

國立中央大學

資 訊 管 理 研 究 所  
碩 士 論 文  
(初稿)

改良式複數模糊類神經系統於時間序列預測之  
應用

Application of Modified Complex Neural Fuzzy  
System in Time Series Forecasting

研 究 生：吳清鈺

指導教授：李俊賢 博士

中 華 民 國 107 年 6 月

## 改良式複數模糊類神經系統於時間序列預測之應用

研究生：吳清鈺

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

本研究針對時間序列預測之問題，提出了一個改良式的複數模糊類神經系統。在模型輸入的選擇上，本研究基於夏農資訊熵(Shannon Entropy)的理論提出了多目標特徵挑選方法，其主要是透過計算特徵對於目標所提供的資訊量，並以一套特徵選取策略針對特徵提供的資訊量多寡進行特徵的挑選。在模型建立上是使用結構學習(Structure learning)的方式，以數據本身客觀地建構模型，改變了傳統以主觀的方式決定模型的建立。複數模糊集合(Complex fuzzy sets, CFSs)比以往傳統的模糊集合具有更佳的解釋能力，運用於模糊類神經系統(Neural fuzzy system)中能夠傳遞更大量的資訊，增加模型預測的效能，並且能夠讓模型同時預測多達6個目標。在模型參數學習階段，將粒子群最佳化演算法(Particle swarm optimization, PSO)與遞迴最小平方估計法(Recursive least squares estimator, RLSE)結合形成PSO-RLSE複合式演算法。PSO用於模型前鑑部之參數優化，而RLSE則負責模型後鑑部之參數更新，此種分而擊之的概念能夠降低模型參數的維度，提升找到模型參數最佳解的機率，並能降低模型整體的訓練時間。根據模型輸出與目標間的誤差，本研究提出了一種誤差回饋的方式，藉由將誤差回饋到模型的後鑑部以改良模型預測時間序列的準確度。本篇研究使用財經市場的時間序列進行多目標預測的實驗，並從實驗結果顯示本論文所提出之模型與其他文獻相比有較好的預測結果。

**關鍵字：**多目標特徵挑選、結構學習、複數模糊類神經系統、PSO-RLSE 複合式演算法、誤差回饋、時間序列預測

## **Application of Modified Complex Neural Fuzzy System in Time Series Forecasting**

Graduate student: Chingyu Wu

Advisor: Dr. Chunshien Li

Department of Information Management, National Central University, Taiwan

### **Abstract**

This study proposes a modified complex fuzzy neural system for the problem of time series prediction. In the selection of model input, this study proposes a multi-targets feature selection method based on the theory of Shannon entropy. The main purpose is to select features by calculating the amount of information provided by the features for the target, and using a set of feature selection strategies to do the multi-targets feature selection. In the establishment of the model, we change the traditional subjective way to objective way by using the training data itself to construct the model setting. Complex fuzzy sets (CFSs) have better interpretive capabilities than traditional fuzzy sets. They can deliver a larger amount of information than traditional fuzzy sets in neural fuzzy systems, increase the effectiveness of model prediction, and let the model predict up to 6 targets at the same time. In the model parameter learning phase, Particle swarm optimization (PSO) and Recursive least squares estimator (RLSE) are combined to form a PSO-RLSE hybrid algorithm. PSO is used to optimize the parameters of the model's premise part, while RLSE is responsible for the parameters of the model's consequent part. This concept of divide-and-conquer can reduce the dimension of the model parameters, increase the probability of finding the best solution of the model parameters, and reduce the overall training time of the model. This study proposes an error feedback method according to the errors between the model outputs and the targets. We increase the model prediction accuracy by feeding the error back to the model's consequent part. This study uses the time series of the financial market to conduct the multi-targets forecasting experiments. The experimental results show that the model proposed in this paper has better forecasting abilities than other literatures.

**Keywords** : multi-targets feature selection, structure learning, complex neural fuzzy system, PSO-RLSE hybrid algorithm, error feedback, time series forecasting

## 目錄

章節	頁碼
摘要 .....	i
Abstract.....	ii
目錄 .....	iii
圖目錄 .....	v
表目錄 .....	vi
第一章 緒論 .....	1
1.1 研究背景與目的 .....	1
1.2 研究方法 .....	2
1.3 論文架構 .....	3
第二章 文獻探討 .....	4
2.1 特徵選取 .....	4
2.2 模糊集合理論與模糊規則 .....	5
2.3 複數模糊集合理論 .....	8
2.4 模糊類神經系統 .....	8
2.5 機器學習演算法 .....	9
2.5.1 粒子群最佳化演算法 .....	11
2.5.2 遞迴最小平方估計法 .....	11
第三章 系統設計與架構 .....	13
3.1 多目標特徵挑選 .....	13
3.2 結構學習 .....	15
3.3 多目標複數模糊類神經模型 .....	16
3.4 複合式 PSO-RLSE 演算法 .....	21
第四章 實驗實作與結果 .....	23
4.1 實驗 1: 單目標股票指數預測-SSECI(2006) .....	23
4.2 實驗 2: 雙目標股票指數預測- DJIA(2004) 、TAIEX(2004) .....	27

4.3 實驗 3: 四個目標股票指數預測- TAIEX(2006)、DJIA(2006)、Nikkei(2006)、FTSE(2006)	
.....	34
第五章 實驗結果討論 .....	42
第六章 結論 .....	49
6.1 結論 .....	49
6.2 未來研究方向 .....	50
參考文獻 .....	51

## 圖目錄

圖	頁次
圖 1 冷氣機的模糊規則範例 .....	7
圖 2 模型預測與目標之誤差圖 .....	17
圖 3 誤差機率密度分布圖 .....	17
圖 4 模型訓練與預測結果(SSECI) .....	26
圖 5 訓練結果與目標之誤差(SSECI) .....	26
圖 6 模型學習曲線(實驗一).....	27
圖 7 模型訓練與預測結果(DJIA).....	31
圖 8 模型訓練與預測結果(TAIEX).....	32
圖 9 訓練結果與目標之誤差(DJIA).....	32
圖 10 訓練結果與目標之誤差(TAIEX).....	33
圖 11 模型學習曲線(實驗二).....	33
圖 12 模型訓練與預測結果(TAIEX).....	38
圖 13 模型訓練與預測結果(DJIA).....	38
圖 14 模型訓練與預測結果(Nikkei) .....	39
圖 15 模型訓練與預測結果(FTSE) .....	39
圖 16 訓練結果與目標之誤差(TAIEX).....	40
圖 17 訓練結果與目標之誤差(DJIA).....	40
圖 18 訓練結果與目標之誤差(Nikkei) .....	41
圖 19 訓練結果與目標之誤差(FTSE) .....	41
圖 20 模型學習曲線(實驗三) .....	42
圖 21 誤差機率密度分布圖(實驗一：SSECI) .....	43
圖 22 誤差機率密度分布圖(實驗二：DJIA).....	44
圖 23 誤差機率密度分布圖(實驗二：TAIEX).....	45
圖 24 誤差機率密度分布圖(實驗三：TAIEX).....	46
圖 25 誤差機率密度分布圖(實驗三：DJIA).....	47
圖 26 誤差機率密度分布圖(實驗三：Nikkei) .....	47
圖 27 誤差機率密度分布圖(實驗三：FTSE).....	48

## 表目錄

表	頁次
表 1 模型參數配置(實驗一) .....	24
表 2 演算法參數配置(實驗一) .....	24
表 3 10 次實驗之結果(實驗一) .....	25
表 4 與其他文獻之比較(實驗一) .....	25
表 5 模型參數配置(實驗二) .....	28
表 6 演算法參數配置(實驗二) .....	29
表 7 10 次實驗之結果(實驗二) .....	30
表 8 與其他文獻之比較(實驗二) .....	30
表 9 與其他文獻之比較(實驗二) .....	31
表 10 模型參數配置(實驗三) .....	35
表 11 演算法參數配置(實驗三) .....	35
表 12 10 次實驗之統計結果(實驗三) .....	36
表 13 與其他文獻之比較(實驗三) .....	37
表 14 誤差機率密度分布參數(實驗一) .....	44
表 15 誤差機率密度分布參數(實驗二) .....	45
表 16 誤差機率密度分布參數(實驗三) .....	46

## 第一章 緒論

### 1.1 研究背景與目的

德國於 2011 年在漢諾瓦工業博覽會上提出了工業 4.0 的概念，並在 2013 年將工業 4.0 訂為國家發展政策。隨後，美國也提出了基於工業 4.0 概念的計畫，旨在將人、數據與智能設備銜接，並以交換數據的方式推動工業轉型。工業 4.0 將虛擬網路與實體連接，通過資訊技術整合工業生產過程，以達到更高效率的生產系統。主要方法是將工廠的機具設備、控制器與虛擬網路相連接，將各種在工業現場所擷取到的巨量資料，透過網路上傳到雲端資料中心執行大數據分析，並利用人工智慧等技術強化生產效率，達到智慧生產能力。自從工業 4.0 的概念被提出後，全球也掀起了工業 4.0 的熱潮，許多大型企業公司也開始將這樣的概念運用於工廠製造之中。

在工業 4.0 中所使用到的人工智慧技術最早開始於 1943 年到 1956 年，其是由許多不同領域的科學家探討建造人工大腦的可能性，並在 1956 年達特茅斯學院所舉行的會議上正式確立了人工智慧這門研究領域。但自從確立人工智慧的研究領域後其發展卻不如預期，當時的計算機計算速度尚未提升、儲存空間有限、數據量更是不足夠，因此不足以解決任何實際的問題。時至今日，因為科技快速的發展，使得硬體儲存裝置的成本降低且資料量迅速的增加，在人工智慧的研究上也得以拓展。2015 年時，Google 公司開發的 AlphaGo 以人工智慧的深度學習技術成功地擊敗了世界頂尖的職業圍棋棋士，再次開啟了眾人對於人工智慧相關研究的熱忱。許多企業開始結合人工智慧的技術應用於各個產業中，眾多學者也積極地針對人工智慧領域進行研究。Malliaris 和 Salchenberger[1]利用類神經網路的技術針對 S&P 500 進行波動性之預測，結果證實類神經網路模型較其他模型更能夠準確預測出股價的波動性。Khashei 和 Bijari [2]使用了一種結合自回歸整合移動平均與類神經網路的混合式模型，並與一般的類神經網路相比，結果表明其更能夠精準地針對時間序列進行預測。學者 Jang [3]提出了自適應網路的模



糊推理系統(Adaptive-network-based fuzzy inference system, ANFIS)，ANFIS 是在自適應網路框架下以 If-Then 的規則來預測時間序列，與類神經網路和早期的模糊建模方式相比，擁有較精確的預測能力。

此外，近年大多研究都是以一個模型預測一個目標，但是在科技進步以及需求的增加下，此方式進行預測並無太大的效益。若是能夠以一個模型以及一組參數，進行多目標預測，便能更容易看出輸入資料對於各個目標之間的關係，協助決策進行。通常模型會經過機器學習來得出一組最佳化的參數，以此方法加強模型的預測精準度。但是當機器學習演算法面臨求解的參數維度過大時，會影響演算法的活動性，以至於落入區域最佳解的問題。因此，本研究除了進行機器學習以提高模型精準度，另外藉由模糊類神經模型的輸出與目標間產生之誤差，提出了一種誤差回饋的方式來優化模型參數並提高模型精準度。

## 1.2 研究方法

本研究在模型的輸入特徵挑選上是以夏農的資訊理論為基礎，計算出特徵資料的資訊熵(Entropy)，並以特徵的資訊熵進一步計算出特徵的資訊量，根據特徵資訊量的多寡定訂一個多目標的特徵挑選策略來選擇最後模型的輸入特徵。

本篇論文主要是以模糊類神經模型來進行時間序列的預測。時間序列通常由長期趨勢、循環變動、季節變動及不規則變動所組成，因此其具有自然雜亂、非平穩等特性。自然雜亂的特性讓觀測者無法透過過去的行為以及資訊來充分掌握過去與未來價格之間的關連性，因此時間序列的預測其實是一個極具挑戰性的問題。模糊類神經模型中的模糊規則系統是使用 Takagi-Sugeno (T-S)型態的 If-then 模糊規則系統，使模型能夠更易於人們理解。在模糊集合的應用上則是採用複數模糊集合(Complex fuzzy set, CFS)，其

將一般模糊集合的歸屬度由 $[0, 1]$ 區間延伸至複數平面的單位圓盤(The unit disc of the complex plane)。使用複數模糊集合，元素可獲取複數歸屬度(Complex-valued membership degree)，亦即可以同時產生實部歸屬度以及虛部歸屬度。複數模糊集合能夠使模型進行多目標的預測，並同時提高了模型的預測準確率，因此在本篇論文的研究模型中採用複數模糊集合以形成複數模糊類神經模型。在模型效能的驗證上則是使用財務金融領域裡的時間序列資料來進行模型的訓練與預測。

在機器學習的演算法方面，本研究為了要最佳化模型的系統參數，因此將系統參數分成兩個子集合，分別為前鑑部(Premise part)參數及後鑑部(Consequence part)參數。前鑑部參數的優化是以粒子群最佳化演算法(Particle swarm optimization, PSO)來進行，期望能夠快速的搜尋到最佳解；後鑑部則是以遞迴最小平方估計法(Recursive least-squares estimation, RLSE)來調整。使用遞迴最小平方估計法能夠以線性計算求解的方式在短時間內計算出最佳近似解，是一個非常有效率且非常準確的演算法。本研究利用複合式的方法(PSO-RLSE)來達到參數快速收斂的目的，並預計能夠得到良好的模型預測準確性。針對模型產生之誤差，本研究提出一個新的誤差回饋方法。將模型產生的誤差回饋到模型後鑑部中，藉此方法來降低模型輸出與目標之間的誤差，提升模型預測的準確性。

### 1.3 論文架構

本篇論文共有六章，第一章為研究緒論，介紹研究的背景與目的以及使用何種方法來完成。第二章為文獻探討，探討的文獻將針對研究所使用到的理論、模型或演算法進行文獻回顧，包括特徵選取、模糊集合理論與模糊規則、複數模糊集合理論、模糊類神經系統、粒子群最佳化演算法以及遞迴最小平方估計法。第三章為系統設計與架構，介紹本論文的特徵挑選方法、結構學習方法、本研究之模型將如何運作、如何利用複合式的演算法於模型之中。第四章為實驗實作與結果，實驗將採用財金市場的歷史資料進行

模型的訓練與預測，預測的結果將與其他文獻研究結果相互比較。第五章為實驗結果與研究方法的討論，對實驗的結果進行討論。第六章則是本篇論文的結論、貢獻以及未來的研究方向。

## 第二章 文獻探討

### 2.1 特徵選取

特徵選取(Feature selection)的目的是在訓練資料建模前，從原有的特徵集合中過濾掉與目標不相關或冗餘的特徵，以便挑選出最佳的特徵子集來減少學習演算法計算時的複雜度、提高機器學習的效率，並進一步地提升模型預測的準確率。特徵選取方法目前可以大致分為三類，分別是包裝法(Wrapper)、嵌入法(Embedded)及過濾法(Filter)。

包裝法根據某些準則從候選特徵集合中挑選出各種不同的特徵子集，再將這些特徵子集依序輸入至同一個模型中，並使用測試資料集來進行測試。透過計算測試資料集上的錯誤次數來對特徵子集進行評分，並根據各特徵子集的分數判斷是否選擇或刪除特徵。包裝法將最終的模型性能作為評價函數，因此往往能為特定類型的模型找到性能最好的特徵子集，但若特徵挑選完成後使用的模型與特徵挑選時使用的模型不一致時，其效果可能會不如預期。由於包裝法需要將每個特徵子集輸入至模型中以得到每一子集的評分，因此需要耗費大量計算資源才能完成。常見의包裝法如循序向後選取法(sequential backward selection, SBS)及循序向前選擇法(sequential forward selection, SFS)。SBS[4]由 Marill 和 Green 在 1963 年提出，是一種由上往下的特徵選取搜索方法。主要概念是將所有特徵逐一淘汰，直到剔除特徵後的特徵子集評分最高即停止。SFS[5]則是由 Whitney 在 1971 年提出，是一種由下往上的特徵選取搜索方法。主要概念是令特徵子集從空集開始，每次選擇一個特徵加入特徵子集中，直到該子集的評分最高即停止。此兩種方法皆是貪婪演算法，因此可能找到局部最優的特徵子集，但不一定是全域最佳特徵子集。

嵌入法直接將特徵選取方法嵌入到訓練模型之中，在訓練模型的過程中同時對各個特徵進行評估，並依據特徵的評分結果對特徵進行挑選或剔除。由於嵌入法直接將特徵選取方法嵌入到訓練模型之中，因此其與包裝法一樣會相依於模型。常見的方法如構建線性模型的套索方法(least absolute shrinkage and selection operator, LASSO)，此方法對回歸係數加入了懲罰項，因此會使不重要的特徵評分結果趨近於 0，不為 0 的特徵則為最後選入的特徵。在文獻[6]中，學者等人突破了 Lasso 僅能使用在線性模型的侷限性，其將 Lasso 方法轉化成與支援向量機(Support vector machine, SVM)類似的形式，使其能夠擴展到非線性模型的特徵挑選上，並透過實驗找出了與非線性相關的少數主要特徵，驗證了其有效性。

過濾法與包裝法類似，都是運用搜尋的機制來找到適當的特徵子集，然而，其並非使用模型性能作為評價函數，而是直接以特徵對目標之間的關係進行評估。例如使用關聯程度(correlation)、共同資訊量(mutual information)，以及熵(entropy)等統計或資訊理論指標，透過間接衡量資料自身特性的方法來評斷是否為重要特徵，因此不會有相依於模型的問題且計算複雜度也較低，常見的方法像是方差選擇法或相關係數法。使用方差選擇法時必須先計算出各個特徵的方差，接著挑選方差大於門檻值的特徵當作選入的特徵；相關係數法則是計算出各個特徵對目標之間線性相關程度的量，並以此來進行特徵挑選。在文獻[7]中，學者 Tourassi 等人使用互資訊用於電腦輔助診斷模型的特徵挑選上，並從實驗結果表明這個該方法能夠有效的選擇出最佳的特徵子集，也與其他特徵選擇法相比來證明此方法的穩健性。

## 2.2 模糊集合理論與模糊規則

在一般集合中，元素的隸屬度(Membership degree)是以二元方式來評估，即元素與集合的關係只有完全不隸屬(記為 0)或完全隸屬(記為 1)之關係。然而，人類的思考和推

理往往是不精確、不完整或概括式的模糊信息。例如，對於一些人而言，20 歲是年輕人，而另一些人則是 30 歲，這個“年輕”的概念沒有清晰的界限，因此清晰的集合很難表達人的思想和專業知識。

