為所有特徵變數總增益資訊量的平均值。若 $\rho(\Phi_k) > \rho_{th}$ 則以 $n_{tmp}$ 紀錄符合門檻值的特徵個數，且 $1 \leq n_{tmp} \leq |\Omega|$ ，同時設定最終特徵池（Final pool, FP）中特徵變數個數的上限 $n_U$ 與下限 $n_L$ ，在介於 $n_U$ 、 $n_L$ 之間取最大值作為最終特徵池的特徵變數個數 $n_{FP}$ 。定義最終特徵池的特徵個數後，將 $\rho(\Phi_k)$ 進行排序取前 $n_{FP}$ 個特徵變數記入最終特徵池，這些最終特徵池內的特徵進行還原即可作為機器學習的輸入使用。

### 3.2 球型複數模糊集合

根據學者 Li and Chiang [28]所提出的複數類神經模糊系統可得知，在類神經模糊系統中使用複數模糊集合相較於傳統模糊集合具有更好的適應力，然而現實中的資料通常都是複雜且多樣的，故模型應該要盡可能的容納多個輸入才能呈現

資料的全貌。本研究中所使用的球型複數模糊集合（Sphere complex fuzzy set, SCFS）的概念即是以複數模糊集合（Complex fuzzy set）為基礎加以延伸，將歸屬程度從複數平面上的單位圓盤再度加以擴展至複數空間中的單位球體（Unit sphere of complex plane）上，使該型態的歸屬程度能夠比複數模糊集合承載更多維度的資訊量。

在此將  $r$  定義為歸屬程度之振幅值； $\theta_1$  和  $\theta_2$  為歸屬程度之相位值。振幅值使用高斯函數的歸屬程度計算而得，再對振幅值進行一階、二階偏微分取得相位值  $\theta_1$  和  $\theta_2$ ，接著將振幅值分別與兩個相位值組合成不同的歸屬程度表示如下：

$$\begin{cases} u_1 = r \cdot \cos\theta_1 \cdot \cos\theta_2 \\ u_2 = r \cdot \sin\theta_1 \cdot \cos\theta_2 \\ u_3 = r \cdot \sin\theta_2 \end{cases} \quad (30)$$

將上述的歸屬程度與複數的概念結合後進行向量化（Vectorization），產生出球型複數模糊集合之歸屬程度：

$$\vec{\mu}_{i,j} = \begin{bmatrix} \mu_{i,j}^{(1)} \\ \mu_{i,j}^{(2)} \end{bmatrix} = \begin{bmatrix} u_1 + ju_2 \\ u_1 + ju_3 \end{bmatrix} \quad (31)$$

其中  $j = \sqrt{-1}$ ，然而上式（30）僅僅是多種組合方式的其中一種，若增加其排列的多樣性之後與式（31）結合，則可以產生多個複數歸屬程度，故球型複數類神經模糊系統可吸收更多資訊量，並且進行多目標的輸出。

### 3.3 球型複數類神經模糊系統

本研究採用自適應網路的架構，總共有六個階層，依序為：輸入層（Input layer）、球型複數模糊集合層（Sphere complex fuzzy set layer）、前鑑部層（Premise layer）、正規化層（Normalization layer）、後鑑部層（Takagi-Sugeno layer）與輸出層（Output layer），系統架構如圖 3 所示，各層的功能將於以下進行詳細的說明。

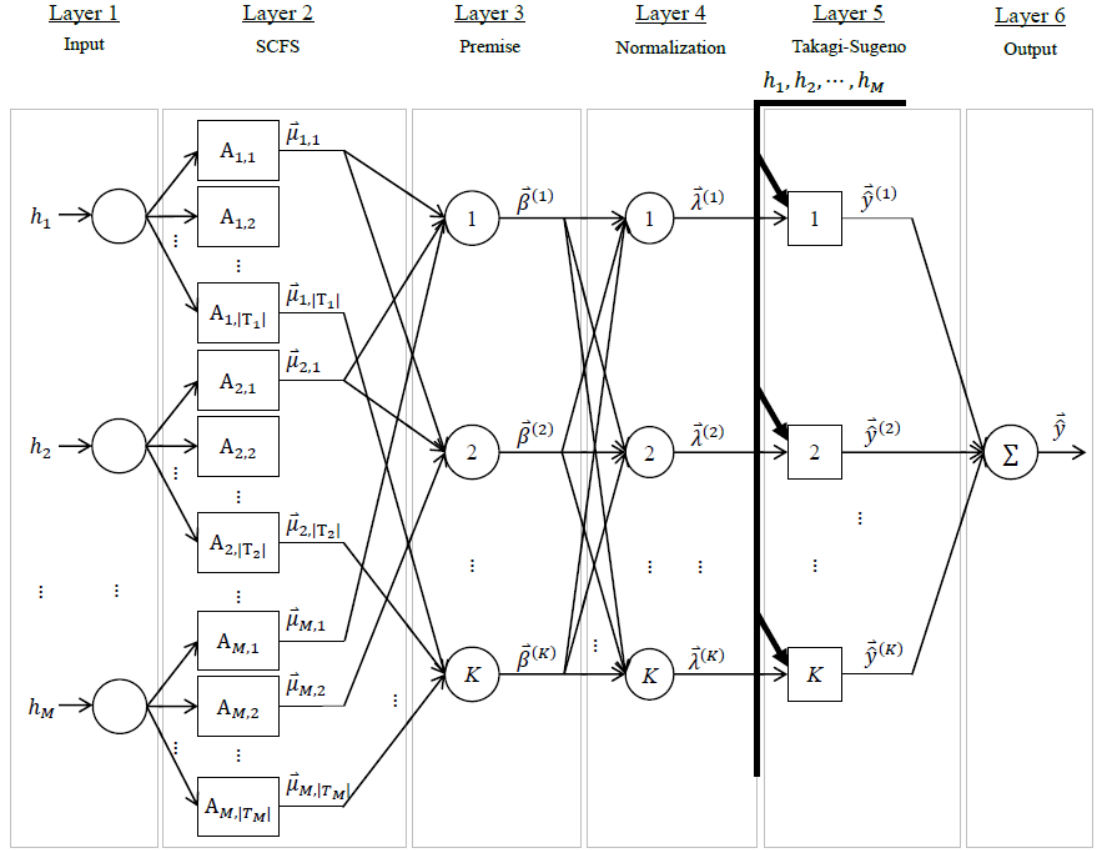


圖 3：球型複數類神經模糊系統架構

### 3.3.1 輸入層（Input layer）

輸入層的作用即是接收輸入值並將其傳入下一階層，輸入值表示如下：

$$\vec{h} = [h_1, h_2, \dots, h_M]^T \quad (32)$$

其中  $M$  為輸入值的個數。

### 3.3.2 球型複數模糊集合層（Sphere complex fuzzy layer）

此階層中的所有節點皆為一球型複數模糊集合，輸入值進入此階層後透過高斯函數與偏微分個別計算振幅值和相位值，接著利用式（30）（31）輸出該球型模糊集合的歸屬程度。

### 3.3.3 前鑑部層 (Premise layer)

前鑑部層中的一個節點即為一條規則中所有子規則的組合，這些組合可以建構矩陣 (Construction matrix, CM) 的方式呈現之，目的是計算出輸入值對每條規則的啟動強度 (firing-strengths)，以符號  $\beta$  表示，由每個組合中所對應的歸屬程度取交集而得，其計算公式如下：

$$\vec{\beta}^{(k)} = \prod_{m=1}^M \vec{\mu}_{m,CM_{k,m}} \quad (33)$$

其中  $\vec{\mu}_{m,CM_{k,m}}$  為第  $m$  個模糊集合對第  $k$  條規則之歸屬程度。

### 3.3.4 正規化層 (Normalization layer)

此階層取得前一層計算出的啟動強度後，將其與所有啟動強度的總和進行正規化，正規化後的值以  $\lambda$  表示，且  $\lambda$  值的總合為 1，計算公式如下：

$$\vec{\lambda}^{(k)} = \frac{\vec{\beta}^{(k)}}{\sum_{k=1}^K \vec{\beta}^{(k)}} \quad (34)$$

### 3.3.5 後鑑部層 (Tagaki-Sugeno layer)

此階層以 T-S 模糊規則中的後鑑部為基礎，與  $\lambda$  值相乘後得到此規則的輸出，其計算公式如下：

$$\vec{y}^{(k)} = \vec{\lambda}^{(k)}(a_0^{(k)} + a_1^{(k)}h_1 + \cdots + a_M^{(k)}h_M) \quad (35)$$

### 3.3.6 輸出層 (Output layer)

將每條規則的輸出進行加總後得到模糊推論系統的輸出，公式如下：

$$\vec{\hat{y}} = \sum_{k=1}^K \vec{y}^{(k)} \quad (36)$$

### 3.4 WOA-RLSE 複合式學習演算法

在本文第二章中所提到的PSO-RLSE複合式學習演算法採用PSO與RLSE分別進行前鑑部與後鑑部參數的訓練，WOA-RLSE複合式學習演算法則是將PSO以鯨群演算法取代之。前鑑部的參數子集合包含模糊集合的參數，由WOA負責訓練前鑑部參數建構出模糊集合之後，接著計算啟動強度並且進行正規化作為RLSE的輸入使用；後鑑部參數子集合則是近似函數的所需參數，由RLSE訓練之。根據迭代不斷的學習與調整最終得到能夠使輸出與目標的誤差最小化的參數，使測試資料得以根據學習而得的參數進行目標預測，參數空間的分割則能夠更有效率地找到最佳解，即為WOA-RLSE複合式學習演算法的主要目的，其演算法流程圖如圖4所示。



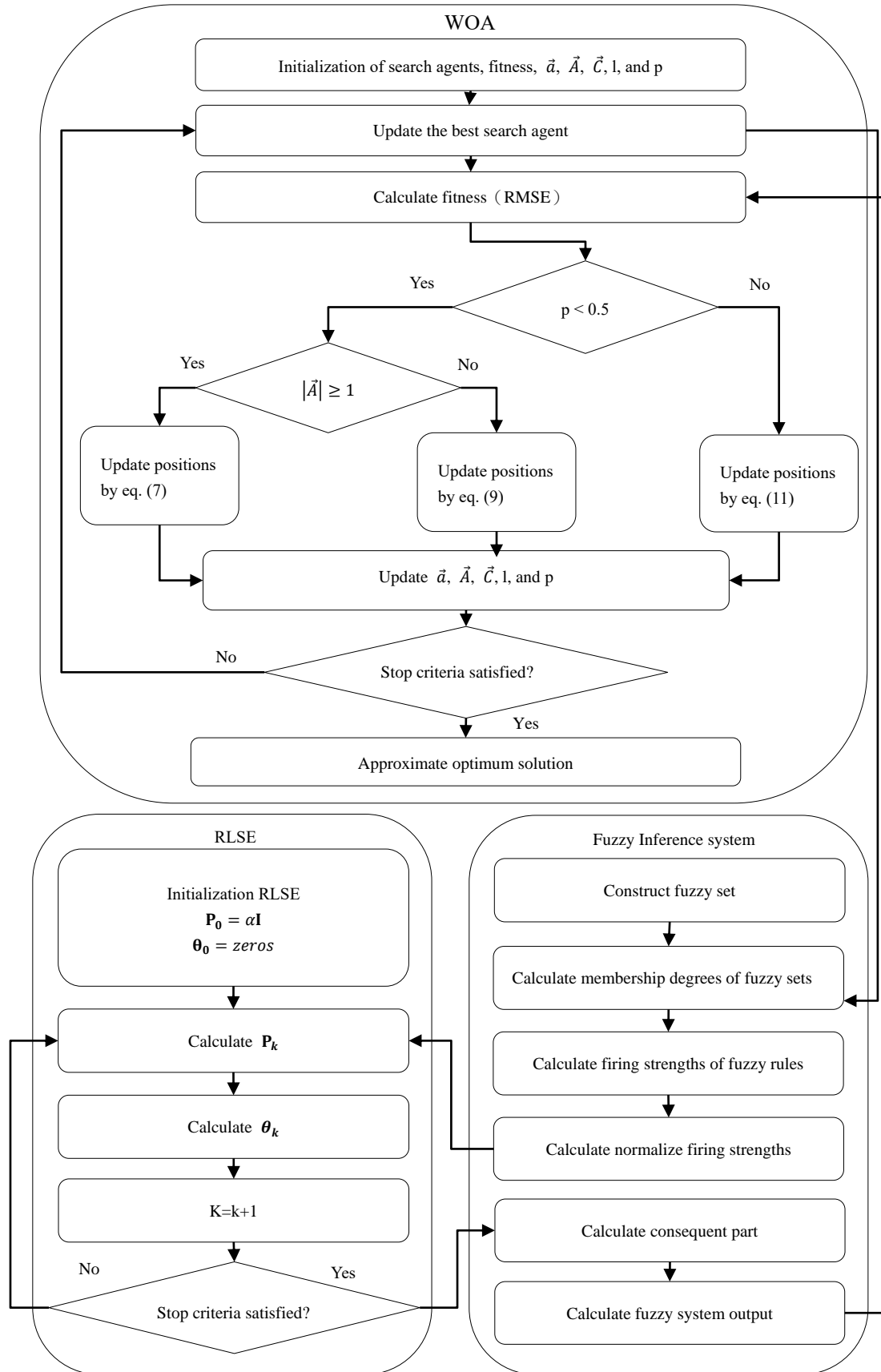


圖 4：WOA-RLSE 複合式學習演算法流程圖

由於本文中使用球型複數模糊集合計算而得的歸屬程度為複數形態，模型的輸出 $y$ 亦為複數，故在多目標預測當中的目標也必須為複數型才能夠計算誤差，並透過誤差值評估參數的學習效能，因此複數型的目標值則以下列方式建構而成：

$$\begin{cases} \mathbf{target}^{(1)} = \mathbf{t}^{(1)} + \mathbf{t}^{(2)} * j \\ \mathbf{target}^{(2)} = \mathbf{t}^{(3)} + \mathbf{t}^{(4)} * j \\ \vdots \\ \mathbf{target}^{(|TS|/2)} = \mathbf{t}^{(|TS|-1)} + \mathbf{t}^{(|TS|)} * j \end{cases} \quad (37)$$

其中 $\mathbf{target}$ 表示為模型的目標； $\mathbf{t}$ 為目標變數； $|TS|$ 為目標變數的個數； $j = \sqrt{-1}$ 。

此外，本研究使用均方根誤差（Root-mean-square error, RMSE）以評估目標與實際輸出之間的誤差值，並將其作為WOA-RLSE複合式演算法的效能指標，其計算方式如下：

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^m e_i^2}{m}} \quad (38)$$

## 第四章 實驗內容

由於不同股票市場之間的收盤價可能具有很大的差異，為確保較大的數據不會對相對小的數據造成影響並且已知正規化後的數據可以維持模型的可靠性與精準度[15, 21]，故本研究中的所有實驗數據會先透過正規化，轉化為介於[0,1]之間的值。正規化公式則採用下式：