1965 年，模糊集合的概念由 L. A. Zadeh [8] 提出，其將一般集合的隸屬度概念擴展，以形成模糊集合。對比於一般集合，模糊集合允許元素部分隸屬於集合，也就是說，元素隸屬於模糊集合的程度介於 0 和 1 之間，歸屬程度越接近 0 表示隸屬於集合的程度越低，反之，若歸屬程度越接近 1 表示隸屬於集合的程度越高。根據隸屬函數(Membership function)，從一般集合到模糊集合的隸屬概念之擴展，在數學上可以表達為整數隸屬度到實數隸屬度的擴展，也就是說，模糊集合將隸屬函數的範圍從 $\{0,1\}$ 擴展到單位實數區間 $[0,1]$  來描述元素 $h$ 屬於集合 $A$ 的歸屬程度，表示如下。

$$A = \{h, \mu_A(h) | h \in H, \mu_A(h) \in [0,1]\} \quad (1)$$

其中 $\mu_A(\cdot)$ 為隸屬函數， $\mu_A(h)$ 為 $h$ 對於 $A$ 的隸屬程度。

相對於傳統非 0 即 1 的推論方式，模糊規則利用不同集合間的歸屬函數取得與輸入相對應的歸屬程度，然後綜合各規則的模糊關聯性(Fuzzy relations)得到適當的推論。模糊的關聯性運算“and”有許多種的運算子，常見的像是取最大值(Max)、最小值(Min)或相乘(Multiply)，一般較常使用相乘來進行運算，因相乘是最快速且容易理解的。模糊規則可由專家提供或是知識擷取來建立，If-Then 的模糊規則如圖 1，圖 1 以室溫及溫差做為輸入變數來判斷冷氣機的風速。

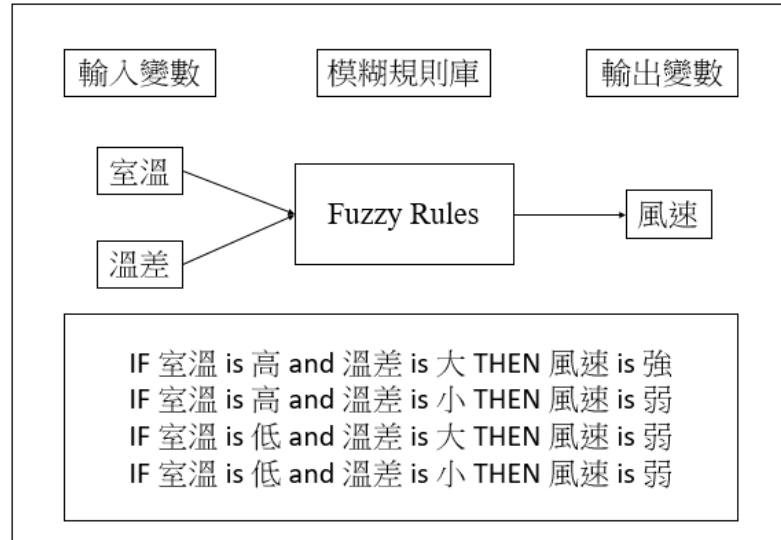


圖 1 冷氣機的模糊規則範例

模糊規則常見的形式有兩種，分別是 Mamdani[9]型態的模糊規則以及 T-S(Takagi Sugeno)型態的模糊規則。Mamdani 型態的模糊規則又稱為語意式模糊規則，是學者 Mamdani 等人在 1974 年研究發表中以研究室所製作的蒸汽引擎模型為對象去驗證模糊控制。T-S 型態的模糊規則是在 1985 年時由 Takagi 與 Sugeno[10]兩位學者所提出的，並且也成功的運用於各種機械的控制器中。兩種型態的模糊規則最主要的差異在於 If-Then 規則中 Then part 的反應，Mamdani 型態的模糊規則 Then part 輸出為一模糊量，而 T-S 型態的模糊規則則是一個線性函數，如方程式(2)、(3)。

Mamdani 型態的模糊規則

$$R^{(q)}: IF \ x_1 \text{ is } A_1^{(q)}(h_1) \text{ and } x_2 \text{ is } A_2^{(q)}(h_2) \text{ and } \dots \text{ and } x_M \text{ is } A_M^{(q)}(h_M),$$

$$\text{Then } Z^{(q)} = B^{(j)} \quad (2)$$

T-S 型態的模糊規則

$$R^{(q)}: IF \ x_1 \text{ is } A_1^{(q)}(h_1) \text{ and } x_2 \text{ is } A_2^{(q)}(h_2) \text{ and } \dots \text{ and } x_M \text{ is } A_M^{(q)}(h_M),$$

$$\text{Then } Z^{(k)} = a_0^{(k)} + a_1^{(k)}h_1 + \dots + a_M^{(k)}h_M \quad (3)$$

## 2.3 複數模糊集合理論

在2002年時學者Ramot[11]等人提出了複數模糊集合(Complex fuzzy sets, CFSs)的概念，將一般模糊集合的歸屬度由 $[0, 1]$ 區間延伸至複數平面的單位圓盤(The unit disc of the complex plane)，換言之，歸屬度就在複數平面的單位圓盤之中。使用複數模糊集合，元素可獲取複數歸屬度(Complex-valued membership degree)，亦即，可以同時產生實部歸屬度以及虛部歸屬度。複數的歸屬程度能夠負載著更大量的訊息，並且反映隱藏在真實世界背後未知的訊息，更重要的是加入複數概念後能夠提升模糊集合的映射能力。假設 $S$ 為一複數模糊集合，其可表示成以下形式。

$$S = \{(h, \mu_s(h)) | h \in U\} \quad (4)$$

其中 $U$ 代表複數模糊集合之論域。而歸屬函數定義如下。

$$\begin{aligned} \mu_s(h) &= r_s(h) \exp(j\omega_s(h)) = \text{Re}(\mu_s(h)) + j\text{Im}(\mu_s(h)) \\ &= r_s(h) \cos(\omega_s(h)) + jr_s(h)\sin(\omega_s(h)) \end{aligned} \quad (5)$$

其中 $j = \sqrt{-1}$ ； $h$ 為複數模糊集合下的數值變數； $r_s(h) \in [0, 1]$ 是複數歸屬度的振幅函數； $\omega_s(h)$ 則是相位函數； $\text{Re}(\cdot)$ 和 $\text{Im}(\cdot)$ 分別表示實部和虛部。當 $\omega_s(h)$ 等於零時，則為傳統模糊集合，可被視為複數模糊集合的特例。

## 2.4 模糊類神經系統

類神經網路主要的原理是模仿生物的神經系統，透過每層所建立的神經元(Neuron)相互連結所組成。這些神經元以平行且分散的方式運算，再經由不斷的學習獲得個別神經元的權重參數，因此對於不明確的輸入也能得到相對應的正確輸出。但類神經網路仍然有一些缺點，其中最大的缺點是類神經網路如同一個黑盒子，對於人類而言非常不易瞭解其內容；模糊推論系統(Fuzzy inference system, FIS)則是用 If-Then 的規則形式來提取人類的經驗和知識。學者 Lin 和 Lee [12]將類神經網路多層分散式的結構和模糊系統

的模糊推理能力互相結合，使得模糊類神經推論系統(Neuro-fuzzy inference system)能夠擁有優秀的推理能力，進而從觀測數據中捕獲前後因果的潛在關係，提升其預測準確性，也增加了模型內部的透明性，改善了類神經模型的缺點。

自從模糊類神經系統被提出後就開始被廣泛的運用，如經濟、科學、商業、管理等方面。Rojas et al. [13]利用模糊類神經系統與自回歸移動平均模型的混合式方法對時間序列進行預測，並觀察此方法對於預測時間序列之可靠性。Sharifian et al. [14]利用模糊類神經系統對風電進行預測，並在結果中驗證了此方法對於風電預測的有效性和適用性，為電力系統控制中心如何精確進行風電預測提供了實用的解決方案。Tang et al. [15]提出了一種改良式的模糊類神經網路的建模方法用於預測多步前進的行車速度。結果表明此方法之預測性能優於傳統模型，並且隨著預測時間步長的增加，此模型之預測誤差上升速度仍較其他模型緩慢。Kristjanpoller 和 Michell[16]利用 ANFIS 確定每個市場指標的個體影響，並使用類神經網路來提高廣義自回歸條件異方差模型預測的預測精度，結果表明這種方法能夠提供更準確的波動率估計。Yolcu 和 Alpaslan[17]使用了模糊類神經系統並搭配粒子群演算法以優化參數，並將此方法用於十二個不同的時間序列，進行了二十四次測量，證明了該方法優越的預測性能。Hou et al.[18]將差分算法與倒傳遞演算法相結合，提出了一種改進的差分倒傳遞演算法來提高模糊類神經系統對於時間序列預測的準確性。實驗結果表明，該算法的收斂速度和預測精度皆優於傳統的模糊類神經系統。

## 2.5 機器學習演算法

人工智慧的發展從一開始的以「推理」為重點，到以「知識」為重點，再到最後以「學習」為重點。顯然，機器學習是實現人工智慧的一個方式，我們可以透過機器學習來解決人工智慧中的問題。機器學習是從大量的資料中進行訓練並自動分析獲得規律，再利用這些規律對未知的資料進行預測。因此機器學習演算法主要是設計和分析一些讓電腦可以自動學習的演算法。



機器學習可分為四種不同的類別，分別是監督學習、無監督學習、半監督學習以及增強學習。監督學習(Supervised learning)是從給定的訓練資料集中學習出一個模式，而此訓練資料集須包括特徵和目標。當新的資料輸入進模型時，模型即可根據這個模式來預測結果，常見的監督學習演算法有回歸分析和統計分類。非監督學習(Unsupervised learning)與監督學習相比，其訓練資料集沒有任何目標，故機器在學習時並不知道其分類結果是否正確。訓練時僅須對機器提供輸入，它會自動從這些輸入中找出潛在的規則，常見的無監督學習演算法為分群演算法。半監督學習(Semi-supervised learning)則是介於監督學習與無監督學習之間。增強學習(Reinforcement learning)則是透過觀察環境而行動，並會隨時根據新進的資料逐步修正，最終得到正確的結果。

機器學習的演算法會直接影響到機器學習的效能，因此使用何種演算法是非常重要的，常見的機器學習演算法像是基因演算法 (Genetic Algorithm, GA)、蟻群演算法(Ant Colony Optimization, ACO)、禁忌搜索演算法 (Tabu Search, TS)、人工蜂群演算法(Artificial bee colony, ABC)等。基因演算法是模擬生物演化過程的方式來解決最佳化的問題；蟻群演算法是模擬蟻群尋找食物的過程，透過螞蟻找到食物時會分泌賀爾蒙的方式來找出問題的最佳解；禁忌搜索演算法是模擬人類的記憶功能，利用短期記憶記錄禁忌名單並使用長期記憶來記錄較佳的解，透過此兩種方式來找出最佳解；人工蜂群演算法是模擬蜜蜂尋找食物的過程所提出之演算法，透過個別蜜蜂分工進行不同活動，並進行蜂群訊息的共享和交流，從而找到問題的最優解。較常用於類神經網路的演算法則是倒傳遞法(Backpropagation)，倒傳遞法利用類神經網路輸出之結果與目標的誤差，反向調整類神經網路的權重參數，借由此方法來找出最佳解，但倒傳遞法並不容易找到最佳解且非常容易陷入區域最小解，因此其使用率在類神經網路中逐漸地降低。本篇論文研究也將倒傳遞法替代為粒子群演算法，並與遞迴最小平方估計法結合形成複合式的演算法用於本研究的複數模糊類神經模型之中。

### 2.5.1 粒子群最佳化演算法

粒子群最佳化演算法(Particle swarm optimization, PSO)[19]最早是由Kennedy和Eberhart兩位學者於1995年提出，其概念源自於鳥群的覓食行為。根據粒子群最佳化演算法的原理，每個粒子都可視為一隻鳥，鳥群則可視為粒子群，而食物的位置即是最佳解位置。在搜尋空間中，每個粒子會依據三種搜尋記憶來改變其搜尋方向，分別是個體粒子當前的位置、個體粒子搜尋到的最佳解位置以及群體粒子的最佳經驗。粒子除了會依照自己曾搜尋過之最佳位置移動，亦會受到群組之最佳經驗的影響而改變移動方向，以此達成快速搜尋的目的。個別粒子之速度與位置更新公式如下：

$$V_i(k+1) = \omega V_i(k) + c_1 \xi_1 (Pbest_i(k) - L_i(k)) + c_2 \xi_2 (Gbest(k) - L_i(k)) \quad (6)$$

$$L_i(k+1) = L_i(k) + V_i(k+1) \quad (7)$$

其中， $L_i(k)$ 為在第 $k$ 次迭代時第 $i$ 顆粒子的位置； $Pbest_i(k)$ 為在第 $k$ 次迭代第 $i$ 顆粒子的最佳位置； $Gbest(k)$ 為在第 $k$ 次迭代時群裡所有粒子的最佳位置； $\{c_1, c_2\}$ 為粒子的學習速率參數； $\omega$ 為速度的慣性參數； $\{\xi_1, \xi_2\}$ 為介於 $[0,1]$ 之間的隨機亂數，目的是要避免陷入區域最小解，增加找到全域最佳解的機率。

粒子群最佳化演算法可以相當快地去逼近最佳解，因此可以有效地對系統的參數進行優化。粒子群最佳化演算法的本質是利用當前位置、全局極值和個體極值 3 個訊息指導粒子下一步迭代位置，充分利用自身經驗和群體經驗調整自身的狀態是粒子群算法具有優異性的關鍵。

### 2.5.2 遞迴最小平方估計法

遞迴最小平方估計法[20](Recursive least squares estimator, RLSE)是從最小平方估計法(Least squares estimator, LSE)加入遞迴的概念而來，適用於線性回歸模型的最佳化問題。最小平方估計法的概念是從已知的訓練資料中以數學計算式直接找出一條最佳的方程

式解，雖然這個方程式有可能不會是最佳解，但會是最佳的近似解。遞迴最小平方估計法不像粒子群最佳化演算法一樣需要時間去訓練找到最佳解，而是透過數學計算式來找到最佳近似解，因此能夠在短時間內就計算出來。給定一訓練資料集 $\{(u, y)^{(i)}, i = 1, 2, \dots, N\}$ ，其線性回歸模型之定義如下所示：

$$y = \sum_{i=1}^N \theta_i f_i(u) + \varepsilon \quad (8)$$

其中 $u$ 為模型輸入； $y$ 為目標值； $\{f_i(\cdot), i = 1, 2, \dots, N\}$ 是 $u$ 的未知函數； $\{\theta_i, i = 1, 2, \dots, N\}$ 為須被估計之未知參數； $\varepsilon$ 為模型誤差。N組數據所構成之線性方程式集合定義如下：

$$\begin{aligned} y_1 &= \theta_1 f_1(u_1) + \theta_2 f_2(u_1) + \dots + \theta_n f_n(u_1) + \varepsilon_1 \\ y_2 &= \theta_1 f_1(u_2) + \theta_2 f_2(u_2) + \dots + \theta_n f_n(u_2) + \varepsilon_2 \\ &\vdots \\ y_N &= \theta_1 f_1(u_N) + \theta_2 f_2(u_N) + \dots + \theta_n f_n(u_N) + \varepsilon_N \end{aligned} \quad (9)$$

將其以矩陣形式表示如下：

$$Y = A\theta + E \quad (10)$$

其中 $Y$ 為模型輸出； $A$ 為輸入矩陣； $\theta$ 為需被估計之參數向量； $E$ 為誤差；分別如下所示。

$$A = \begin{bmatrix} f_1(u_1) & f_2(u_1) & \dots & f_n(u_1) \\ f_1(u_2) & f_2(u_2) & \dots & f_n(u_2) \\ \vdots & \vdots & \vdots & \vdots \\ f_1(u_N) & f_2(u_N) & \dots & f_n(u_N) \end{bmatrix} \quad (11)$$

$$\theta = [\theta_1 \quad \theta_2 \quad \dots \quad \theta_n]^T \quad (12)$$

$$Y = [y_1 \quad y_2 \quad \dots \quad y_N]^T \quad (13)$$

$$E = [\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_N]^T \quad (14)$$

使用LSE時，必須一次輸入一個資料集才能估計出一組參數；RLSE可以透過遞迴性、循序性的方式進行，僅需要輸入一筆資料即可獲得一組參數，因此能夠根據每一次的輸入進行參數的調整。與LSE相比，RLSE更能夠精確地估計未知的參數，並降低使用LSE所帶來的龐大計算量。對單一目標使用RLSE時，參數計算方式如下所示：

$$P_{k+1} = P_k - \frac{P_k a_{k+1} a_{k+1}^T P_k}{1 + a_{k+1}^T P_k a_{k+1}} \quad (15)$$

$$\theta_{k+1} = \theta_k + P_{k+1}a_{k+1}(y_{k+1} - a_{k+1}^T\theta_k) \quad (16)$$

其中， $k$ 為迭代數， $k = 0, 1, 2, \dots, N-1$ ； $P_k$ 為增益矩陣； $\theta_k$ 代表第 $k$ 次迭代時的估計參數； $\theta_n$ 為最佳參數解；(10)中第 $k$ 組 $[A, Y]$ 可標示為 $[a_k^T, y_k]$ 。在開始遞迴式估計前， $\theta_0$ 和 $P_0$ 需被初始化， $\theta_0$ 設定為零向量，而 $P_0 = \alpha I$ ，其中 $\alpha$ 是一極大整數， $I$ 為單位矩陣，其大小為估計參數數量之大小。

### 第三章 系統設計與架構

#### 3.1 多目標特徵挑選

本篇論文的多目標特徵挑選是使用三種特徵挑選類別中的過濾法，基於夏農資訊熵(Shannon entropy)[21]來計算目標資料集的資料亂度，再透過加入特徵後目標資料集的亂度變化以計算出特徵提供給目標資料集的資訊量，並以特徵提供給目標資料集的資訊量多寡來挑選特徵。

首先，我們需要對時間序列資料集進行差分運算，目的是要了解資料集的漲跌狀況，之後將這些差分資料處理成一個資料矩陣，如以下所示：

$$\begin{matrix} & f_1 & f_2 & \dots & f_n & t_1 & t_2 & \dots & t_N \\ \begin{bmatrix} x_{f_1,1} & x_{f_2,1} & \dots & x_{f_n,1} & y_{t_1,1} & y_{t_2,1} & \dots & y_{t_N,1} \\ x_{f_1,2} & x_{f_2,2} & \dots & x_{f_n,2} & y_{t_1,2} & y_{t_2,2} & \dots & y_{t_N,2} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ x_{f_1,s} & x_{f_2,s} & \dots & x_{f_n,s} & y_{t_1,s} & y_{t_2,s} & \dots & y_{t_N,s} \end{bmatrix} \end{matrix} \quad (17)$$

其中， $f_i$ 代表第 $i$ 個候選特徵變數， $i = 1, 2, \dots, n$ ； $x_{f_i,j}$ 代表 $f_i$ 的第 $j$ 筆資料， $j = 1, 2, \dots, s$ ； $t_i$ 代表第 $i$ 個目標變數， $i = 1, 2, \dots, N$ ； $y_{t_i,j}$ 代表 $t_i$ 的第 $j$ 筆資料， $j = 1, 2, \dots, s$ 。