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (39)$$

其中 $x$ 為原始資料， $x^*$ 為正規化後的資料， $x_{max}$ 、 $x_{min}$ 為原始資料的最大與最小值[9, 11, 15]。

本研究中導入了特徵挑選與球型複數類神經模糊系統兩大架構，以一個月份的資料預測後一天股價，執行實驗十次、每次迭代 1000 回的方式，設計出以下三個實驗個別探討本文之模型架構的優劣。除此之外，模型當中規則數的數量與參數量有直接的關係，為了控制參數與數據之間的比例不會導致過度凝合

（Overfitting）的現象進而影響到模型的精確度，故需要篩選建構矩陣以減少規則數，此作法稱之為區塊挑選（Block selection）。其作法是將一區塊內所有輸入的啟動強度進行累加，若累加值超過所給定的門檻值，則判斷該區塊對於模型有足夠的影響力，並選取該區塊的模糊規則進行訓練。本研究將區塊挑選的門檻值定義為所有區塊啟動強度和的平均值，並且從中挑選啟動強度和最高的前 2-4 條規則作為模型的訓練參數。

### 4.1 特徵的擷取與影響

實驗一當中使用上海證券交易所綜合股價指數（SSEC）、香港恒生指數（HSI）、日經平均指數（Nikkei 225）、那斯達克綜合指數（NASDAQ）自 2018 年 1 月 1 日至 2018 年 12 月 31 日的股票收盤價，取交集共 217 筆交易資料作為模型輸入進行四目標預測，候選特徵數為 30 得 187 筆資料對，其中取 83% 作為訓練資料集，共 156 筆，其餘為測試資料。

本實驗將候選特徵依其貢獻指數排列之後，分別透過順序與倒序的方式取前 2、4、6、8 個特徵，亦即貢獻指數表現最佳與最劣的 2、4、6、8 個特徵進行預測，同取規則數 3，用以檢驗貢獻指數的有效性以及特徵數量對預測結果的影響。在此將貢獻指數表現最佳與最劣的兩個候選特徵與目標變數的機率密度函數進行作圖，可參考圖 5-12。