夏農所提出的熵理論原本是運用於熱力學中表示無序的程度，但在資訊理論裡，熵則是用來表示資料的混亂程度，夏農將隨機事件的熵定義如下：

$$H(Y) = \int p_d(y) \log\left(\frac{1}{p_d(y)}\right) dy \quad (18)$$

其中， $H(Y)$ 代表隨機變數 $Y$ 之資訊熵； $p_d(y)$ 為事件 $y$ 發生之機率；而條件資訊熵的定義如下：

$$H(Y|X) = \iint p_d(y|x)p_d(x) \log \frac{1}{p_d(y|x)} dydx \quad (19)$$

其中， $H(Y|X)$ 為已知隨機變數 $X$ 情況下，隨機變數 $Y$ 之資訊熵； $p_d(y|x)$ 為已知事件 $x$ 發生的條件下事件 $y$ 發生的機率； $p_d(x)$ 為事件 $x$ 發生的機率。互資訊(Mutual information)[21]是透過計算隨機變數 $Y$ 在知道某隨機變數 $X$ 後，它的資訊混亂程度下降了多少，而下降的量就是隨機變數 $X$ 提供給隨機變數 $Y$ 的資訊量。因此互資訊可簡單的以原本隨機變數 $Y$ 之資訊熵減掉隨機變數 $Y$ 在已知隨機變數 $X$ 後的資訊熵計算出來，其公式如下：

$$I(X, Y) = H(Y) - H(Y|X) \quad (20)$$

本研究後續再計算互資訊時，另外分成兩種條件來計算，分別是在隨機變數 $X$ 為正情況下與隨機變數 $Y$ 產生之互資訊，及在隨機變數 $X$ 為負情況下與隨機變數 $Y$ 產生之互資訊。如下所示：

$$I(X^+, Y) = H(Y) - H(Y|X^+) \quad (21)$$

$$I(X^-, Y) = H(Y) - H(Y|X^-)$$

以這兩種情況下所計算出來的兩種互資訊，分別乘上相對應的 $x$ 機率密度分佈期望值再加總即為最後的影響資訊量(Influence Information)[43]。因本篇研究所使用的模型能夠同時預測到多個目標，所以在特徵挑選的策略上會針對多個目標進行特徵挑選[43]。最後就將挑選出的特徵變數與目標變數整理成資料對即可作為模型的訓練資料集(Training data, TD)，表示如下：

$$TD = \{(\vec{h}, \vec{y})^{(i)}, i = 1, 2, \dots, |TD|\} \quad (22)$$

其中， $\vec{h}$ 為 $FP$ 中特徵變數所對應的輸入向量； $\vec{y}$ 為目標向量； $|TD|$ 代表訓練資料集之大小。特別注意到本研究之目標向量是以複數型態來表示。也就是說，當模型要進行兩個

實數目標預測時，本研究會將第一個目標放在實部的位置，第二個目標放在虛部的位置，因此兩個實數目標就會得到一個複數型態的目標，以此類推。

### 3.2 結構學習

以往研究中，許多模型內部的節點及規則數量都是由使用者自行決定。若節點或規則數量太少，會使模型的資料容量變得太小，進而影響模型輸出的結果；若節點及規則數量太多，則會造成模型計算效能上的負擔。因此，決定節點及規則的數量是非常不容易的。結構學習的目的是想透過資料本身來建立研究模型之初始設置，包括模糊集合的初始參數及模糊規則的數量。

分群演算法是自我建構概念的一種方法，較常被使用的分群演算法有K-Means分群法、模糊C-Means分群法(Fuzzy C-Means Clustering, FCM)、減法聚類分群法(Subtractive Clustering, SC)。K-Means演算法[22]是依據資料點彼此之間的距離來進行分群的，與群中心越接近的資料點，越容易被分至同一群。FCM[23][24]是基於模糊邏輯的分群演算法，它依照每個資料點隸屬於某群的程度做為分群的依據。SC[25]於1994年由Chiu所提出，基本概念是將每個資料點視為一潛在的分群中心，並把附近資料點的密度當作該資料點是否可成為群中心的參考標準。選擇擁有最高密度的資料點成為群中心，之後降低該群中心周圍的點成為新的群中心的機率，並持續上述步驟直到終止條件被滿足。

在結構學習中，本研究使用SC對模型每一維度上的輸入進行分群，因此可得到每一維度輸入的分群數量。接著以分群結果在相對應維度的輸入上建立模糊集合，並根據模糊規則建立出格狀的輸入空間(Input space)。本研究將所有的訓練資料輸入到Input space上，每筆資料即會在每個方格上計算出一個啟動強度。將每個方格上所計算出的啟動強度加總，目的是想了解方格上資料點的分佈情況。計算每個方格加總的啟動強度平均值及標準差，最後以平均值加上一倍的標準差作為挑選前鑑部之門檻值，若方格加總的啟動強度大於門檻值，此方格即成為模型的前鑑部。使用門檻值的原因是為了限制參數的數量，以免需要學習的參數數量太多，造成模型計算複雜度升高以及機器學習找尋參數最佳解變得不容易。本研究使用SC進行模糊類神經系統的前鑑部的結構學習，自動決定分群群數，讓分群結果能更貼近真實資料的分布情形。

### 3.3 多目標複數模糊類神經模型

模糊類神經模型具有優秀的學習能力，甚至對未知的輸入也能得到正確的輸出，是一種適應性良好的模型，因此經常使用於預測時間序列中。本研究以複數模糊類神經模型為基礎，結合誤差回饋機制以形成一改良的複數模糊類神經模型。模型大致可以分成兩個區塊，分別是前鑑部與後鑑部，前鑑部會將模型的輸入透過複數模糊集合轉換成歸屬程度，並利用歸屬程度計算出啟動強度(Firing strength)；後鑑部則是將啟動強度與每條規則進行計算得到模型最終輸出，且後鑑部的型態是使用 T-S 型態的模糊推論系統。模型的模糊規則表達如下：

Rule  $i$ :

$$\begin{aligned} \text{If } x_1 \text{ is } A_1^{(i)}(h_1) \text{ and } x_2 \text{ is } A_2^{(i)}(h_2) \text{ and } \dots \text{ and } x_M \text{ is } A_M^{(i)}(h_M) \\ \text{Then } z^{(i)} = a_0^{(i)} + \sum_{j=1}^M a_j^{(i)} h_j \end{aligned} \quad (23)$$

其中， $i$ 代表是第 $i$ 條模糊規則， $i = 1, 2, \dots, Q$ ； $x_M$ 是第 $M$ 個輸入的語意變數； $h_M$ 是第 $M$ 個輸入的數值變數； $A_M^{(i)}(h_M)$ 代表複數模糊集合中第 $i$ 條規則的第 $M$ 個球型模糊集合； $z^{(i)}$ 則是第 $i$ 條規則的輸出； $a_j^{(i)}$ 是第 $i$ 條規則的第 $j$ 個後鑑部參數。

本研究根據訓練集訓練出來的模型預測輸出與目標之間產生的誤差進行分析。在此以第一個實驗的模型預測SSECI股價所產生之誤差為例，如圖2所示。

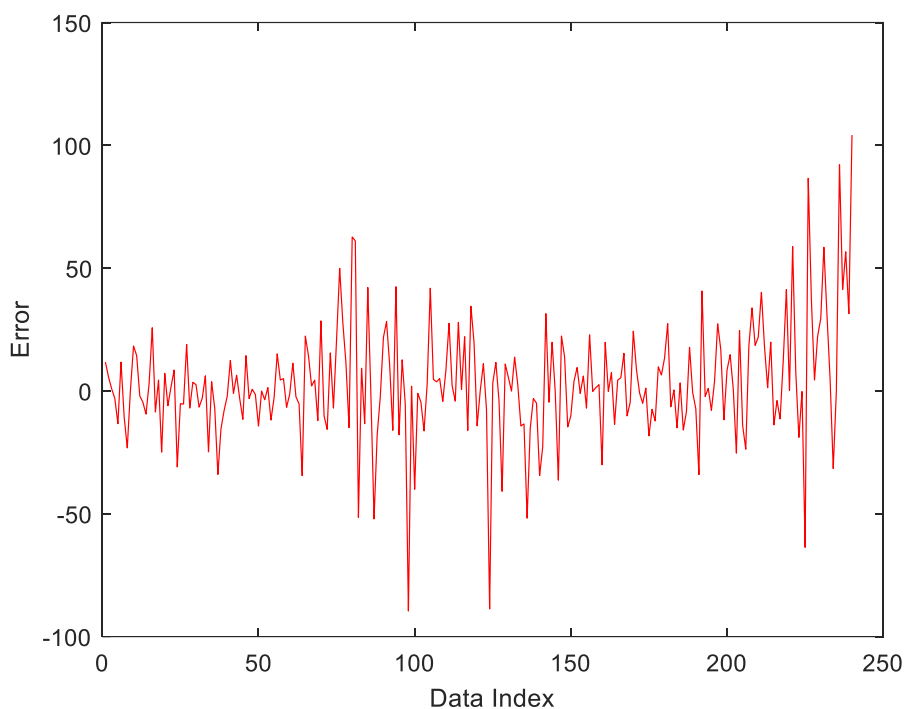


圖 2 模型預測與目標之誤差圖

本研究藉由分析這些誤差以降低誤差之產生，而這個過程即為誤差分析。誤差分析是改進模型預測輸出一個非常重要的方法，因此本研究進一步地將這些誤差繪製成誤差機率密度分布圖，如圖3所示。

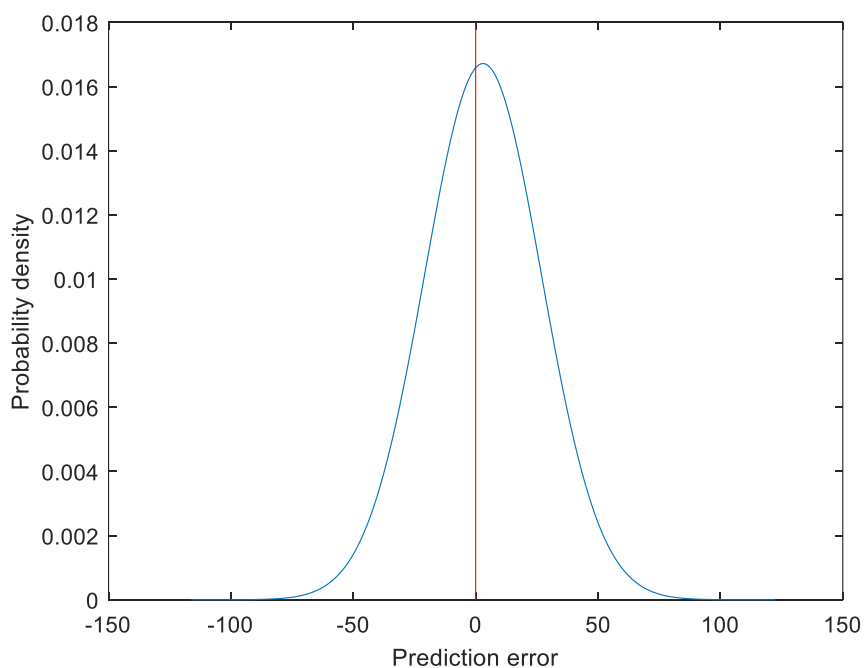


圖 3 誤差機率密度分布圖



本研究發現誤差之機率密度分布為常態分佈，而誤差散布在-150到150之間，並主要集中在-50到50之間。另外本研究發現此分布圖之中心值為3.0002並非為零，因此可得知模型長期預測下來的輸出其實是低估於現實世界中的目標值(此例為低估，但其他案例可能為高估)。因此本研究根據此一特性提出一個誤差回饋的機制，校正本研究建立的模型。以下詳細說明此模型之架構，而誤差回饋方法會在後鑑部層中詳細介紹。

*Layer 1:* 此層稱為輸入層，依照多目標特徵挑選策略選出的特徵變數，將相對應的數值傳遞給下一層，而輸入向量 $\vec{h}$ 表示如下：

$$\vec{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_M \end{bmatrix}_{M \times 1} \quad (24)$$

*Layer 2:* 此層為模糊集層，在此階層中每個輸入皆會輸入進相對應的複數模糊集合中。在本研究中使用的模糊集合是複數型態的高斯模糊集合，其歸屬函數如下所示：

$$\mu_c(h, m, \sigma, \lambda) = r(h, m, \sigma) \exp(j\omega(h, m, \sigma, \lambda)) \quad (25)$$

其中

$$r(h, m, \sigma) = \exp \left[ -\frac{(h - m)^2}{2\sigma^2} \right] \quad (26)$$

$$\omega(h, m, \sigma, \lambda) = -\exp \left[ -\frac{(h - m)^2}{2\sigma^2} \right] \left( \frac{h - m}{\sigma^2} \right) \lambda \quad (27)$$

其中， $\mu_c$ 為複數歸屬程度； $j = \sqrt{-1}$ ； $h$ 是數值變數(Base variable or numerical variable)； $m$ 為中心參數； $\sigma$ 為寬度參數； $\lambda$ 為是相位參數； $r$ 為函數振幅； $\omega$ 為相位函數，為 $r$ 的一階導函數乘上相位參數 $\lambda$ 。複數模糊集合是模型能夠實做出多目標的關鍵，複數的歸屬程度能夠另外解構成3個複數值以形成一個向量如(28)。將實部與虛部同時各掛載一個

目標，本研究之模型至多就能達到 6 個目標的預測。

$$\overrightarrow{\mu_c} = [\mu_c \text{ real}(\mu_c) \text{ imag}(\mu_c)] \quad (28)$$

$\text{real}()$ 表示取實數值; $\text{imag}()$ 表示取虛部值。在本層中所有模糊集合的中心參數、寬度參數與相位參數，皆稱為模型的前鑑部參數。

*Layer 3*: 此層為啟動強度層，針對模糊規則之前鑑部進行歸屬度之邏輯運算，例如AND、OR、NOT或PRODUCT等。本研究使用PRODUCT運算得出每條規則之啟動強度，以第*i*條規則為例，其啟動強度表示如下：

$$\vec{\beta}^{(i)} = \begin{bmatrix} \beta_1^{(i)} \\ \beta_2^{(i)} \\ \vdots \\ \beta_3^{(i)} \end{bmatrix} \quad (29)$$

$$\beta_c^{(i)} = \prod_{j=1}^M \mu_{c_j}^{(i)} \quad (30)$$

其中， $\mu_{c_j}^{(i)}$ 為第*i*個前鑑部的第*j*個複數模糊集合中相對應的第*c*個歸屬程度； $\beta_c^{(i)}$ 為第*i*個前鑑部啟動強度向量中的第*c*個啟動強度； $\vec{\beta}^{(i)}$ 為第*i*個前鑑部的啟動強度向量。

*Layer 4*: 此層為正規化層，將每條模糊規則之啟動強度進行正規化的計算，以便了解每條模糊規則對於總系統輸出之貢獻程度。以第*i*條規則為例，其啟動強度表示如下：

$$\vec{\lambda}^{(i)} = \frac{\vec{\beta}^{(i)}}{\sum_{i=1}^Q \vec{\beta}^{(i)}} \quad (31)$$

其中 $\vec{\lambda}^{(i)}$ 是第*i*條規則正規化啟動強度向量。

*Layer 5*: 此層為後鑑部層，根據第四層產生之正規化啟動強度來計算每條模糊規則的後鑑部輸出。本研究在後鑑部層中，提出一個新穎的誤差回饋方式來嘗試改善模型輸出之

結果，使模型輸出能更加貼近現實世界中的數值。誤差回饋方法是將誤差輸入回模型的後鑑部，讓誤差進入到後續參數學習的部分。因為將誤差輸入至後鑑部，因此須根據模型之目標個數來增加後鑑部的參數個數，並對公式(23)進行如下改良。

Rule  $i$ :

$$\begin{aligned} \text{If } x_1 \text{ is } A_1^{(i)}(h_1) \text{ and } x_2 \text{ is } A_2^{(i)}(h_2) \text{ and } \dots \text{ and } x_M \text{ is } A_M^{(i)}(h_M) \\ \text{Then } z^{(i)} = a_0^{(i)} + \sum_{j=1}^{M^*} a_j^{*(i)} h_j^* \end{aligned} \quad (32)$$

當沒有進行誤差回饋機制時， $M^* = M$ ，因此 $h_j^* = h_j$ ，也就是說後鑑部的輸入向量大小並沒有改變，因此後鑑部的參數數量也不會改變；當進行誤差回饋機制時， $M^* = M + |TS|$ ，也就是說當 $1 \leq j \leq M$ 時， $h_j^* = h_j$ ，但當 $M+1 \leq j \leq M+|TS|$ 時， $h_j^* = e_j$ 。因本研究之誤差回饋機制是將每一目標與模型輸出產生之誤差輸入進後鑑部，因此後鑑部的輸入向量會根據目標的個數增加，相對應的後鑑部參數數量也會因此增加。本研究透過此方法重新訓練出一組參數來使得模型輸出能更加貼近目標。而後鑑部之輸出以第 $i$ 條規則為例，表示如下：

$$\vec{\hat{y}}^{(i)} = \vec{\lambda}^{(i)} z^{(i)} \quad (33)$$

其中， $\vec{\hat{y}}^{(i)}$ 為第 $i$ 條規則的輸出； $z^{(i)}$ 為第 $i$ 條規則的後鑑部。

Layer 6: 此層為輸出層，將第五層產生之輸出加總計算，以形成模型之預測輸出。如下所示：

$$\vec{\hat{y}} = \sum_{i=1}^Q \vec{\hat{y}}^{(i)} \quad (34)$$

其中 $\vec{\hat{y}}$ 為模型之總預測輸出。

假設有 $i$ 組歷史觀測值，經由模型計算可得 $i$ 組預測值，進而得到 $i$ 組預測誤差，表示如下：

$$\vec{e}^{(i)} = \vec{y}^{(i)} - \vec{\hat{y}}^{(i)} \quad (35)$$

其中， $i = 1, 2, \dots, |TD|$ ； $\vec{y}^{(i)}$ 為第 $i$ 組模型之目標； $\hat{\vec{y}}^{(i)}$ 為第 $i$ 組模型之預測輸出， $\vec{e}^{(i)}$ 是第 $i$ 組模型目標與輸出之誤差。此 $|TD|$ 組預測誤差可用來計算出性能指標。性能指標之功能為判斷模型輸出的性能，並讓機器學習演算法依據這個性能指標調整模型參數。本研究使用均方根誤差(Root mean square error, RMSE)做為計算的效能指標，另外因本研究提出之模型能夠同時預測多個目標，因此計算出來的誤差將會是一個複數值，所以在RMSE(36)中的平方計算必須使用共軛轉置來相乘，RMSE之定義如下：

$$\text{RMSE} = \sqrt{\frac{1}{|TD|} \sum_{i=1}^{|TD|} \vec{e}^{(i)*} \vec{e}^{(i)}} \quad (36)$$

### 3.4 複合式 PSO-RLSE 演算法

在本篇論文中所使用到的機器學習法是將粒子群最佳化演算法結合遞迴最小平方估計法來合成複合式的演算法，藉由分而擊之(Divide and conquer)的概念，將粒子群最佳化演算法用於最佳化前鑑部的參數，而遞迴最小平方估計法則是用於後鑑部參數的最佳化。此複合學習方式結合粒子群最佳化演算法群體智慧演化的搜尋方法，以及遞迴式最小平方估計法線性模型最佳化的效率，可以達到快速收斂的目的，並具有良好的估計效果。詳細參數學習方法如以下步驟所示。

步驟一 粒子群最佳化演算法粒子數、粒子位置和最大迭代數初始化。在模型中，運用SC分群後所產生之最佳分群中心與標準差作為第一顆粒子之初始位置，其餘皆使用初始位置加上亂數隨機產生。

步驟二 計算每個前鑑部之啟動強度，並將啟動強度正規化。

步驟三 利用遞迴式最小平方估計法更新模型之後鑑部參數，(15)、(16)中的 $a$ 定義如下。

$$a_{k+1} = [aa^1(k+1) \ aa^2(k+1) \ \dots \ aa^Q(k+1)] \quad (37)$$

$$aa^i(k+1) = \begin{bmatrix} \lambda^{1,i} & h_1(k+1)\lambda^{1,i} & \dots & h_M(k+1)\lambda^{1,i} \\ & \vdots & & \\ \lambda^{N,i} & h_1(k+1)\lambda^{N,i} & \dots & h_M(k+1)\lambda^{N,i} \end{bmatrix} \quad (38)$$

因為本研究是在多目標的情況下進行預測，因此需將 RLSE 參數計算方式修改成在多目標的情況下仍可進行的型式，如下所示：

$$P_{k+1} = P_k - \frac{P_k a_{k+1}}{I + a_{k+1}^T P_k a_{k+1}} a_{k+1}^T P_k \quad (39)$$

$$\theta_{k+1} = \theta_k + P_{k+1} a_{k+1} (y_{k+1} - a_{k+1}^T \theta_k) \quad (40)$$

其中  $\theta_k$  為第  $k$  次迭代的後鑑部參數； $y_{k+1}$  為模型目標； $P_k$  為  $\theta_k$  的增益矩陣；在公式(39)中的  $I$  為單位矩陣，但是大小則是根據目標之數量決定。

步驟四 系統參數由 PSO-RLSE 決定後，得出模型預測值與誤差。

步驟五 重複步驟二到步驟四，並依照 RMSE 公式計算出每個粒子的成本函數值。

步驟六 更新 PSO 的  $Pbest$  及  $Gbest$ ，並以  $Gbest$  更新前鑑部參數。

步驟七 若滿足終止條件，參數學習流程即結束，否則回到步驟二繼續參數學習流程。終止條件為迭代的次數，當迭代次數用完時即完成學習，迭代次數則由使用者自行設定。

## 第四章 實驗實作與結果

### 4.1 實驗 1: 單目標股票指數預測-SSECI(2006)

本次實驗的目的是為了測試模型以單一目標的方式來進行訓練與預測的效能。實驗一將從 YAHOO 財經資訊網站收集上海證券交易所綜合股價指數(SSECI)一段時間的收盤價進行單一目標的模型預測訓練。上海證券交易所綜合股價指數是反映上海證券交易所掛牌股票總體走勢的統計指標，該指數反應了香港的經濟走勢，也可視為亞洲地區重要的指數之一。

本實驗從 2006 年 1 月 1 日起至 2006 年 12 月 31 日止，搜集 271 筆每日上海證券交易所綜合股價指數收盤資料。在資料前處理時，會將此指數之收盤資料處理成資料的漲跌，並以資料集的漲跌進行訓練與預測，表示為 $\{y_1(t), t = 1, 2, \dots, 270\}$ ， $t$ 為時間指標。接著會挑出該指數資料集的前 30 天做為候選特徵，因此共有 30 個特徵。此 30 個特徵為上海證券交易所綜合股價指數的候選特徵變數，表示為 $\{f_i, i = 1, 2, \dots, 30\}$ ， $i$ 值越大表示特徵變數越靠近目標。之後對此指數進行資訊影響矩陣的運算，得出上海證券交易所綜合股價指數之 $IIM_{31 \times 31}$ ，再利用特徵選取策略挑出特徵 $f_7$ 、 $f_{14}$ 、 $f_{23}$ 、 $f_{26}$ 。對上述資料整理形成特徵值與目標值之資料矩陣，表示為 $\{(\vec{f}(k), \vec{t}(k)), k = 1, 2, \dots, 240\}$ ，其中， $\vec{f}(k)$ 是模型的特徵輸入向量， $\vec{t}(k)$ 是目標向量。以一月至十月的資料作為訓練資料，十一及十二月的資料做為測試資料，並重複進行十次的實驗，最後將實驗結果與其他文獻模型進行比較。

在模型中，每個模型輸入的高斯複數模糊集合數量是使用減法分群法(Subtractive clustering)來對模型輸入進行分群。在本實驗中每個模型輸入都是分為三群，因此總共會有 81 個待挑選的前鑑部，並利用門檻值的篩選以及上下界的設立(最少要有 4 個前鑑部，最多不超過 15 個前鑑部)，最終得到了 9 個前鑑部，詳細的模型參數配置如表 1。

表 1 模型參數配置(實驗一)

Features selected as inputs to the model	$\{f_7, f_{14}, f_{23}, f_{26}\}$
Number of Gaussian complex fuzzy sets (GCFSs) of each input	3、3、3、3
Number of premise parts	9
Number of premise parameters	36
Number of consequent parts	9
Number of consequent parameters	54
Number of model outputs	1

在本實驗中使用的機器學習演算法為粒子群最佳化演算法與遞迴最小平方估計法合成的複合式演算法，在此次實驗總共設定了 15 顆粒子，粒子的初始位置是由 SC 對訓練資料進行分群所得出的參數，因次每顆粒子都有 36 個維度，各個維度代表著模型前鑑部的各個參數，迭代次數設定為 50 次迭代，演算法配置如表 2。

表 2 演算法參數配置(實驗一)

<b>PSO</b>	
Number of particle dimensions	36
Number of particles	15
$(\omega, c_1, c_2)$	(0.85, 2, 2)
$\xi_1, \xi_2$	random[0,1]
Number of learning iterations	50
<b>RLSE</b>	
$\theta_0$	54×1 Zero vector
$P_0$	$\alpha \cdot I$
$\alpha$	$10^7$
$I$	54×54 Identity matrix

本研究將模型重複跑 10 次以達到統計上的公平性，如表 3 所示，並在之後的各圖表中列出十次中最佳的一次結果。效能與其他文獻比較如表 4。圖 4 為模型預測輸出比較圖，圖 5 為模型預測輸出誤差圖，圖 6 為模型學習曲線圖。

表 3 10 次實驗之結果(實驗一)

#	總 RMSE	SSECI	
	訓練階段	訓練階段	測試階段
1	19.89358	19.81749	36.56889
2	20.37994	20.27536	37.09097
3	20.11064	19.98181	36.83894
4	20.31251	20.00827	37.3884
5	19.78419	19.44353	33.79935
6	20.23911	20.21365	36.15826
7	19.88268	19.81142	37.05421
8	20.26453	20.24769	36.99292
9	20.1671	19.99176	37.19274
10	19.97215	19.81838	34.97742

表 4 與其他文獻之比較(實驗一)

Method	SSECI(RMSE)	
	Training phase	Testing phase
Huarng et al.'s method[26]	-	75.0643
Cheng et al.'s method[27]	-	64.0943
Chen's method[28]	-	42.6612
Lee et al.'s method[29]	-	82.0055
Egrioglu et al.'s method[30]	-	114.9601
Wang et al.'s method[31]	-	379.5415
Bas et al.'s method[32]	-	221.1243
Yolcu et al.'s method[33]	-	226.9612
Zhang et al.'s method[34]	-	36.5687
(Proposed) without feedback	20.1165	37.1975
(Proposed) with feedback	19.4435	33.7993



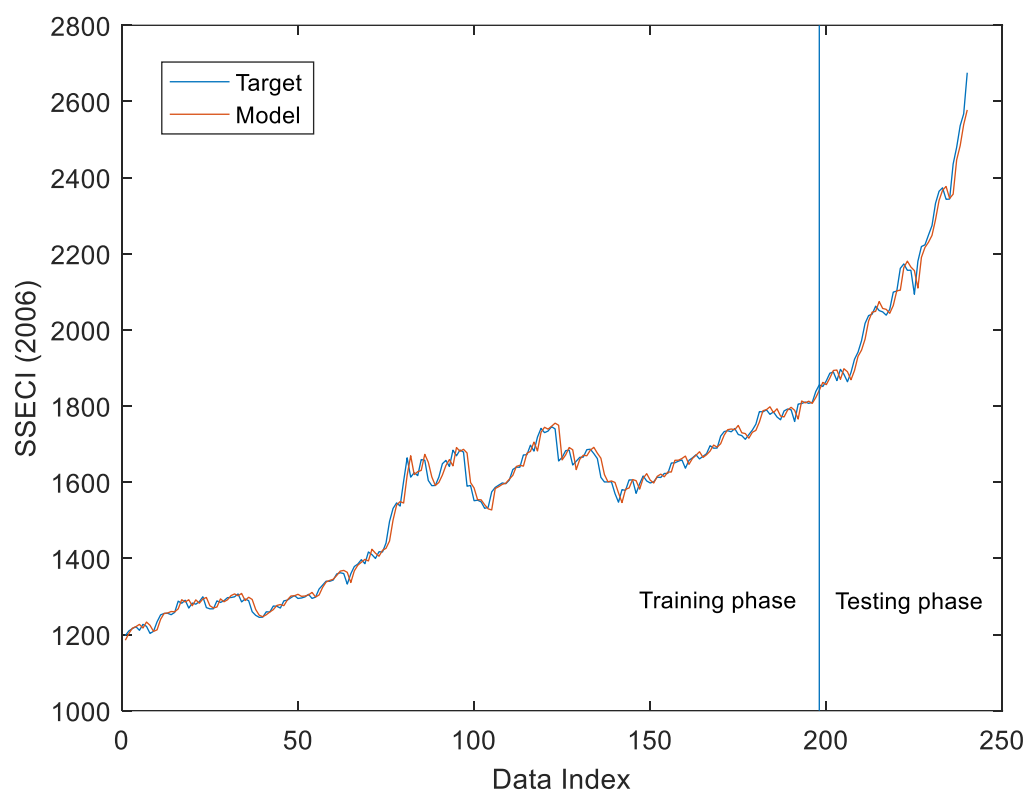


圖 4 模型訓練與預測結果(SSECI)

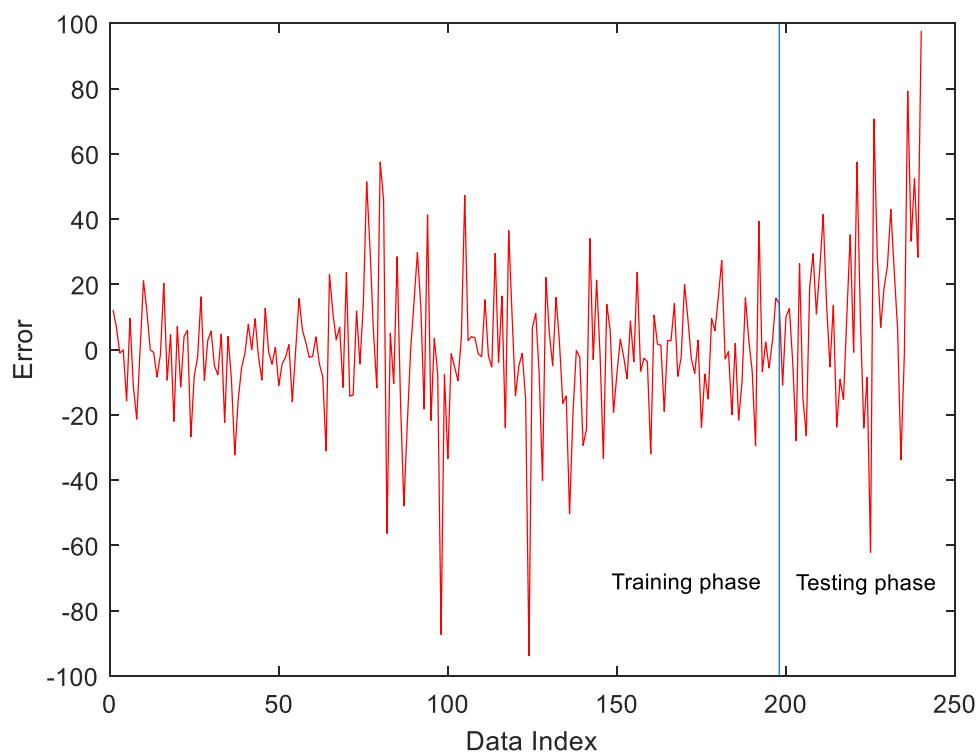


圖 5 訓練結果與目標之誤差(SSECI)

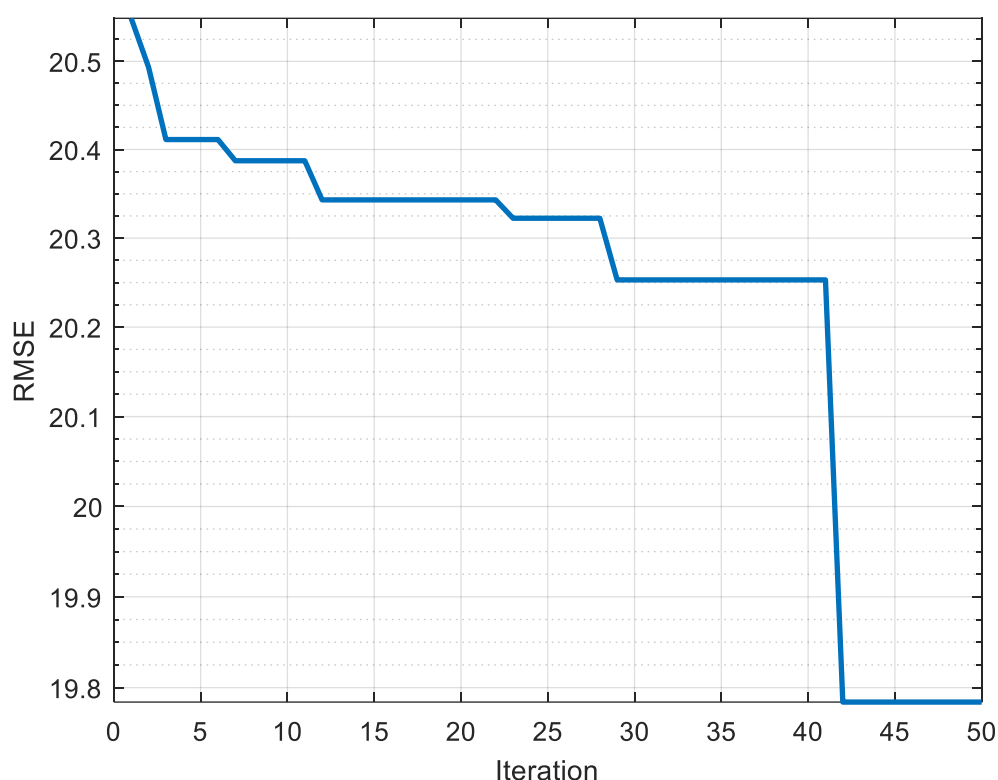


圖 6 模型學習曲線(實驗一)

## 4.2 實驗 2: 雙目標股票指數預測- DJIA(2004) 、TAIEX(2004)

本次實驗的目的是為了測試模型以兩個目標的方式來進行訓練與預測的效能。實驗二將從 YAHOO 財經資訊網站收集道瓊工業指數(DJIA)與台股加權指數(TAIEX)一段時間的收盤價進行兩個目標的模型預測訓練。道瓊工業指數為華爾街日報和道瓊公司所創造的股票市場指數，其為美國 30 間知名的上市公司加權平均出來的指數，是最悠久的美國市場指數之一，因此在美國市場甚至擁有很大的影響力及代表性;台股加權指數為台灣證券交易所所編製的股價指數，該指數反應了整體台灣的經濟走勢，也可視為亞洲地區重要的指數之一。

本實驗從 2004 年 1 月 1 日起至 2004 年 12 月 31 日止，搜集 271 筆每日道瓊工業指數收盤資料以及每日台灣加權指數收盤資料。在資料前處理時，會將兩指數之收盤資料

處理成資料的漲跌，並以資料集的漲跌進行訓練與預測，表示為 $\{y_1(t), y_2(t), t = 1, 2, \dots, 270\}$ ， $t$ 為時間指標。接著會挑出各指數資料集的前 30 天做為候選特徵，因此兩個指數即有 60 個特徵。其中，前 30 個特徵為道瓊工業指數的特徵變數，31 到 60 個特徵為台股加權指數的特徵變數。候選特徵變數表示為 $\{f_i, i = 1, 2, \dots, 60\}$ ， $i$ 值越大表示特徵變數越靠近目標。之後對兩指數進行資訊影響矩陣的運算，得出道瓊工業指數之  $IIM_{61 \times 61}$  以及台灣加權指數之  $IIM_{61 \times 61}$ ，再利用多目標特徵選取策略挑出特徵  $f_{16}$ 、 $f_{40}$ 、 $f_3$ 、 $f_{41}$ 。對上述資料整理形成特徵值與目標值之資料矩陣，表示為 $\{(\vec{f}(k), \vec{t}(k)), k = 1, 2, \dots, 240\}$ ，其中， $\vec{f}(k)$ 是模型的特徵輸入向量， $\vec{t}(k)$ 是目標向量。以一月至十月的資料作為訓練資料，十一及十二月的資料做為測試資料，並重複進行十次的實驗，最後將實驗結果與其他文獻模型進行比較。

在模型中，每個模型輸入的高斯複數模糊集合數量是使用減法分群法(Subtractive clustering)來對模型輸入進行分群。在本實驗中每個模型輸入都是分為三群，因此總共會有 81 個待挑選的前鑑部，並利用門檻值的篩選以及上下界的設立(最少要有 4 個前鑑部，最多不超過 15 個前鑑部)，最終得到了 9 個前鑑部，詳細的模型參數配置如表 5。

表 5 模型參數配置(實驗二)

Features selected as inputs to the model	$\{f_{16}, f_{40}, f_3, f_{41}\}$
Number of Gaussian complex fuzzy sets (GCFSs) of each input	3、3、3、3
Number of premise parts	9
Number of premise parameters	36
Number of consequent parts	9
Number of consequent parameters	63
Number of model outputs	1

在本實驗中使用的機器學習演算法為粒子群最佳化演算法與遞迴最小平方估計法合成的複合式演算法，在此次實驗總共設定了 15 顆粒子，粒子的初始位置是由 SC 對訓練資料進行分群所得出的參數，因此每顆粒子都有 36 個維度，各個維度代表著模型前

鑑部的各個參數，迭代次數設定為 50 次迭代，演算法配置如表 6。

表 6 演算法參數配置(實驗二)

<b>PSO</b>	
Number of particle dimensions	36
Number of particles	15
$(\omega, c_1, c_2)$	(0.85,2,2)
$\xi_1, \xi_2$	random[0,1]
Number of learning iterations	50
<b>RLSE</b>	
$\theta_0$	63×1 Zero vector
$P_0$	$\alpha \cdot I$
$\alpha$	$10^7$
$I$	63×63 Identity matrix

本研究將模型重複跑 10 次以達到統計上的公平性，如表 7 所示，並在之後的各圖表中列出十次中最佳的一次結果。效能與其他文獻比較如表 8 及表 9。圖 7、圖 8 為模型預測輸出與目標的比較圖，圖 9、圖 10 為模型預測輸出與目標間的誤差圖，圖 11 為模型學習曲線圖。

表 7 10 次實驗之結果(實驗二)