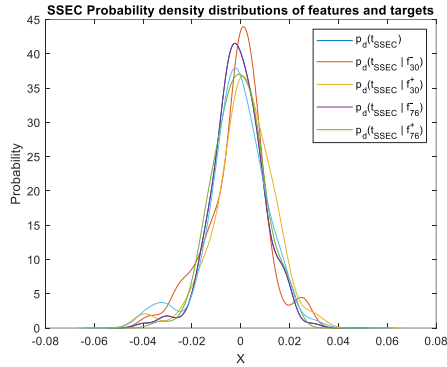


圖 5：2018 年 SSEC 目標與特徵  $f_{30}$ 、 $f_{76}$  條件下的目標機率密度分布（順序選取）

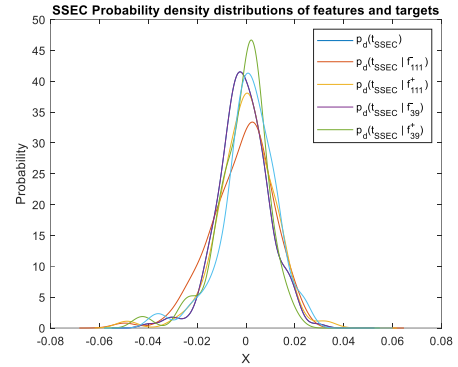


圖 6：2018 年 SSEC 目標與特徵  $f_{111}$ 、 $f_{39}$  條件下的目標機率密度分布（倒序選取）

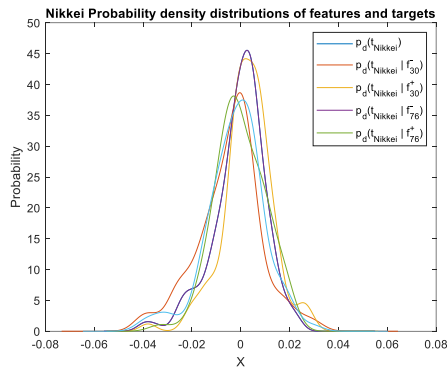


圖 7：2018 年 Nikkei 225 目標與特徵  $f_{30}$ 、 $f_{76}$  條件下的目標機率密度分布（順序選取）

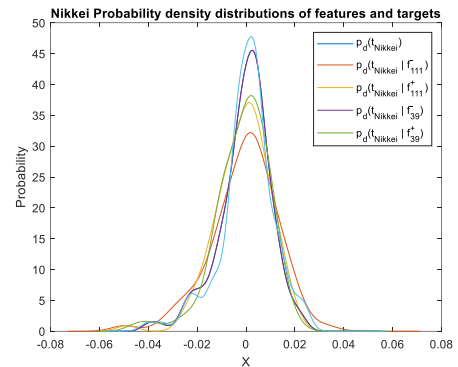


圖 8：2018 年 Nikkei 225 目標與特徵  $f_{111}$ 、 $f_{39}$  條件下的目標機率密度分布（倒序選取）

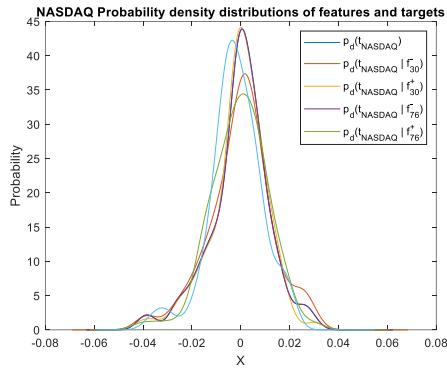


圖 9：2018 年 NASDAQ 目標與特徵  $f_{30}$ 、 $f_{76}$  條件下的目標機率密度分布（順序選取）

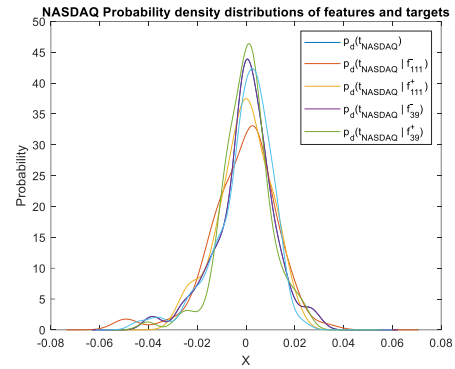


圖 10：2018 年 NASDAQ 目標與特徵  $f_{111}$ 、 $f_{39}$  條件下的目標機率密度分布（倒序選取）

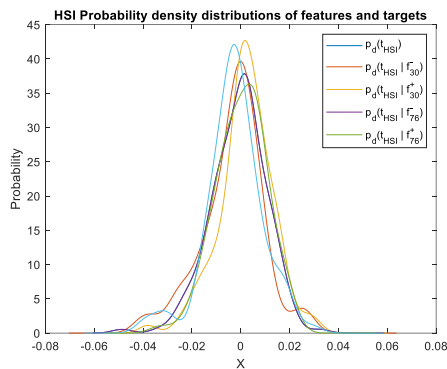


圖 11：2018 年 HSI 目標與特徵  $f_{30}$ 、 $f_{76}$  條件下的目標機率密度分布（順序選取）

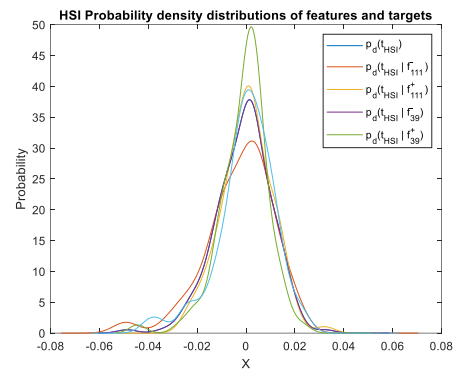


圖 12：2018 年 HSI 目標與特徵  $f_{111}$ 、 $f_{39}$  條件下的目標機率密度分布（倒序選取）

圖 5-12 當中標示目標變數本身與目標變數受到貢獻指數最高的  $f_{30}$ 、 $f_{76}$  其正負值影響後的機率密度函數（圖 5、7、9、10），以及目標變數本身與目標變數受到貢獻指數最低的  $f_{111}$ 、 $f_{39}$  其正負值影響後的機率密度函數（圖 6、8、10、12），透過公式（16）的計算可藉由機率密度之間的差評估特徵對目標的資訊量，成為特徵挑選的主要依據。由圖中可觀察到目標變數受到表現最佳的特徵影響後，其機率密度函數的圖形具有抽高的現象，反之則下降。

如前文所述特徵挑選的方式又分為順序挑選與倒序挑選兩種，順序挑選（descending order）即選擇貢獻指數表現最佳的 2、4、6、8 個特徵，反之倒序挑選（ascending order）則選擇貢獻指數表現最劣的 2、4、6、8 個特徵。實驗結果列於表 5，RMSE 最低者以粗體示之。

表 5：四目標預測特徵多寡之效能比較（RMSE）

		RMSE				
		SSEC	HSI	Nikkei 225	NASDAQ	Average
2 Features	Descending	<b>35.1970</b>	<b>350.6400</b>	<b>222.6781</b>	<b>94.1295</b>	<b>175.6612</b>
	Ascending	35.7287	366.9145	253.3162	94.5546	187.6285
4 Features	Descending	<b>34.8774</b>	<b>328.2104</b>	<b>232.2399</b>	96.4573	<b>172.9463</b>
	Ascending	37.7493	386.4353	242.8330	<b>88.6865</b>	188.926
6 Features	Descending	<b>34.3217</b>	<b>325.1491</b>	<b>222.2573</b>	<b>97.0546</b>	<b>169.6957</b>
	Ascending	35.0286	372.4195	249.4580	98.9589	188.9663
8 Features	Descending	<b>33.3978</b>	<b>326.2835</b>	<b>210.2565</b>	92.2382	<b>165.544</b>
	Ascending	35.8468	380.9012	252.2235	<b>88.7380</b>	189.4274

由上表 5 當中可得知，特徵數較少的 RMSE 均高於特徵數較高者，亦即 RMSE 隨著特徵數目的增加而減少，故數量的多寡確實會影響模型的輸出結果；此外，順序挑選的 RMSE 其平均值優於倒序挑選，即驗證貢獻指數的有效性。

## 4.2 中國金融市場預測

本實驗以 Zhang et al. [18] 的研究為基礎，取 2000 至 2006 年上海證券交易所綜合股價指數（SSEC）的年度股價資訊做為資料來源，由於該文獻中使用 1 至 10 月的股市收盤價作為訓練資料集相當於年度資料的 83%，故本實驗遵循該設定取資料對的 83% 作為訓練資料集，其餘為測試資料集。由於 2005 年 SSEC 符合增益資訊量的特徵數為 4，故挑選後的特徵數取 4，其餘年份挑選後特徵數為 10，規則數 3 的條件下，實驗設定如表 6 所示：

表 6：實驗 2 單目標預測實驗設定

	2000	2001	2002	2003	2004	2005	2006
Number of original data	259	245	260	260	261	259	240
Number of data pairs	229	215	230	230	231	229	210
Number of training data	191	179	191	191	192	191	175
Number of rules	3	3	3	3	3	3	3
Number of premise parameters	24	10	36	40	38	6	24
Number of consequent parameters	33	9	33	33	33	15	33