#	總 RMSE	DJIA		TAIEX	
	訓練階段	訓練階段	測試階段	訓練階段	測試階段
1	118.7277	70.8099	63.88057	95.3007	56.01075
2	117.862	72.50762	64.37455	92.91989	55.64566
3	118.7923	70.72761	62.03837	95.44224	57.55793
4	118.7864	75.92931	67.84511	91.35071	54.61158
5	117.7235	73.56229	68.21016	91.9098	54.57453
6	118.0785	74.50345	64.06841	91.60663	58.4241
7	116.016	71.78638	56.69803	91.13956	56.79757
8	119.3136	73.33933	65.48269	94.11207	60.0473
9	114.6252	66.65247	60.13028	93.25435	56.64147
10	118.7811	71.964	61.94589	94.49939	56.75504

表 8 與其他文獻之比較(實驗二)

Method	DJIA(RMSE)	
	Training phase	Testing phase
SVR [35](two models, each with single output)	-	71.49
ANFIS [35] (two models, each with single output)	-	68.00
RBF [35] (one model with two outputs)	-	148.11
CNFS-ARIMA [35](one model with two outputs)	-	66.55
(Proposed) without feedback	71.9852	68.3225
(Proposed) with feedback	66.6524	60.1302

表 9 與其他文獻之比較(實驗二)

Method	TAIEX(RMSE)	
	Training phase	Testing phase
Multivariate heuristic fuzzy time-series model [36]	-	73.49
Fuzzy time series model [37-39]	-	84
Fuzzy time series model with FLRGs [40]	-	58.89
SVR [35] (two models, each with single output)	-	58.81
RBF [35](two models, each with single output)	-	102.86
CNFS-ARIMA[35] (one model with two outputs)	-	55.56
(Proposed) without feedback	95.5535	61.7063
(Proposed) with feedback	93.2343	56.6414

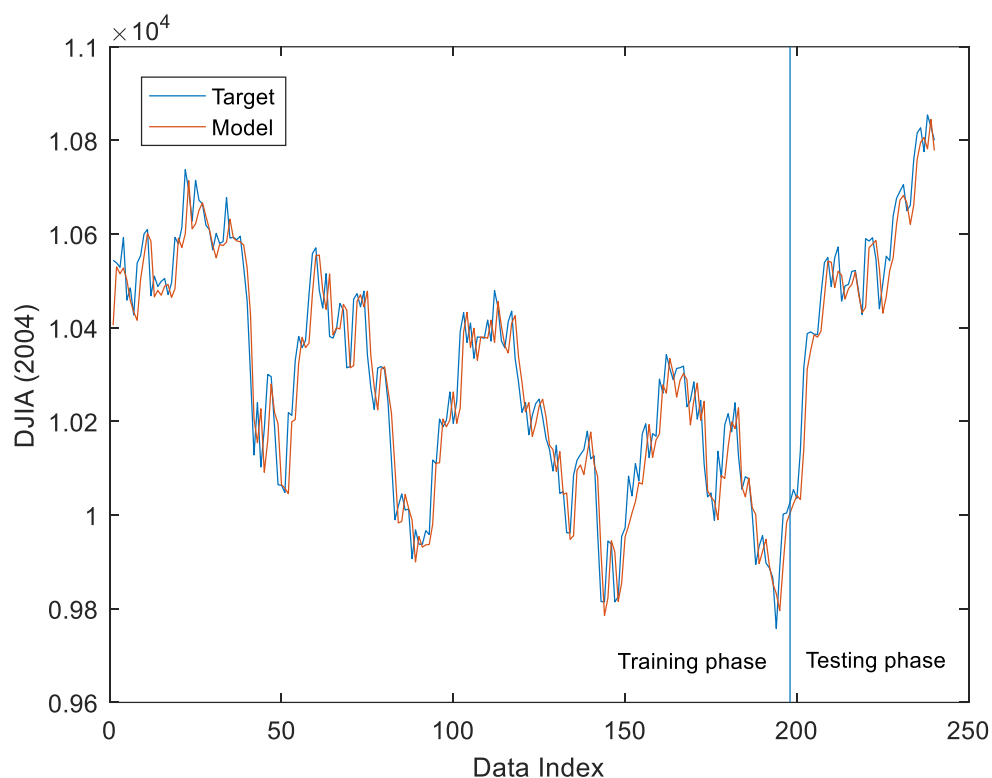


圖 7 模型訓練與預測結果(DJIA)

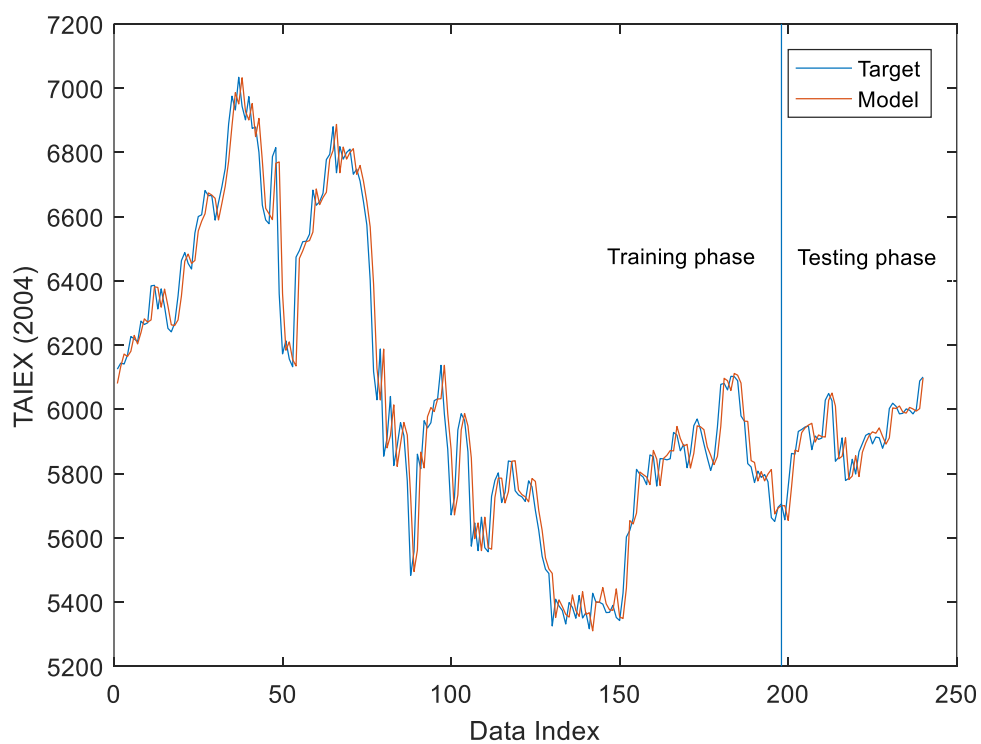


圖 8 模型訓練與預測結果(TAIEX)

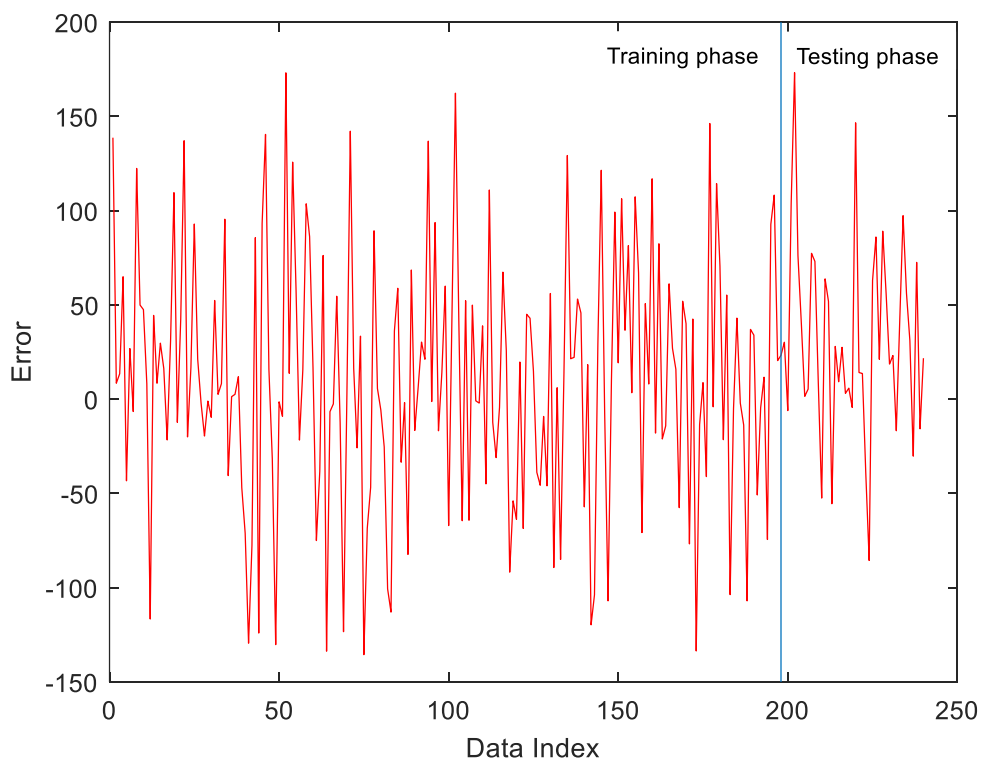


圖 9 訓練結果與目標之誤差(DJIA)

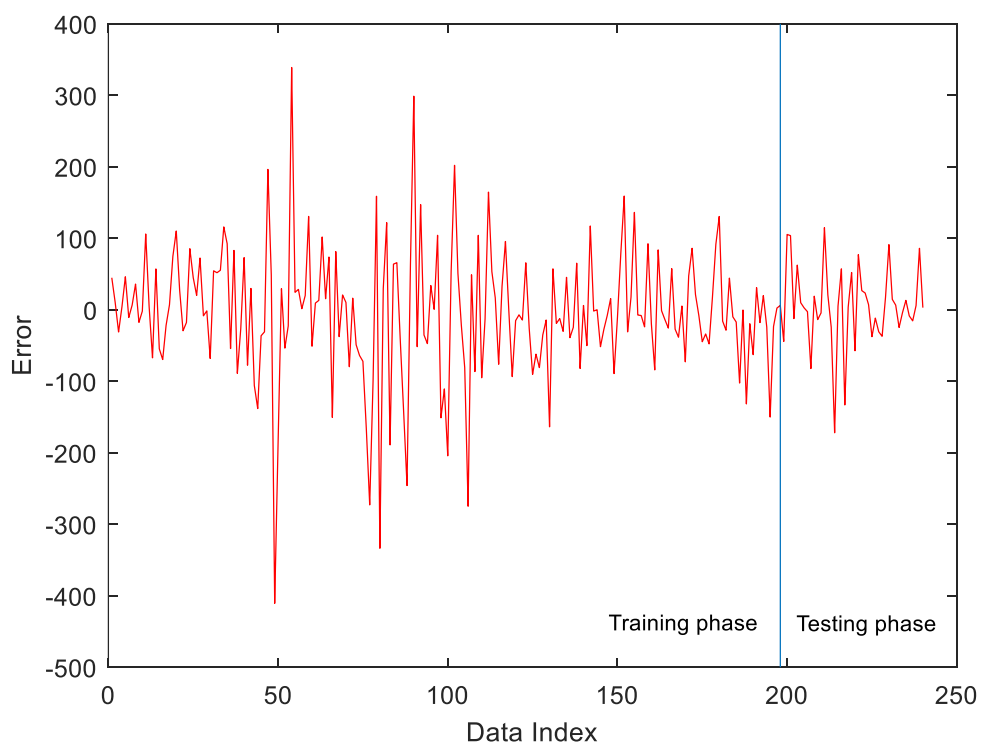


圖 10 訓練結果與目標之誤差(TAIEX)

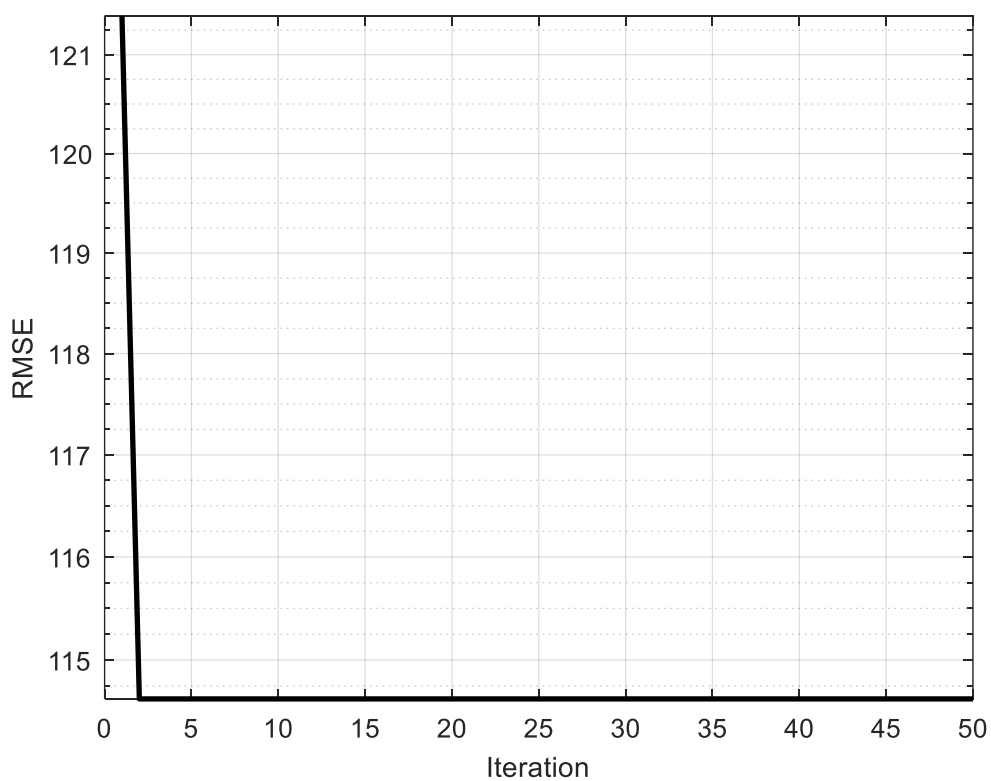


圖 11 模型學習曲線(實驗二)



### 4.3 實驗 3: 四個目標股票指數預測- TAIEX(2006)、DJIA(2006)、Nikkei(2006)、FTSE(2006)

本次實驗是從 YAHOO 財經資訊網站收集台灣加權指數(TAIEX)、道瓊工業指數(DJIA)、日經平均指數(Nikkei)及英國富時指數(FTSE)一段時間的收盤價進行四個目標之預測。台股加權指數為台灣證券交易所所編製的股價指數，該指數反應了整個台灣的經濟走勢，也可視為亞洲地區重要的指數之一；道瓊工業指數為華爾街日報和道瓊所創造的市場指數，在美國市場甚至是全球的市場都有很大的影響力以及代表性；日經指數為日本經濟新聞社所推出的股價指數，代表了 225 品種的股價指數，反映了整個日本地區的經濟狀況；英國富時指數為富時集團所制定的，其中包含了倫敦證券交易所上市的前 100 大公司的股價，該指數也表現了整個英國的經濟狀態。

本實驗從 2006 年 1 月 1 日起至 2006 年 12 月 31 日止，搜集 256 筆每日台灣加權指數收盤資料、每日道瓊工業指數收盤資料、每日日經平均指數收盤資料及每日英國富時指數收盤資料。在資料前處理時，會將此四指數之收盤資料處理成資料的漲跌，並以資料集的漲跌進行訓練與預測，表示為 $\{(y_1(t), y_2(t), y_3(t), y_4(t)), t = 1, 2, \dots, 255\}$ ， $t$ 為時間指標。接著會挑出各指數資料集的前 30 天做為候選特徵，因此四個指數即有 120 個特徵。其中，前 30 個特徵為台灣加權指數的特徵變數，31 到 60 個特徵為道瓊工業指數的特徵變數，61 到 90 個特徵為日經平均指數的特徵變數，91 到 120 個特徵為英國富時指數的特徵變數。候選特徵變數表示為 $\{f_i, i = 1, 2, \dots, 120\}$ ， $i$ 值越大表示特徵變數越靠近目標。之後對此四個指數進行資訊影響矩陣的運算，得出台灣加權指數之 $IIM_{121 \times 121}$ 、道瓊工業指數之 $IIM_{121 \times 121}$ 、日經平均指數之 $IIM_{121 \times 121}$ 、英國富時指數之 $IIM_{121 \times 121}$ ，再利用多目標特徵選取策略挑出特徵 $f_{66}$ 、 $f_{68}$ 、 $f_{63}$ 、 $f_{64}$ 。對上述資料整理形成特徵值與目標值之資料矩陣，表示為 $\{(\vec{f}(k), \vec{t}(k)), k = 1, 2, \dots, 225\}$ ，其中， $\vec{f}(k)$ 是模型的特徵輸入向量， $\vec{t}(k)$ 是目標向量。樣本中以前六個月的資料作為訓練資料，後六個月的資料做為測試資料，並且重複進行十次的實驗，最後實驗結果將與其他文獻模型進行比較。

在模型中，每個模型輸入的高斯複數模糊集合數量是使用減法分群法(Subtractive

clustering)來對模型輸入進行分群。在本實驗中每個模型輸入都是分為三群，因此總共會有 81 個待挑選的前鑑部，並利用門檻值的篩選以及上下界的設立(最少要有 4 個前鑑部，最多不超過 15 個前鑑部)，最終得到了 10 個前鑑部，詳細的模型參數配置如表 10。

表 10 模型參數配置(實驗三)

Features selected as inputs to the model	$\{f_{66}、f_{68}、f_{63}、f_{64}\}$
Number of Gaussian complex fuzzy sets (GCFSs) of each input	3、3、3、3
Number of premise parts	10
Number of premise parameters	36
Number of consequent parts	10
Number of consequent parameters	90
Number of model outputs	2

在本實驗中使用的機器學習演算法為粒子群最佳化演算法與遞迴最小平方估計法合成的複合式演算法，在此次實驗總共設定了 10 顆粒子，粒子的初始位置是由 SC 對訓練資料進行分群所得出的參數，因此每顆粒子都有 36 個維度，各個維度代表著模型前鑑部的各個參數，迭代次數設定為 50 次迭代，演算法配置如表 11。

表 11 演算法參數配置(實驗三)

<b>PSO</b>	
Number of particle dimensions	36
Number of particles	10
$(\omega, c_1, c_2)$	(0.85, 2, 2)
$\xi_1, \xi_2$	random[0,1]
Number of learning iterations	50
<b>RLSE</b>	
$\theta_0$	90×1 Zero vector
$P_0$	$\alpha \cdot I$
$\alpha$	$10^7$
$I$	90×90 Identity matrix

本研究將模型重複跑 10 次以達到統計上的公平性，如表 12 所示，並在之後的各圖表中列出十次中最佳的一次結果。效能與其他文獻比較如表 13。圖 12、圖 13、圖 14、圖 15 為模型預測輸出與目標的比較圖，圖 16、圖 17、圖 18、圖 19 為模型預測輸出與目標間的誤差圖，圖 20 為模型學習曲線圖。

表 12 10 次實驗之統計結果(實驗三)

#	總 RMSE	TAIEX		DJIA		Nikkei		FTSE	
	訓練階段	訓練階段	測試階段	訓練階段	測試階段	訓練階段	測試階段	訓練階段	測試階段
1	357.7505	83.8454	61.48198	62.52778	56.3569	246.6757	159.2491	56.91811	42.78553
2	350.1726	78.85677	61.12006	77.16153	66.75839	233.0038	150.2636	56.87332	41.47662
3	355.9818	81.56714	58.49494	67.09629	61.71022	245.4789	158.4847	49.21652	35.60974
4	357.6192	83.57147	60.41444	61.30106	60.12616	245.9667	159.556	63.27673	44.02481
5	356.6202	74.87223	58.25182	76.88681	70.5114	238.9784	152.5758	70.99492	49.79668
6	362.3978	89.47123	67.56661	71.55962	69.7829	241.7765	165.1841	54.43926	44.9642
7	356.1675	81.50976	60.81115	66.47344	60.18028	243.9687	158.204	58.94648	40.16324
8	349.4076	79.72195	61.27665	65.57607	61.12473	239.595	154.7204	56.561	39.32288
9	362.24	80.42029	60.13431	73.88353	65.70176	246.3424	160.3234	57.80161	40.77505
10	356.8007	85.38438	71.61052	80.27384	76.77685	231.1997	149.1987	62.91414	54.16299

表 13 與其他文獻之比較(實驗三)

Method	TAIEX		DJIA		Nikkei		FTSE	
	<u>Training</u>	<u>Testing</u>	<u>Training</u>	<u>Testing</u>	<u>Training</u>	<u>Testing</u>	<u>Training</u>	<u>Testing</u>
	<u>phase</u>	<u>phase</u>	<u>phase</u>	<u>phase</u>	<u>phase</u>	<u>phase</u>	<u>phase</u>	<u>phase</u>
Huarng et al.'s method[26]	-	82.8194	-	-	-	-	-	-
Cheng et al.'s method[27]	-	99.7014	-	-	-	-	-	-
Chen's method[28]	-	92.4910	-	-	-	-	-	-
Lee et al.'s method[29]	-	84.2257	-	-	-	-	-	-
Egrioglu et al.'s method[30]	-	78.7521	-	-	-	-	-	-
ARIMA[41]	-	-	-	64.43	-	-	-	-
ETS[41]	-	-	-	63.72	-	-	-	-
ABC-RNN[42]	-	116	-	84	-	161	-	41
(Proposed)without feedback	83.5435	62.0253	74.0421	66.0495	246.4806	159.6775	54.48716	39.3211
(Proposed)with feedback	79.7219	61.2766	65.5760	61.1247	239.595	154.7204	56.561	39.3228

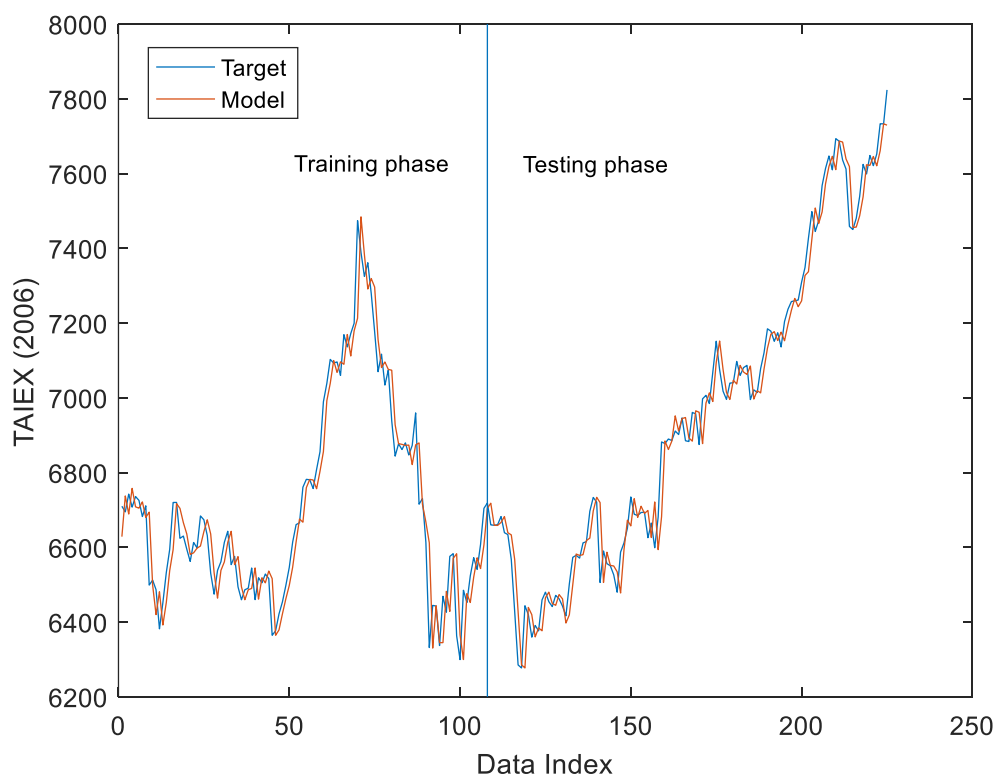


圖 12 模型訓練與預測結果(TAIEX)

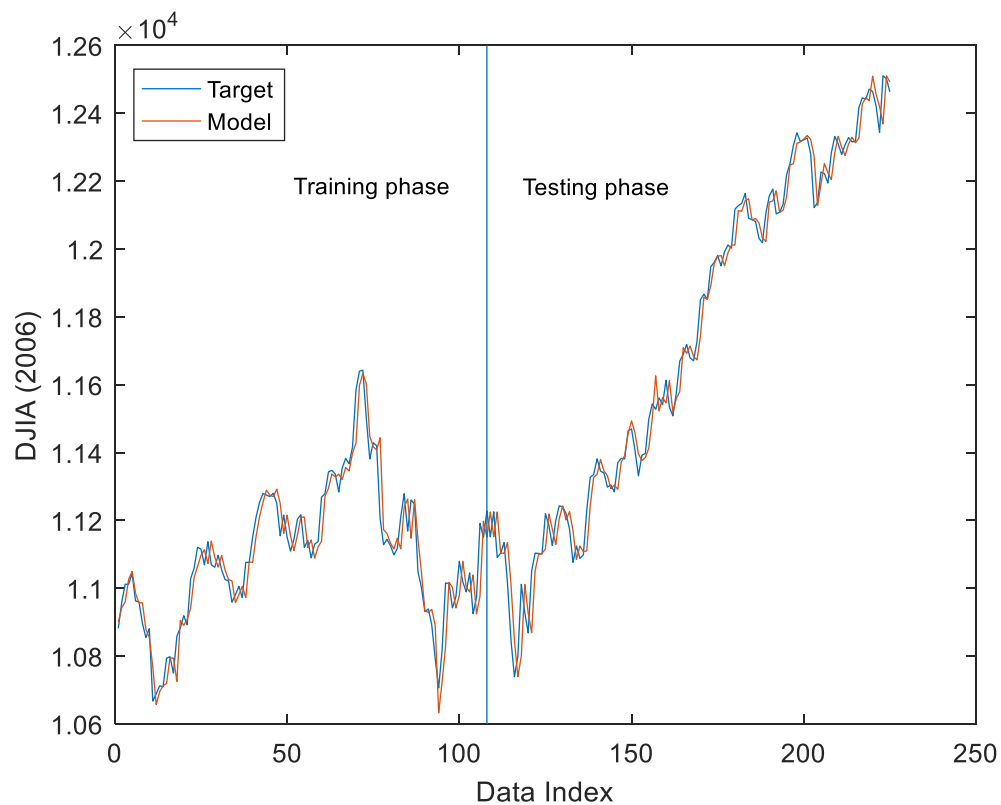


圖 13 模型訓練與預測結果(DJIA)

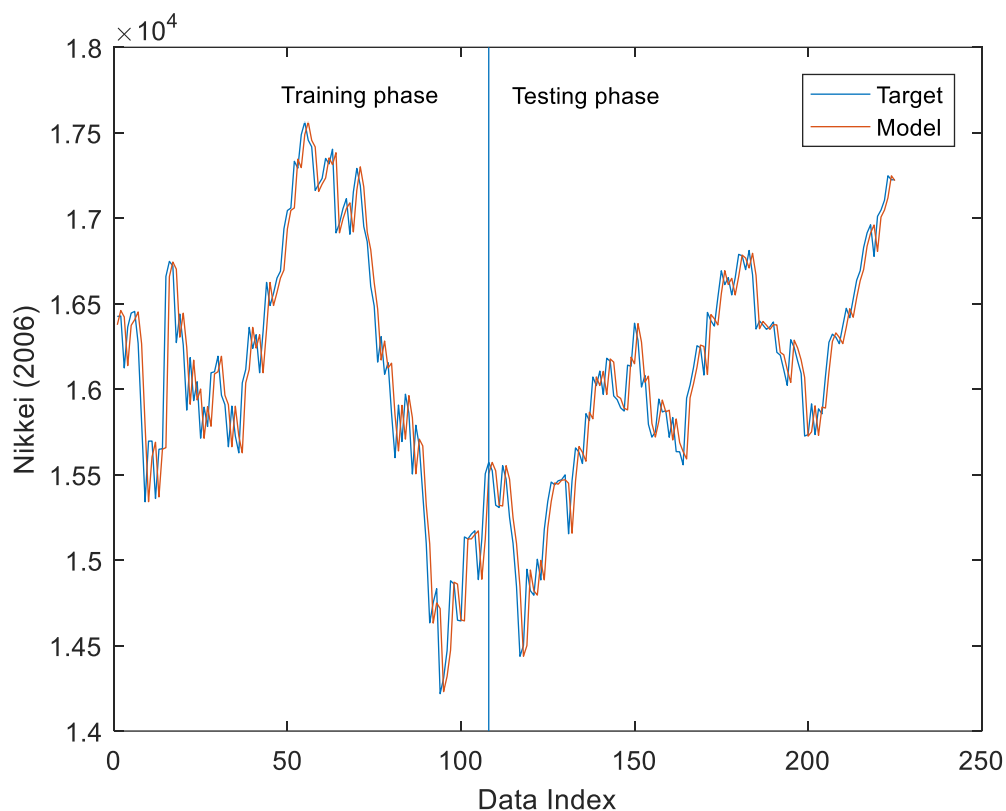


圖 14 模型訓練與預測結果(Nikkei)

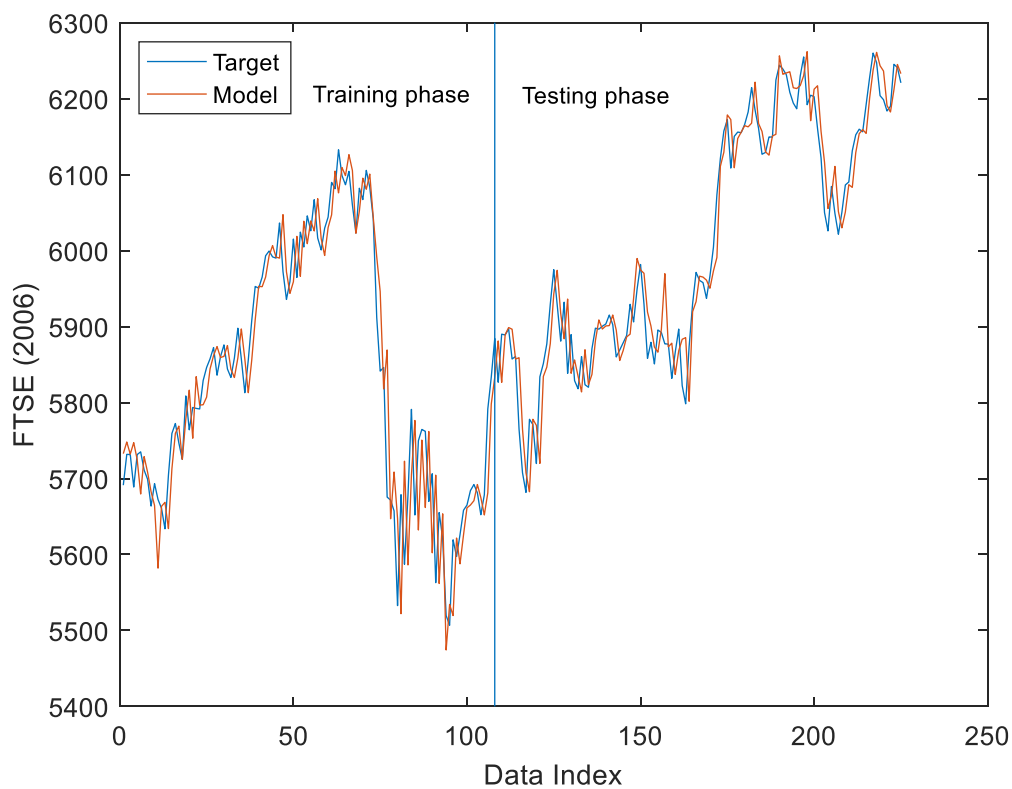


圖 15 模型訓練與預測結果(FTSE)

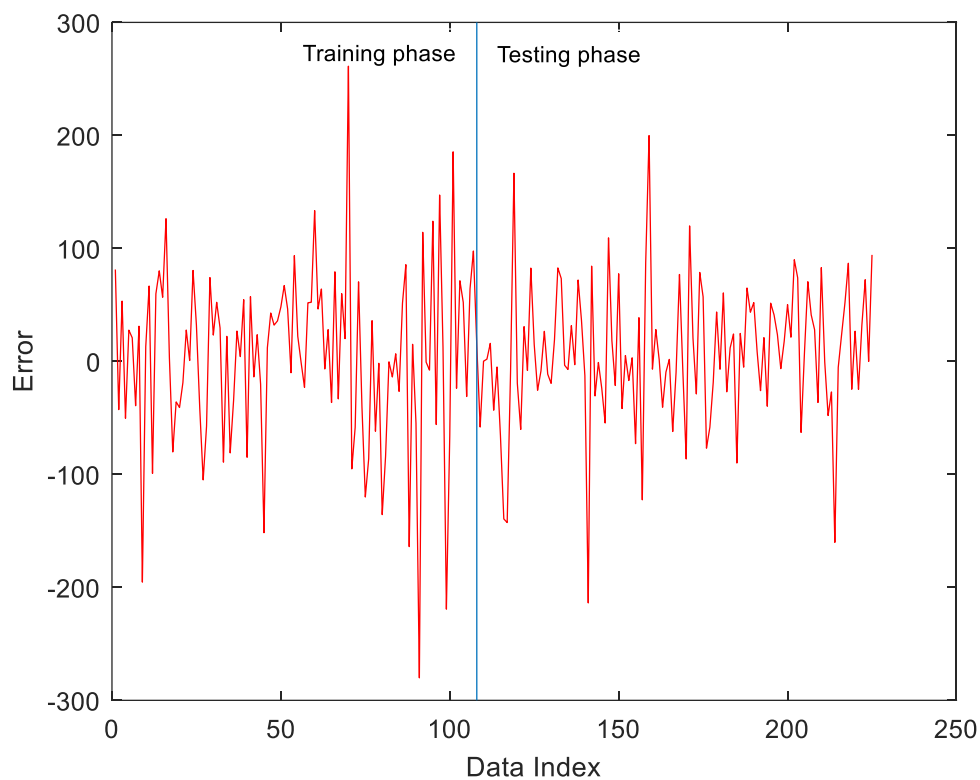


圖 16 訓練結果與目標之誤差(TAIEX)

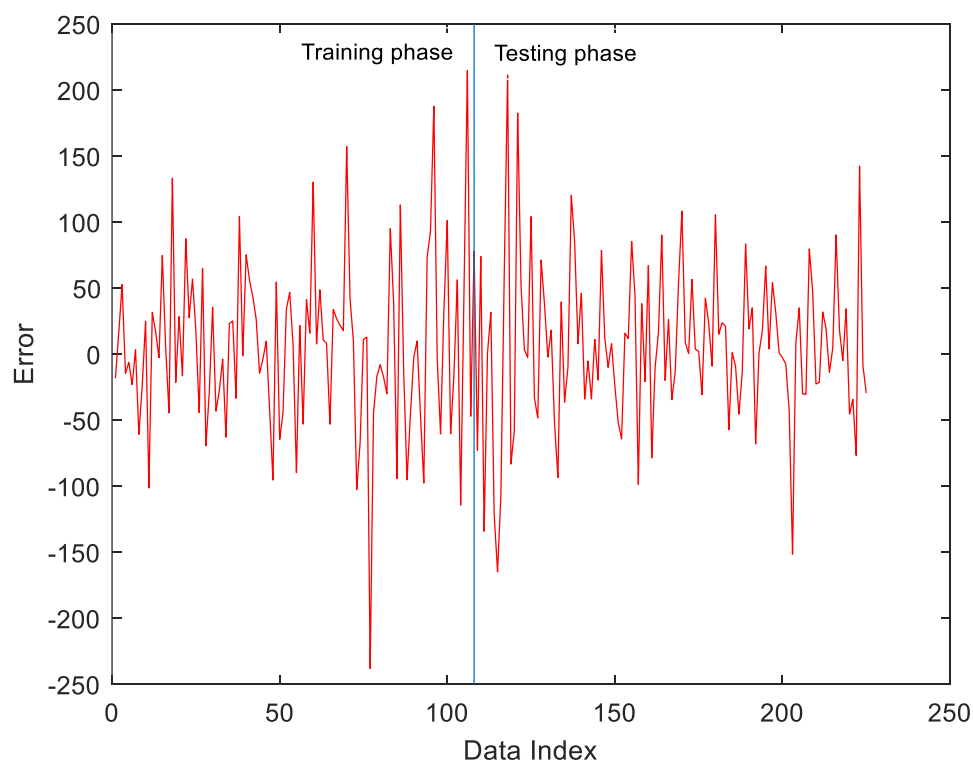


圖 17 訓練結果與目標之誤差(DJIA)

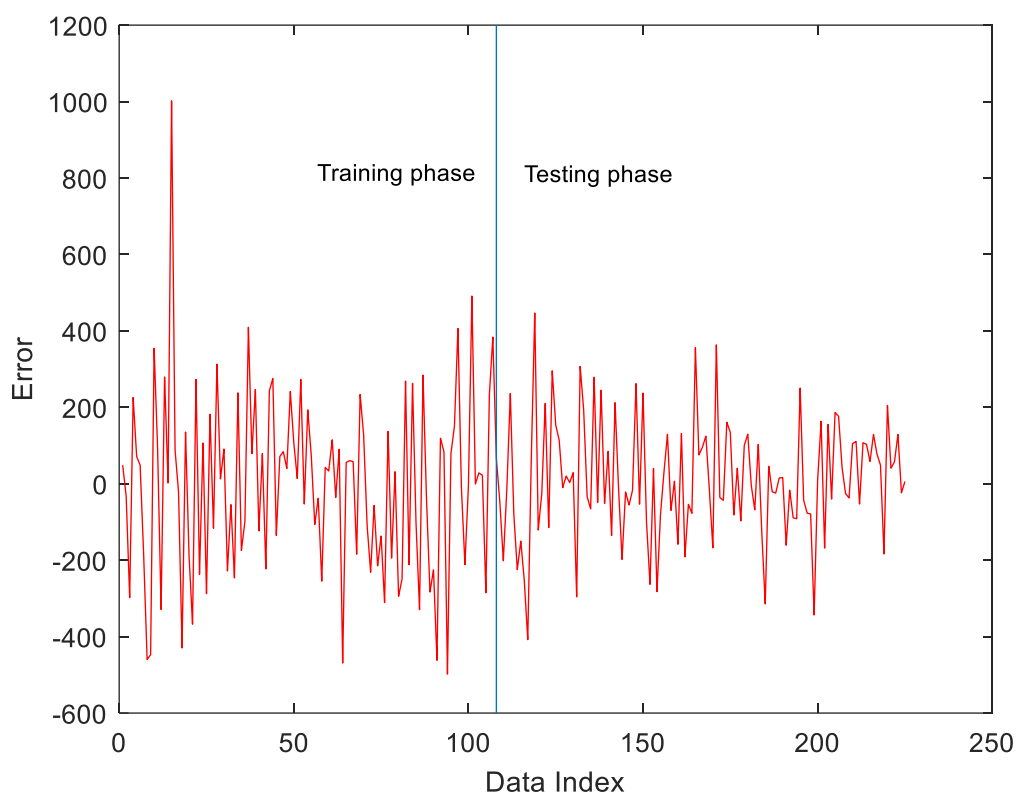


圖 18 訓練結果與目標之誤差(Nikkei)