將貢獻指數表現最佳的兩個候選特徵與目標變數的機率密度函數作圖可得結果如下 (圖 13-19)：

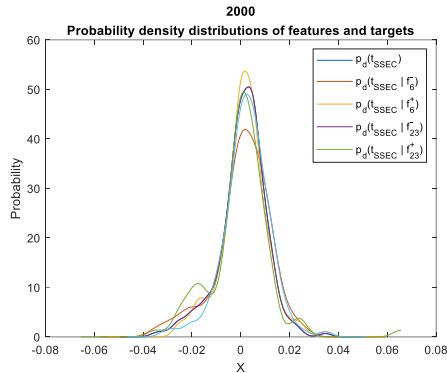


圖 13：2000 年 SSEC 目標與特徵  $f_6$ 、 $f_{23}$  條件下的目標機率密度分布

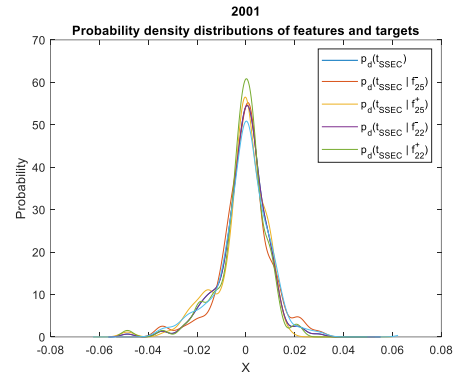


圖 14：2001 年 SSEC 目標與特徵  $f_{25}$ 、 $f_{22}$  條件下的目標機率密度分布

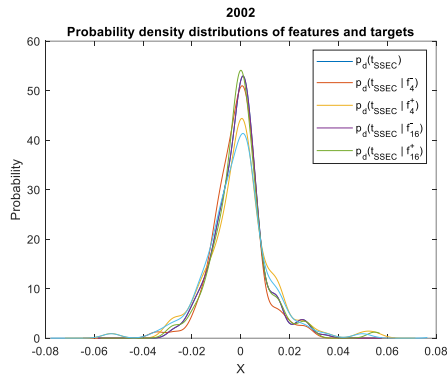


圖 15：2002 年 SSEC 目標與特徵  $f_4$ 、 $f_{16}$  條件下的目標機率密度分布

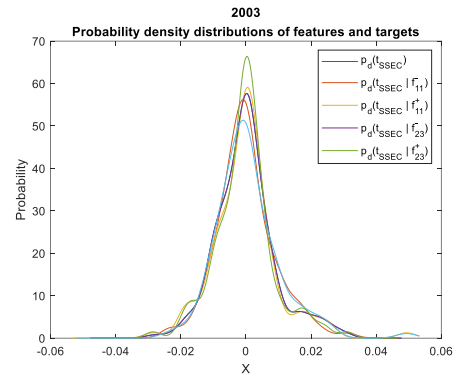


圖 16：2003 年 SSEC 目標與特徵  $f_{11}$ 、 $f_{23}$  條件下的目標機率密度分布

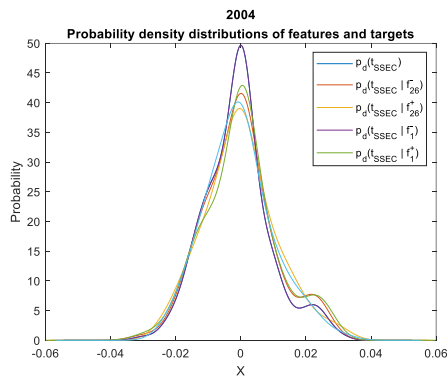


圖 17：2004 年 SSEC 目標與特徵  $f_{26}$ 、 $f_1$  條件下的目標機率密度分布

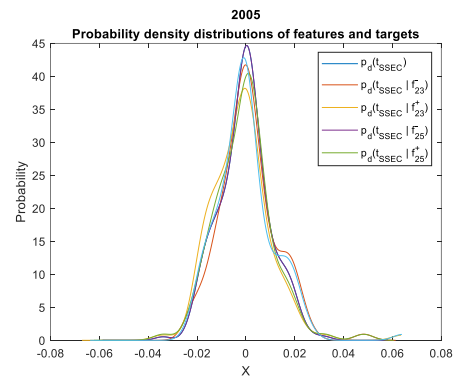
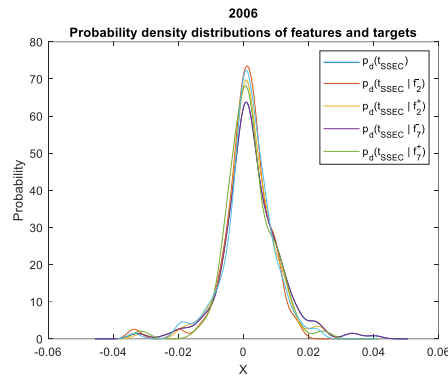


圖 18：2005 年 SSEC 目標與特徵  $f_{23}$ 、 $f_{25}$  條件下的目標機率密度分布

圖 19：2006 年 SSEC 目標與 $f_2$ 、 $f_7$ 條件下的目標機率密度分布

實驗效能（RMSE）比較作表於下表 7，表內數據均由 Zhang et al. [18]提供。

表 7：實驗 2 SSEC 之效能比較（RMSE）

	RMSE							Average
	2000	2001	2002	2003	2004	2005	2006	
Huang and Yu (2005)	23.9147	31.9274	31.9575	21.9938	21.7138	14.6053	75.0643	31.5967
Cheng et al. (2008)	29.4617	33.5855	33.4515	21.6367	32.0092	12.3227	64.0943	32.3659
Chen (2002)	40.768	43.009	57.6315	32.2600	28.4259	16.4664	62.6612	40.1746
Lee et al. (2006)	30.5366	48.4292	45.2494	24.1420	22.3151	12.0581	82.0055	37.8194
Egrioglu et al. (2011)	17.9911	24.0736	26.3361	18.1261	<b>12.5963</b>	<b>5.9938</b>	114.9601	31.4396
Wang et al. (2013)	43.0975	34.0014	26.4196	17.8860	20.1084	11.8674	379.5415	75.9888
Bas et al. (2015)	35.1766	55.1909	55.0887	66.6560	37.5188	27.9020	221.1243	71.3955
Yolcu et al. (2016)	34.0485	51.7665	56.8118	65.4207	33.7176	24.0424	226.9612	70.3955
Zhang et al. [18]	<b>16.2662</b>	<b>20.3227</b>	18.0470	17.7821	13.7292	9.0226	36.5687	18.8198
Proposed method	17.3292	22.1362	<b>15.4426</b>	<b>12.467</b>	15.4890	13.8235	<b>22.7848</b>	<b>16.8633</b>

根據 2000 至 2006 期間的平均 RMSE 可知，本研究中所提出的模型總體效能優於 Zhang et al. [18]所提出的 Type-2 FTS。



### 4.3 中國與國際市場的相互作用

實驗三中分為兩部分，首先本研究將與當前的學術文獻進行預測效能之比較，用以檢測本模型架構的有效性，接著透過使用本研究所提出的模型架構分別預測單目標與多目標股票收盤價，並且觀察兩者之間的效能差異，藉此驗證不同市場之間的連動關係。

#### 4.3.1 多目標預測與比較

在第一部分當中，本研究取 Hsieh et al. [6]、Chen et al. [17]以及 Zhang et al. [18]的研究當中含有 SSEC、HSI、Nikkei 225、DJI 的研究數據並將三篇文件中的股市資料取交集，以 2002 至 2003 年 SSEC、HSI、Nikkei 225、DJI 的共同交易日為資料集來源，取資料對當中的 83%作為訓練資料集，設定挑選後的特徵數為 10 且規則數為 3，實驗設定如表 8 所示：

表 8：實驗 3-1 四目標預測實驗設定

	2002	2003
Number of original data	227	226
Number of data pairs	197	196
Number of training data	158	163
Number of rules	3	3
Number of premise parameters	36	32
Number of consequent parameters	33	33

取貢獻指數最佳的前兩個候選特徵，其目標本身與在特徵影響下該目標變數的機率密度函數作圖如下：(圖 20-27)

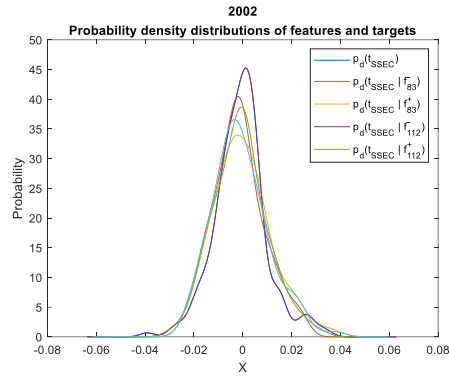


圖 20：2002 年 SSEC 目標與特徵  $f_{83}$ 、 $f_{112}$  條件下的目標機率密度分布

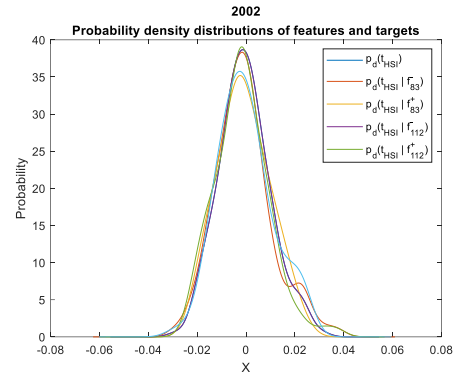


圖 21：2002 年 HSI 目標與特徵  $f_{83}$ 、 $f_{112}$  條件下的目標機率密度分布

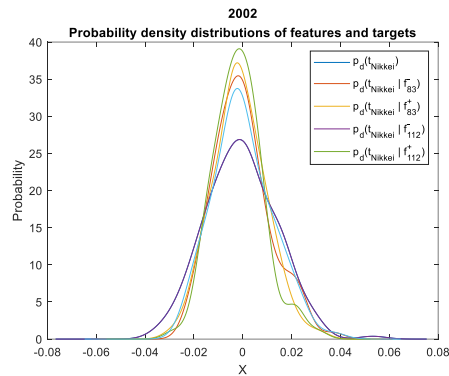


圖 22：2002 年 Nikkei 225 目標與特徵  $f_{83}$ 、 $f_{112}$  條件下的目標機率密度分布

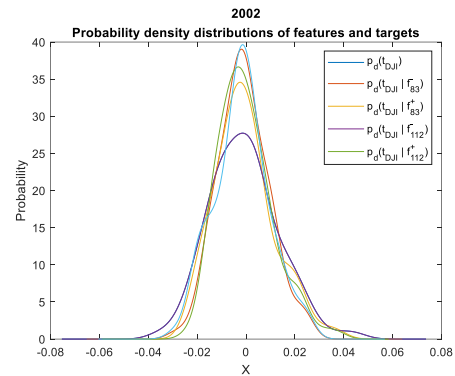


圖 23：2002 年 DJI 目標與特徵  $f_{83}$ 、 $f_{112}$  條件下的目標機率密度分布

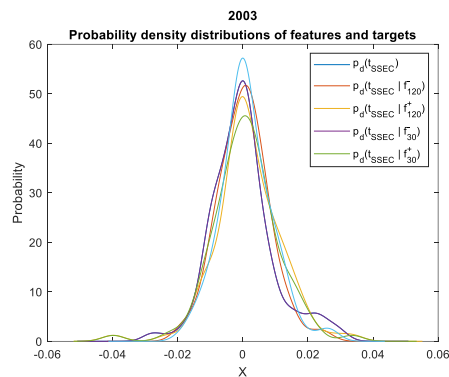


圖 24：2003 年 SSEC 目標與特徵  $f_{120}$ 、 $f_{30}$  條件下的目標機率密度分布

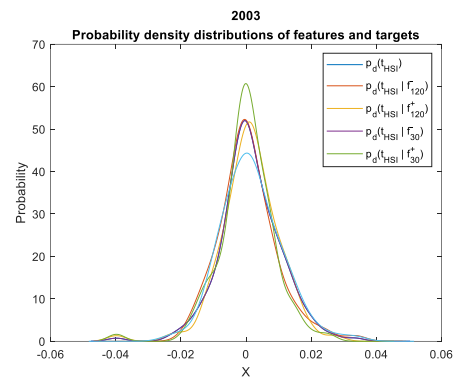


圖 25：2003 年 HSI 目標與特徵  $f_{120}$ 、 $f_{30}$  條件下的目標機率密度分布

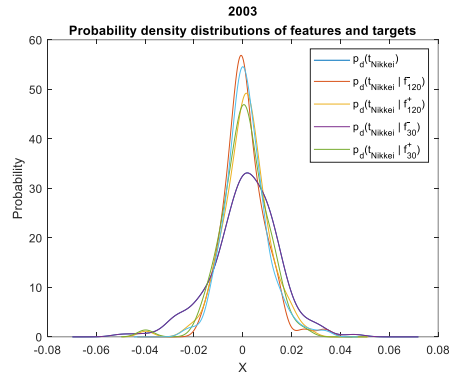


圖 26：2003 年 Nikkei 225 目標與特徵  $f_{120}$ 、 $f_{30}$  條件下的目標機率密度分布

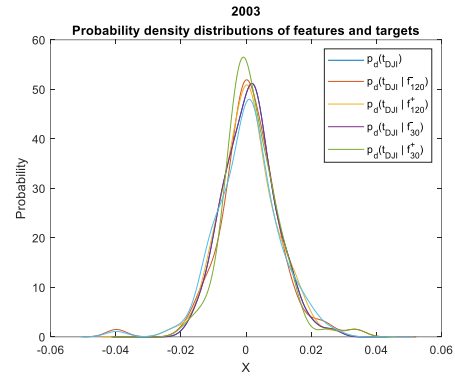


圖 27：2003 年 DJI 目標與特徵  $f_{120}$ 、 $f_{30}$  條件下的目標機率密度分布

實驗結果之 RMSE 列於下表 9：

表 9：實驗 3-1 四目標預測之效能比較 (RMSE)

		RMSE			
		Hsieh et al. [6]	Chen et al. [17]	Zhang et al. [18]	Proposed method
2002	SSEC	-	-	<b>18.0470</b>	20.0222
	HSI	-	118.27	-	<b>105.8136</b>
	Nikkei 225	141	-	-	<b>136.6241</b>
	DJI	132	-	-	<b>129.2588</b>
2003	SSEC	-	-	17.7821	<b>15.1384</b>
	HSI	-	132.67	-	<b>101.6561</b>
	Nikkei 225	177	-	-	<b>127.4672</b>
	DJI	89	-	-	<b>75.5405</b>

將各文獻中的年度資料進行平均之後得實驗效能比較如表 10 所示：

表 10：模型平均之效能比較 (RMSE)

		RMSE			
		Hsieh et al. [6]	Chen et al. [17]	Zhang et al. [18]	Proposed method
SSEC	-	-	-	17.9146	<b>17.5803</b>
HSI	-	-	125.4700	-	<b>103.7349</b>
Nikkei 225	159.0000	-	-	-	<b>132.0457</b>
DJI	110.5000	-	-	-	<b>102.3997</b>

### 4.3.2 多目標與單目標預測之差異性

實驗三的第二部分，本研究使用 2018 年 1 月 1 日至 2018 年 12 月 31 日上海證券交易所綜合股價指數（SSEC）、香港恒生指數（HSI）、日經平均指數

（Nikkei 225）、道瓊工業平均指數（DJI）個別使用本研究所提出的模型進行單目標預測，接著將四個股票市場取交集得共同交易日後作為模型輸入值，進行四目標預測，用以比較同一模型在進行單目標與多目標預測時兩者之間的差異性。

單目標預測實驗中，同樣取每家股市資料集的 83% 作為訓練資料集，其資料對依照原始資料的數量不同而有所差異，挑選後的特徵數為 2，規則數為 3，實驗設定顯示於表 11：

表 11：實驗 3-2 單目標預測實驗設定

	SSEC	HSI	Nikkei 225	DJI
Number of original data	242	245	245	249
Number of data pairs	212	215	215	219
Number of training data	176	179	179	182
Number of rules	3	3	3	3
Number of premise parameters	10	10	10	10
Number of consequent parameters	9	9	9	9

四目標預測中得 SSEC、HSI、Nikkei 225、DJI 共同交易日得 217 筆資料，候選特徵數為 30，故取得 187 筆資料對，其中取 83% 作為訓練資料集，共 156 筆，由於所取資料集的來源較單目標多，故將挑選後的特徵數調整為 10，規則數為 3。實驗設定以表 12 呈現如下：

表 12：實驗 3-2 四目標預測實驗設定

Number of original data	217
Number of data pairs	187
Number of training data	156
Number of rules	3
Number of premise parameters	40
Number of consequent parameters	33