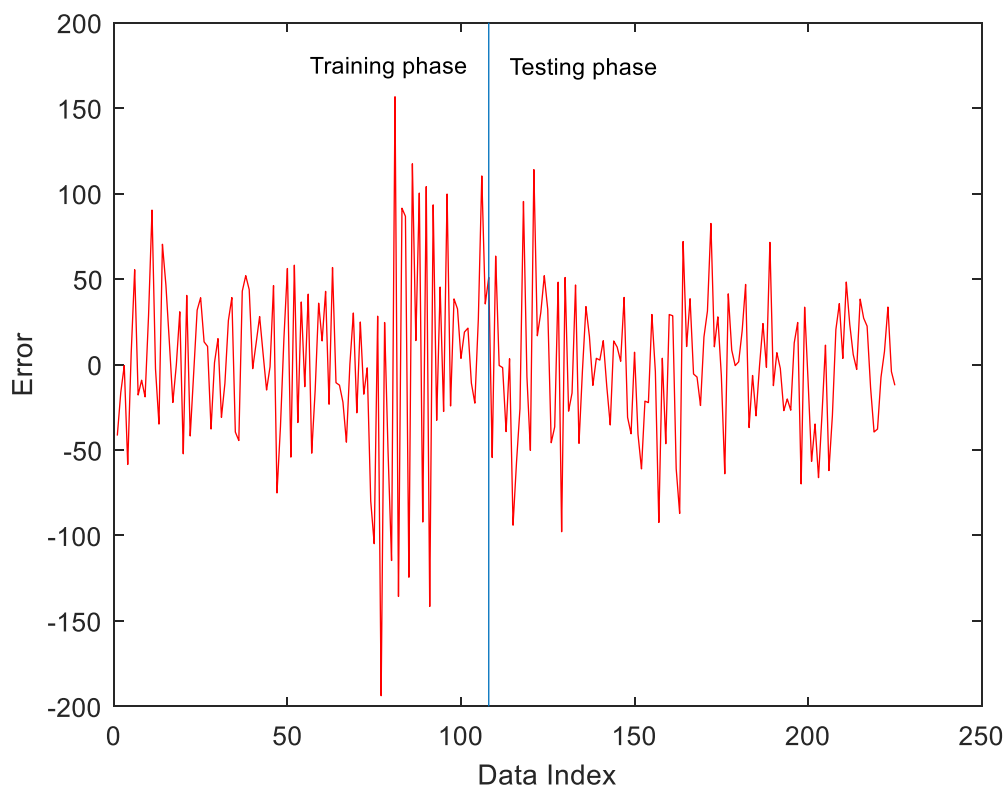


圖 19 訓練結果與目標之誤差(FTSE)



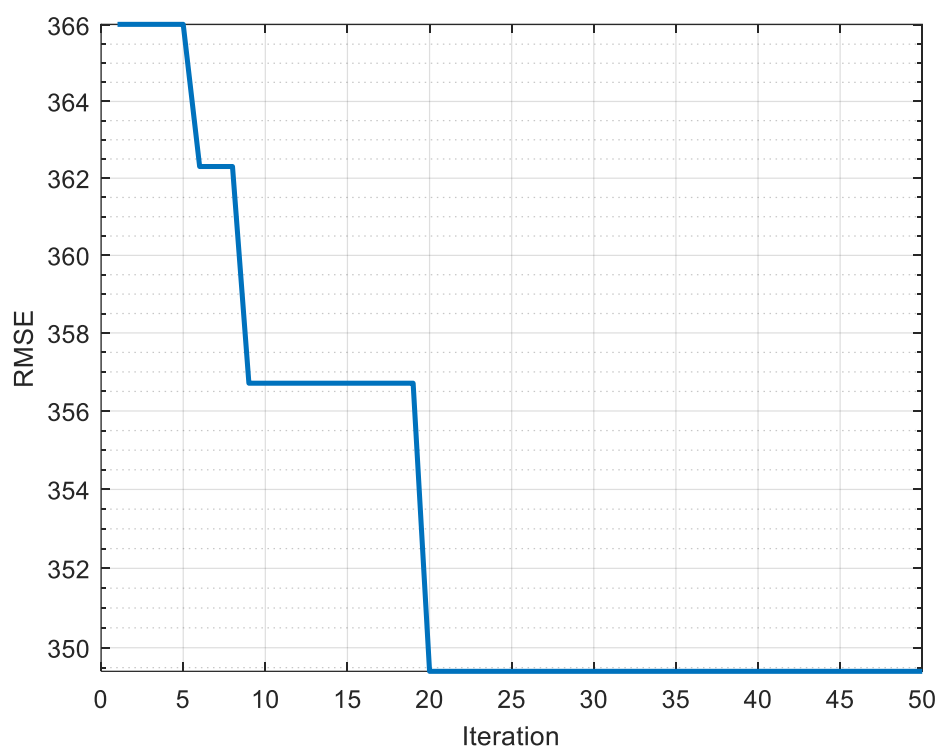


圖 20 模型學習曲線(實驗三)

## 第五章 實驗結果討論

本研究以傳統的模糊類神經網路為基礎提出了複數模糊類神經之模型。在模行內導入了自行建構的概念，使用 SC 分群演算法依照資料的分布情況自動決定分群數，因此能夠自動決定模型的初始參數設定，避免人為設定的干擾與偏頗，增加模型建置的客觀性與簡易性。將一般的模糊集合改為複數模糊集合，因複數模糊集合能夠根據目標數量來做相對應的歸屬程度數量調整，使模型能更加彈性地預測不同數量的目標並提升模型的預測效能。機器學習是使用 PSO-RLSE 的複合式學習演算法來自動調整系統參數，PSO 用於調整模糊規則前鑑部之參數，而 RLSE 則用於後鑑部參數之調整，此種分配方式可以降低個別演算法處理系統參數的負載程度，增進學習效率與收斂速度，提升機器學習時找到全域最佳解的機率，並且最佳化後的參數能使模型在複雜的時間序列預測上帶來良好的預測效能。本研究發現目標與模型輸出之間的誤差機率密度分佈平均值並非為零，因此本研究嘗試將一開始模型所產生的誤差回饋到模型內以改善模型的預測效能。回饋方式是將誤差輸入回模型的後鑑部，讓誤差再一次進入到 RLSE 演算法中，透過此方法能有效降低目標與模型輸出之間的誤差，因此模型在時間序列上能夠擁有良好的預測效

能。為了驗證本研究的可行性，我們進行了三個不同的時間序列預測實驗，從實驗中可以觀察到，本研究提出之方法確實能夠改善原始模型的預測效能，並且與其他文獻所提出之方法相比也有較好的預測效能。

在實驗一中，可以看到本研究一開始時所提出的模型預測結果幾乎都比其他研究所提出之模型要好。進行本研究所提出的誤差回饋方法後，發現 RMSE 又進一步降低，代表模型預測準確度又更加提升。將誤差回饋之模型與其他文獻模型相比後，發現其效能指標皆優於任何其他文獻之模型。觀察有無進行誤差回饋模型的誤差機率密度分佈參數，如圖 21 與表 14 所示，可以看到在 SSECI 目標中，沒有進行誤差回饋機制的誤差機率密度分佈平均值及標準差皆較大，在進行誤差回饋後皆減小。觀察相對應的 RMSE 指標，發現 SSECI 不論在訓練階段或測試階段的 RMSE 皆下降，因此可以推論當進行誤差回饋機制的誤差機率密度分布平均值和標準差下降時，模型的預測準確度會提高。

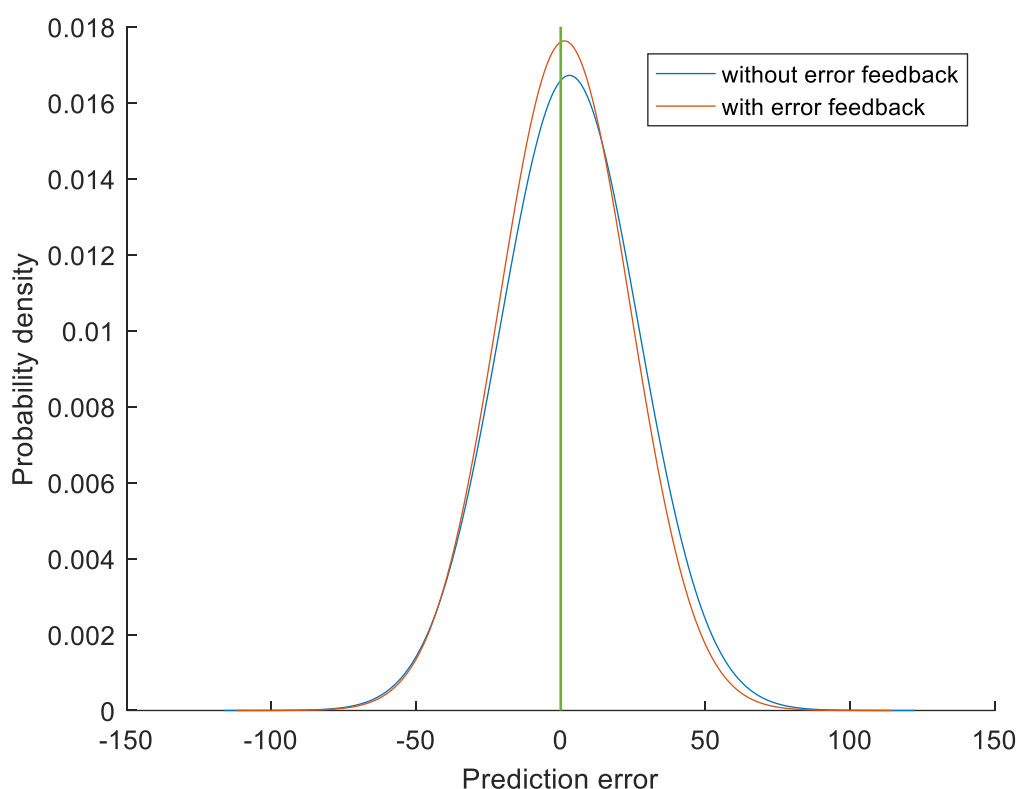


圖 21 誤差機率密度分布圖(實驗一：SSECI)

表 14 誤差機率密度分布參數(實驗一)

	SSECI	
	Mean	Standard deviation
without error feedback	3.0002	23.8615
with error feedback	1.2922	22.6335

在實驗二中，我們觀察了有無進行誤差回饋模型中各個目標的誤差機率密度分佈參數，如圖 22、圖 23、表 15 所示。我們可以看到兩個目標中沒有進行錯誤回饋的模型其平均值較接近零，但標準差卻較大；在進行錯誤回饋後的模型其平均值較遠離零，但標準差皆減小。觀察相對應的 RMSE，發現有進行誤差回饋之模型其在 DJIA 目標及 TAIEX 目標預測中，不論訓練或測試階段的 RMSE 皆優於沒有進行誤差回饋的模型。由此發現 RMSE 之下降與誤差機率密度分佈之標準差有一定程度的關聯性存在。與其他文獻進行比較後，也表現出本研究所提出之有進行誤差回饋之模型表現皆優於任何文獻模型與原模型。

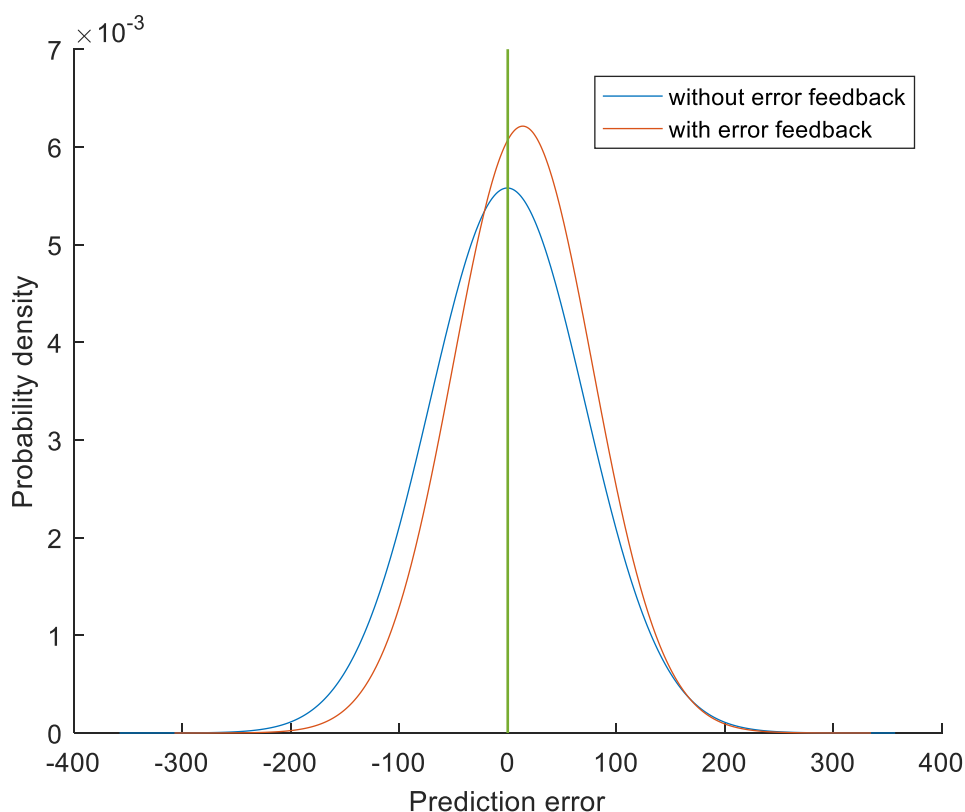


圖 22 誤差機率密度分布圖(實驗二：DJIA)

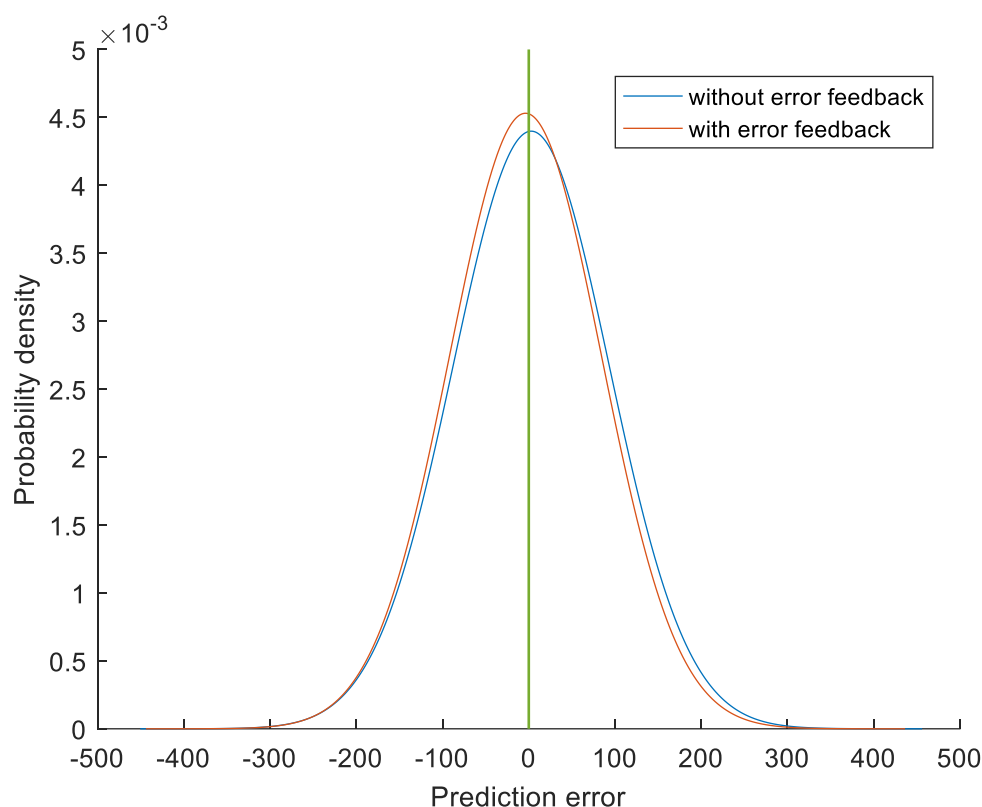


圖 23 誤差機率密度分布圖(實驗二：TAIEX)

表 15 誤差機率密度分布參數(實驗二)

	DJIA		TAIEX	
	Mean	Standard deviation	Mean	Standard deviation
without feedback	-0.3737	71.5060	2.7884	90.6944
with feedback	13.8281	64.2169	-3.7953	88.0560

在實驗三中，我們對四個目標進行了模型預測。觀察本研究所提出之模型與文獻模型比較發現，本研究之原模型在這四個目標預測的 RMSE 效能比較上皆優於文獻所提出之模型。觀察原模型與進行誤差回饋之模型，如圖 24、圖 25、圖 26、圖 27、表 16 所示，本研究發現採用誤差回饋方法時，TAIEX，DJIA 和 Nikkei 這三個目標的誤差機率密度分布的平均值及標準差皆降低；在 FTSE 目標中發現其平均值下降但標準差卻升高。進一步觀察進行誤差回饋模型在 RMSE 效能的表現上，發現在 TAIEX，DJIA 和 Nikkei 這三個目標中不論是訓練階段或測試階段其 RMSE 皆較原模型下降了許多；在 FTSE 目標上其在訓練階段的 RMSE 上升 2.0738，在測試階段 RMSE 也上升了 0.0017，但上升

幅度不大。在此四個目標的效能比較中，我們可以看到進行誤差回饋後的模型與原模型的 RMSE 相比其下降的比例高出了上升比例許多。

表 16 誤差機率密度分布參數(實驗三)

	TAIEX		DJIA	
	Mean	Standard deviation	Mean	Standard deviation
without feedback	5.6821	73.0899	8.5849	69.6265
with feedback	5.4334	70.6815	5.5712	63.1954

	Nikkei		FTSE	
	Mean	Standard deviation	Mean	Standard deviation
without feedback	4.1980	206.3767	3.7474	47.1688
with feedback	3.8767	200.4151	0.5743	48.4745