其預測結果如圖 28-35 所示：

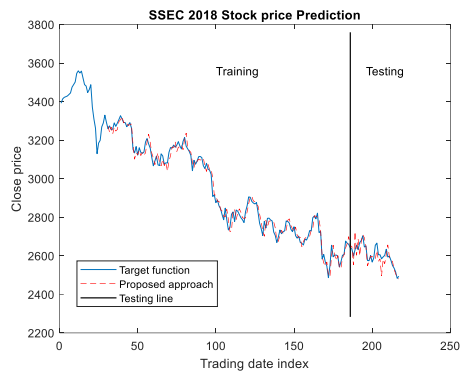


圖 28：SSEC 股市預測線

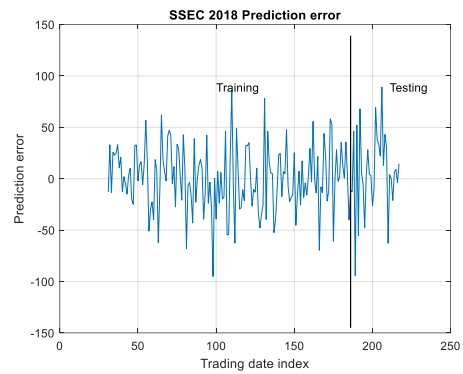


圖 29：SSEC 預測誤差圖

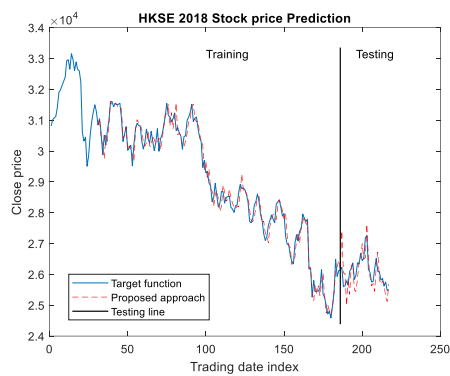


圖 30：HSI 股市預測線

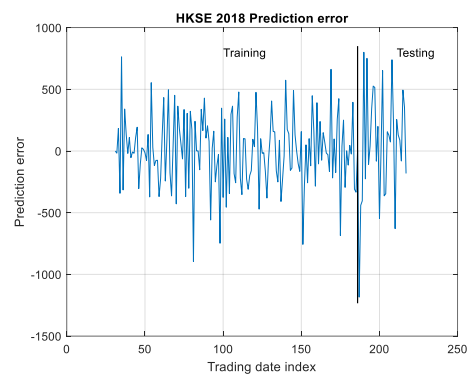


圖 31：HSI 預測誤差圖

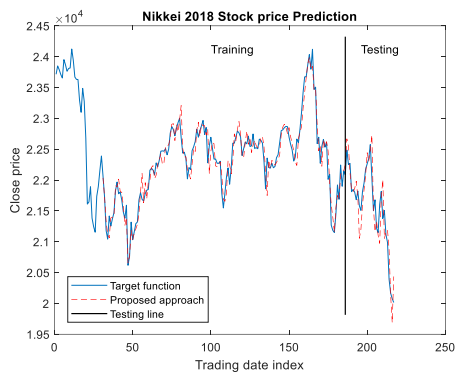


圖 32：Nikkei 股市預測線

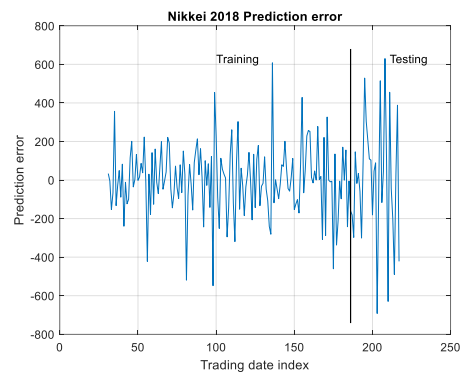


圖 33：Nikkei 預測誤差圖

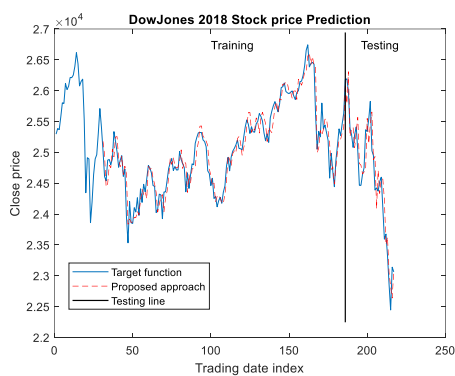


圖 34：DJI 股市預測線

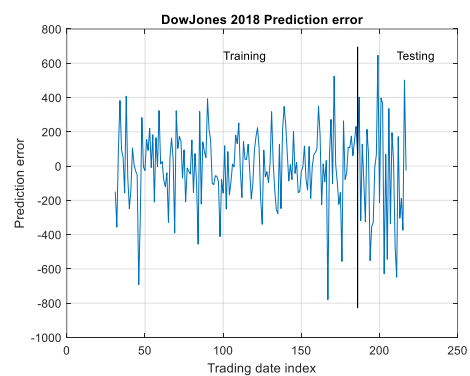


圖 35：DJI 預測誤差圖

與多目標預測與單目標預測之模型效能比較列於下表 13：

表 13：單目標與多目標預測之效能比較（RMSE）

	RMSE			
	SSEC	HSI	Nikkei 225	DJI
1 Target	32.6480	332.7996	238.5166	251.3554
4 Targets	<b>31.6853</b>	<b>306.8782</b>	<b>200.4556</b>	<b>245.1469</b>

由表 13 中可得知，在多目標預測中，SSEC、HSI、Nikkei 225、DJI 個別市場的 RMSE 均優於單目標預測，證實中國與國際市場之間具有相互影響的連動關係，且本研究所提出的模型能夠從多樣且複雜的數據中汲取有效資訊，進而改善多目標金融市場預測的精確度。

## 第五章 結果與討論

本研究提出一個具有特徵挑選的球型複數類神經模糊系統，搭配 WOA-RLSE 複合式學習演算法分別訓練前鑑部與後鑑部的參數，進行單一市場與多目標市場預測，試圖藉此驗證中國與國際市場間的連動性，並透過下列三個實驗驗證本模型的有效性。

實驗一當中本研究透過特徵挑選的數量與順序作為操作變因，用以檢驗特徵挑選架構對於模型預測效能的影響力，故依特徵貢獻指數的排列以順序與倒序方式分別選取 2、4、6、8 個特徵，也就是依據對目標的有效貢獻選取表現最佳與最劣的 2、4、6、8 個特徵，觀察模型學習效能的差異。結果顯示，順序選取特徵的學習效能優於倒序選取特徵的學習效能，證明特徵的貢獻指數能夠正確反映出特徵對於預測目標的貢獻度，達成從多樣化資料中去除雜訊與極端值的目的。除此之外，選取特徵數多的模型其 RMSE 優於特徵數小者，其原因有二，一為特徵數多時所涵蓋的資訊量越多，越能夠反映資料的全貌；二為特徵的數目對模型的整體參數量有直接影響，當模型中的可調整參數量越多則模型的規模愈大提高模型的學習彈性，造成學習效能的提升，但參數量過多可能會增加過度凝合現象的發生機率。

實驗二將本研究所提出的模型與 Zhang et al. [18] 的研究進行比較，採用 2000 至 2006 年海證券交易所綜合股價指數（SSEC）的年度股價資訊作為原始數據，分別依照年度進行單目標預測。在選取特徵數為 10 且規則數為 3 的條件之下，顯示本模型在 2000 至 2006 期間的平均 RMSE 優於 Zhang et al. [18]，證實本研究所提出之模型的有效性。

實驗三分為兩個部分探討之，第一部分當中採用多目標預測與其它學術研究的單目標預測實驗進行比較；第二部分則是使用具有特徵挑選的 SCNFS 模型分別進行單目標與多目標股市預測，用以驗證中國與國際市場之間的相互關係。從實驗結果可得知，在多目標預測下，其個別市場的 RMSE 均優於單目標預測，由此

可知中國市場確實會受到其他國際市場的股價趨勢所影響，反之其他國際市場也同樣會受到中國市場所影響，即證明中國與國際市場具有連動性，亦可證實本研究所提出的模型得以吸收不同來源的資訊，在多目標預測的條件下亦能提升個別市場的預測能力。



## 第六章 結論

本研究中將所提出的多目標時間序列預測模型應用於中國與國際金融市場，透過實驗證實具特徵挑選架構的球型複數類神經模糊系統持有兩大特點：首先，特徵挑選能夠透過判斷特徵對目標的貢獻度有效去除大量多樣化資料中的雜訊與極端值，加以排除數據當中的冗餘性與干擾性。篩選出能夠對目標造成正面影響的特徵則能夠使用較少的數據集為目標變數提供較多有效資訊量，降低模型的運算負擔、減少不必要的浪費，對於多目標預測有極大的幫助。

其二為球型複數類神經模糊系統的運用能夠藉由球型複數模糊集合乘載更多資訊量，透過複數的應用可吸收多維度的資訊進行多目標預測，證實中國與國際市場之間的連動關係，且多國市場的股市資料確實能夠作為單一市場波動的影響因子。同時顯示本研究所提出的模型能夠從不同來源的市場數據當中汲取有用資訊，進而提升多目標股票預測的學習效能。

## 第七章 參考文獻

- [1] Jang, J.-S., “ANFIS: adaptive-network-based fuzzy inference system,” *IEEE transactions on systems, man, and cybernetics*, Vol.23, iss.3, pp.665-685, 1993.
- [2] Jang, J.-S.R., et al., “Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence ” *IEEE Transactions on automatic control*, Vol.42, iss.10, pp.1482-1484, 1997.
- [3] Khashei, M. & M. Bijari, “An artificial neural network (p, d, q) model for timeseries forecasting,” *Expert Systems with applications*, Vol.37, iss.1, pp.479-489, 2010.
- [4] Balkin, S.D. & J.K. Ord, “Automatic neural network modeling for univariate time series,” *International Journal of Forecasting*, Vol.16, iss.4, pp.509-515, 2000.
- [5] Chung, H. & K.-s. Shin, “Genetic algorithm-optimized long short-term memory network for stock market prediction,” *Sustainability*, Vol.10, iss.10, pp.3765, 2018.
- [6] Hsieh, T.-J., et al., “Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm,” *Applied soft computing*, Vol.11, iss.2, pp.2510-2525, 2011.
- [7] Qiu, M. & Y. Song, “Predicting the direction of stock market index movement using an optimized artificial neural network model,” *PloS one*, Vol.11, iss.5, pp.e0155133, 2016.
- [8] Liu, Y., et al., “Stock volatility prediction using recurrent neural networks with sentiment analysis,” *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp.192-201, Springer, 2017
- [9] Pang, X., et al., “An innovative neural network approach for stock market prediction,” *The Journal of Supercomputing*, pp.1-21, 2018.
- [10] Li, X., et al., “An EPC forecasting method for stock index based on integrating empirical mode decomposition, SVM and cuckoo search algorithm,” *Journal of Systems Science and Information*, Vol.2, iss.6, pp.481-504, 2014.
- [11] Wang, W. & H. Nie, “Comparative Study on Forecasting Model for Stock Index Future Price,” *Modern Economy*, Vol.9, iss.04, pp.750, 2018.
- [12] Cheng, C.-H., et al., “Fuzzy time-series based on adaptive expectation model for TAIEX forecasting,” *Expert systems with applications*, Vol.34, iss.2, pp.1126-1132, 2008.
- [13] Xi, L., et al., “A new constructive neural network method for noise processing

- and its application on stock market prediction,” *Applied Soft Computing*, Vol.15, pp.57-66, 2014.
- [14] Bas, E., et al., “A fuzzy time series forecasting method based on operation of union and feed forward artificial neural network,” *American Journal of Intelligent Systems*, Vol.5, iss.3, pp.81-91, 2015.
- [15] Ye, Q. & L. Wei, “The prediction of stock price based on improved wavelet neural network,” *Open Journal of Applied Sciences*, Vol.5, iss.04, pp.115-120, 2015.
- [16] Khuat, T.T., et al., “Forecasting Stock Price using Wavelet Neural Network Optimized by Directed Artificial Bee Colony Algorithm,” *Journal of Telecommunications and Information Technology*, iss.2, pp.43-52, 2016.
- [17] Chen, Y.-S., et al., “A study of ANFIS-based multi-factor time series models for forecasting stock index,” *Applied Intelligence*, Vol.45, iss.2, pp.277-292, 2016.
- [18] Zhang, W., et al., “A multi-factor and high-order stock forecast model based on Type-2 FTS using cuckoo search and self-adaptive harmony search,” *Neurocomputing*, Vol.240, pp.13-24, 2017.
- [19] Wei, Y., et al., “Prediction of Stock Price Trend Based on Wavelet Neural Network and RS Attributes Reduction,” *2017 International Conference on Education, Economics and Management Research (ICEEMR 2017)*, pp.95-98, Atlantis Press, 2017
- [20] Chong, E., et al., “Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies,” *Expert Systems with Applications*, Vol.83, pp.187-205, 2017.
- [21] Chatzis, S.P., et al., “Forecasting stock market crisis events using deep and statistical machine learning techniques,” *Expert Systems with Applications*, Vol.112, pp.353-371, 2018.
- [22] Lei, L., “Wavelet neural network prediction method of stock price trend based on rough set attribute reduction,” *Applied Soft Computing*, Vol.62, pp.923-932, 2018.
- [23] Shastri, M., et al., “Stock Price Prediction using Artificial Neural Model: An Application of Big Data,” *ICST Transactions on Scalable Information Systems*, Vol.6, iss.20, pp.1-8, 2019.
- [24] Cao, W., et al., “Deep modeling complex couplings within financial markets,” *Twenty-Ninth AAAI Conference on Artificial Intelligence*, pp.2518-2524, 2015
- [25] Zadeh, L.A., “Fuzzy sets,” *Information and control*, Vol.8, iss.3, pp.338-353, 1965.
- [26] Ramot, D., et al., “Complex fuzzy sets,” *IEEE Transactions on Fuzzy Systems*, Vol.10, iss.2, pp.171-186, 2002.

- [27] Ramot, D., et al., "Complex fuzzy logic," *IEEE Transactions on Fuzzy Systems*, Vol.11, iss.4, pp.450-461, 2003.
- [28] Li, C. & T.-W. Chiang, "Complex neuro-fuzzy self-learning approach to function approximation," *Asian Conference on Intelligent Information and Database Systems*, pp.289-299, Springer, 2010
- [29] 張克群, et al., "適應性類神經模糊推論系統於財務危機預測之應用," *臺灣銀行季刊*, Vol.3, iss.61, pp.61-75, 2010.
- [30] Takagi, T. & M. Sugeno, "Derivation of fuzzy control rules from human operator's control actions," *IFAC Proceedings Volumes*, Vol.16, iss.13, pp.55-60, 1983.
- [31] Tu, C.-H. & C. Li, "Multiple Function Approximation-A New Approach Using Complex Fuzzy Inference System," *Asian Conference on Intelligent Information and Database Systems*, pp.243-254, Springer, 2018
- [32] Mirjalili, S. & A. Lewis, "The whale optimization algorithm," *Advances in engineering software*, Vol.95, pp.51-67, 2016.
- [33] Li, C. & T. Wu, "Adaptive fuzzy approach to function approximation with PSO and RLSE," *Expert Systems with Applications*, Vol.38, iss.10, pp.13266-13273, 2011.
- [34] 李俊賢 & 江泰緯, "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*, Vol.15, iss.1, pp.137-158, 2013.
- [35] 劉福才, et al., "基于智能优化算法的 TS 模糊模型辨识," *系统工程与电子技术*, Vol.35, iss.12, pp.2643-2650, 2010.
- [36] Tu, C.-H. & C. Li, "A Novel Entropy-Based Approach to Feature Selection," *Asian Conference on Intelligent Information and Database Systems*, pp.445-454, Springer, 2017
- [37] Shannon, C.E., "A mathematical theory of communication," *Bell system technical journal*, Vol.27, iss.3, pp.379-423, 1948.
- [38] Girolami, M. & C. He, "Probability density estimation from optimally condensed data samples," *IEEE Transactions on pattern analysis and machine intelligence*, Vol.25, iss.10, pp.1253-1264, 2003.