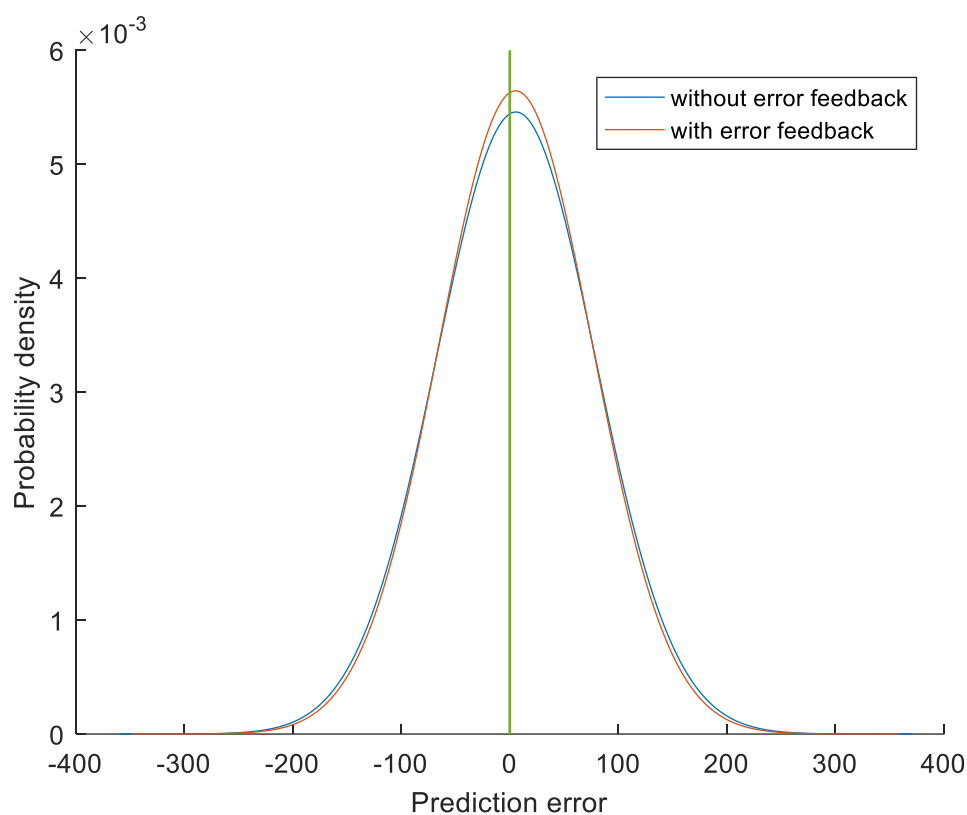


圖 24 誤差機率密度分布圖(實驗三：TAIEX)

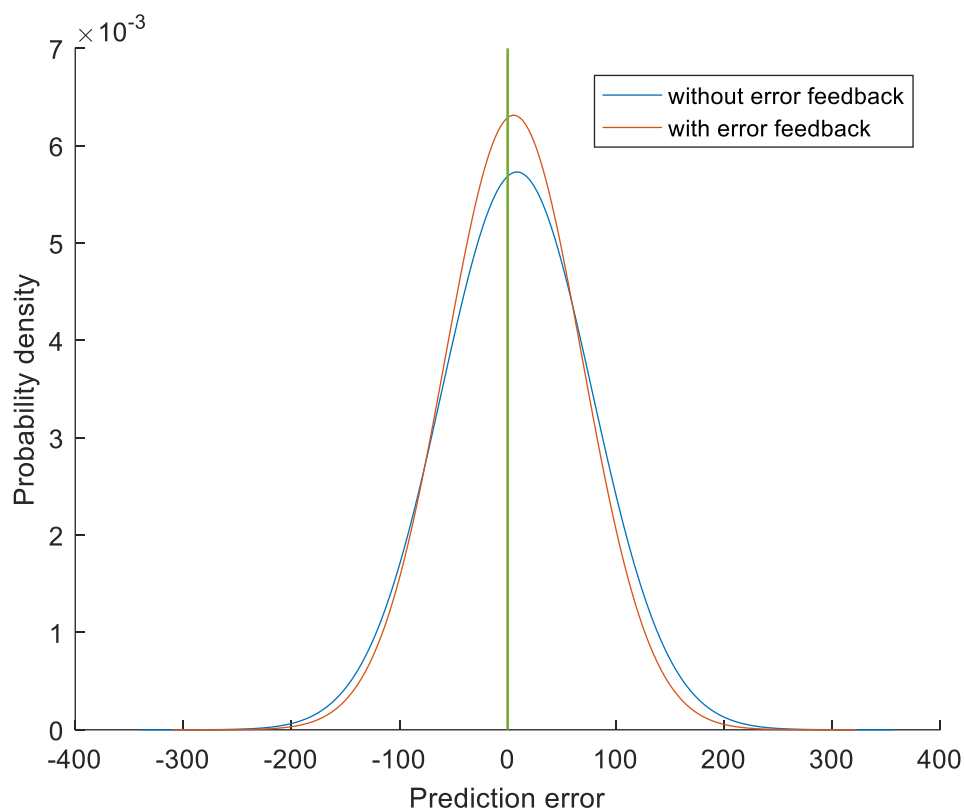


圖 25 誤差機率密度分布圖(實驗三：DJIA)

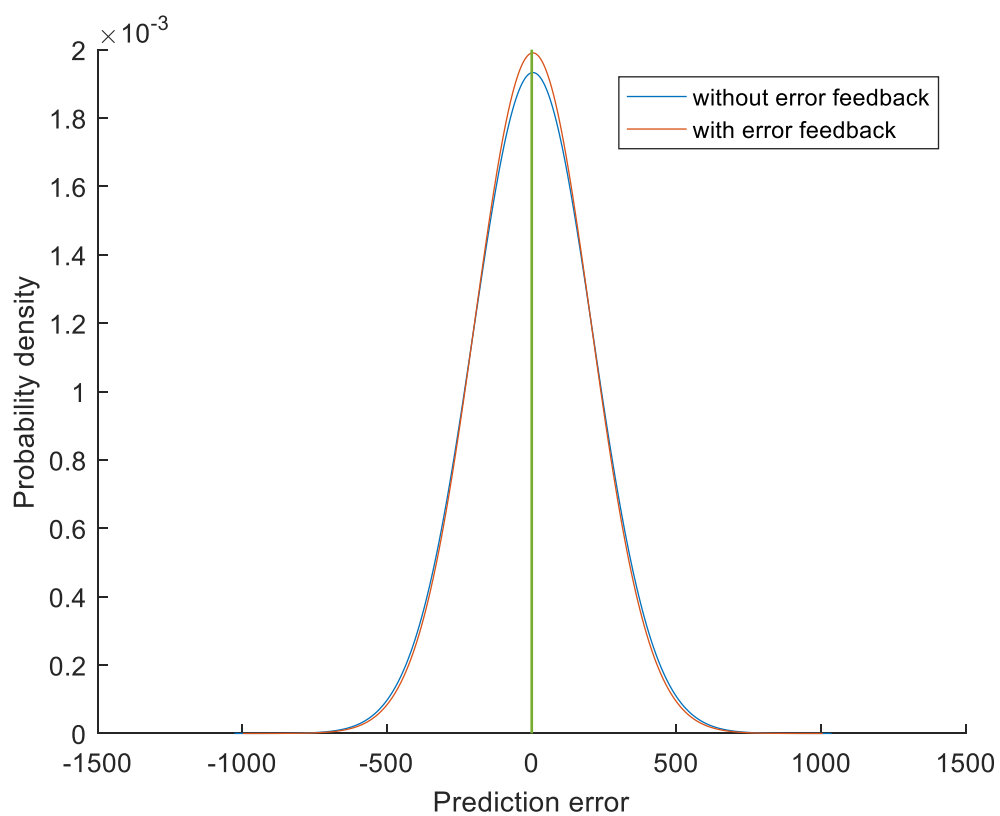


圖 26 誤差機率密度分布圖(實驗三：Nikkei)

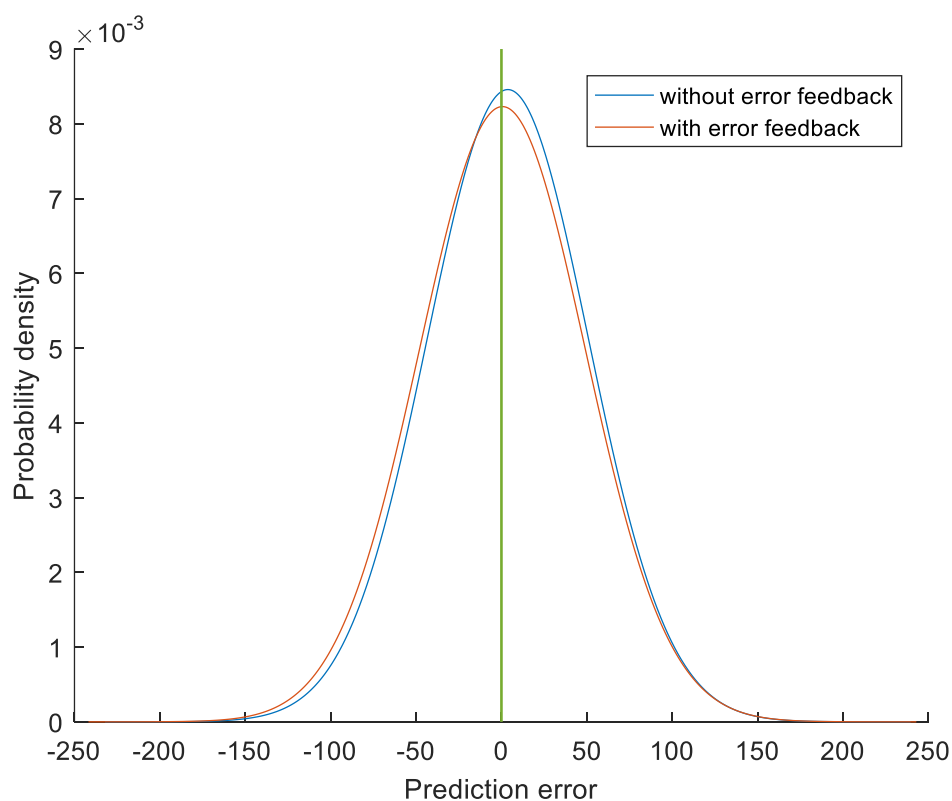


圖 27 誤差機率密度分布圖(實驗三：FTSE)

結合上述三個實驗的結果，可以驗證本研究提出的誤差回饋方法有其可行性。在每一個實驗中有進行誤差回饋方法的模型，其在每個目標的 RMSE 表現上幾乎都是下降的，並且 RMSE 的總下降幅度也比總升高的幅度來得大，由此可證此方法確實能夠有效提高模型的預測能力。在實驗結果中我們進一步觀察了進行誤差回饋的誤差機率密度分布圖之平均值和標準差的變化，發現當標準差下降時，RMSE 會減小；當標準差上升時，RMSE 會增加，因此推論 RMSE 之變化與誤差機率密度分布的標準差有較大的關連性。

## 第六章 結論

### 6.1 結論

隨著資訊儲存的成本降低，使得現今的資訊量以急遽的速度增長，因此特徵選取在現今已是一項不可或缺的技术。本研究提出的特徵選取方法是以夏農資訊熵為基礎，透過計算資料本身的亂度來加以計算特徵提供給目標的資訊量。此種特徵選取方法能從大量的特徵變數中找出提供給目標最大資訊量的特徵，並且能夠有效地減少掉冗餘的特徵變數，具有優秀的特徵選取能力。此種特徵選取方法並不相依於本研究所提出的模型，因此除了運用於本研究所提出的模型外也能運用於其他研究模型之中，在未來研究發展上有廣泛運用的可能性。

在過去的研究文獻中，大部份的研究在建立模型時，都是以研究者主觀的判斷來做決定。本研究提出了結構學習的方式來建立模型，不再依靠人為主觀的判斷來決定，而是以訓練資料本身來建立模型，因此更能增加模型建立的客觀性。

本研究使用的複數模糊集合將歸屬程度從一維的移動空間中擴展到了二維的平面圓盤空間。使用複數模糊集合，元素可獲取複數型歸屬度，亦即，可以同時產生出實部歸屬度以及虛部歸屬度，因此，使用複數模糊集合可以提升模糊類神經模型對於非線性複雜問題的映射能力。本篇論文所使用的模型至多能夠同時預測 6 個目標，在過去的文獻中，較少模型能夠達到同時預測如此大量的目標，因此往後的研究也能多加使用此方法來進行時間序列之模型預測。

在機器學習中是使用 PSO-RLSE 的複合式學習演算法來調整系統參數，先前已有研究證實此種方法能夠使模型在複雜的時間序列預測上帶來良好的預測效能。為了因應本研究之多目標預測模型，因此將 RLSE 演算法改良成能夠針對多目標來進行操作的方法，為往後將 RLSE 應用在多目標預測模型上提供了一種新的使用方法。

本研究根據模型輸出與目標間之誤差提出了一種誤差回饋方法，並在實驗過程中驗證了此方法降低誤差的有效性，使模型在預測時間序列上能夠有良好的預測準確性。在實驗結果中也進一步地發現了誤差機率密度之平均值與標準差與模型在預測效能上有



一定程度的關聯性。

## 6.2 未來研究方向

雖然在本篇論文中所使用的模型與機器學習演算法已經不錯的結果，但仍然有幾點可以在未來繼續加以研究，以持續增加模型的預測準確性。以下列出幾點問題提供未來研究探索：

### 1. 機器學習演算法之改良

粒子群最佳化演算法在粒子的維度過高時會導致搜尋全局最小化的能力下降，因此本篇論文使用了 divide-and-conquer 的概念，將模型後鑑部的參數交由遞迴最小平方法估計法來進行學習，大量地減少粒子群最佳化演算法中粒子的維度數量。但隨著模型輸入以及模糊集合數量的增加，粒子的維度也會隨之增加，導致粒子群最佳化演算法的效能降低，因此針對這個問題未來或許能再利用 divide-and-conquer 的概念將高維度的粒子切割成多顆低維度的粒子，或是再加入其他的演算法來分擔過多模型參數，以此提升學習演算法的效能。

### 2. 誤差機率密度分布參數之探討

在實驗結果中發現到了誤差機率密度分布之平均值、標準差與模型在預測效能上有一定程度的關聯性。在未來的研究上，或許能夠繼續深入探討誤差機率密度分布之平均值、標準差與模型效能的關聯性為何，又或者能夠透過誤差機率密度分布圖提出其他不同的誤差回饋方法來使模型擁有更好的預測精準度。

### 3. T-S 模糊系統

本研究中，採用的規則是源自於 T-S 模糊系統。該方式是以一前鑑部建立一後鑑部。但是若前鑑部數目過多，隨之建立的後鑑部規則數也會相當龐大。加上需要優化的參數數量，對模型是龐大的預測負擔。因此未來可以朝向將前鑑部與後鑑部的建立方式分離，使後鑑部數量不會與前鑑部有太大關係，如此便能使模型建立更為彈性。

## 參考文獻

- [1] M. Malliaris, and L. Salchenberger, "Using neural networks to forecast the S&P 100 implied volatility," *Neurocomputing*, vol. 10, iss. 2, pp. 183-195, March 1996.
- [2] M. Khashei and M. Bijari, "An artificial neural network ( $p, d, q$ ) model for timeseries forecasting," *Expert Systems with applications*, vol. 37, iss. 1, pp. 479-489, January 2010.
- [3] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, iss. 3, pp. 665-685, May 1993.
- [4] T. Marill and D. Green, "On the effectiveness of receptors in recognition systems," *IEEE transactions on Information Theory*, vol. 9, iss. 1, pp. 11-17, January 1963.
- [5] A. W. Whitney, "A direct method of nonparametric measurement selection." *IEEE Transactions on Computers*, vol. 20, iss. 9, pp. 1100-1103, September 1971.
- [6] F. Li, Y. Yang and E. P. Xing, "From lasso regression to feature vector machine," *Advances in Neural Information Processing Systems*, pp. 779-786, January 2005.
- [7] G. D. Tourassi, E. D. Frederick, M. K. Markey and & C. E. Floyd, "Application of the mutual information criterion for feature selection in computer-aided diagnosis," *Medical Physics*, vol. 28, iss. 12, pp. 2394-2402, December 2001.
- [8] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, iss. 3, pp. 338-353, June 1965.
- [9] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, Vol. 7, No. 1, pp. 1-13, 1975.
- [10] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, iss. 1, pp. 116-132, January 1985.
- [11] D. Ramot, R. Milo, M. Friedman and A. Kandel, "Complex fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 10, iss. 2, pp. 171-186, August 2002.
- [12] C. T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Transactions on Computers*, vol. 40, iss. 12, pp. 1320-1336, December 1991.
- [13] I. Rojas, O. Valenzuela, F. Rojas, A. Guillen, L. J. Herrera, H. Pomares, L. Marquez, M. Pasadas, "Soft-computing techniques and ARMA model for time series prediction," *Neurocomputing*, vol. 71, iss. 4-6, pp. 519-537, January 2008.
- [14] A. Sharifian, M. J. Ghadi, S. Ghavidel, L. Li, J. Zhang, "A new method based on

- Type-2 fuzzy neural network for accurate wind power forecasting under uncertain data,” *Renewable Energy*, vol. 120, pp. 220-230, May 2018.
- [15] J. Tang, F. Liu, Y. Zou, W. Zhang and Y. Wang, “An improved fuzzy neural network for traffic speed prediction considering periodic characteristic,” *IEEE Transactions on Intelligent Transportation Systems*, vol.18, iss. 9, pp. 2340-2350, September 2017.
- [16] R. W. Kristjanpoller and V. K. Michell, “A stock market risk forecasting model through integration of switching regime, ANFIS and GARCH techniques,” *Applied Soft Computing*, vol. 67, pp. 106-116, June 2018.
- [17] O. C. Yolcu and F. Alpaslan, “Prediction of TAIEX based on hybrid fuzzy time series model with single optimization process.” *Applied Soft Computing*, vol. 66, pp. 18-33, May 2018.
- [18] Y. Hou, L. Zhao and H. Lu, “Fuzzy neural network optimization and network traffic forecasting based on improved differential evolution,” *Future Generation Computer Systems*, vol. 81, pp. 425-432, April 2018.
- [19] J. Kennedy and R. Eberhart, “Particle swarm optimization,” *IEEE International Conference on Neural Networks*, vol. 4, pp. 1942-1948, November 1995.
- [20] J. S. R. Jang, C. T. Sun and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Upper Saddle River NJ: Prentice Hall, 1997.
- [21] C. E. Shannon, “A Mathematical Theory of Communication,” *The Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, July, October, 1948.
- [22] J. B. MacQueen, “Some methods for classification and analysis of multivariate observations,” *Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 14, pp. 281-297, 1967.
- [23] J. C. Dunn, “A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters,” *Journal of Cybernetics*, vol. 3, iss. 3, pp. 32-57, September 1973.
- [24] J. C. Bezdek, R. Ehrlich, and W. Full, “FCM: the fuzzy c-means clustering algorithm,” *Computers and Geosciences*, vol. 10, iss. 2-3, pp. 191-203, 1984.
- [25] S. L. Chiu, “Fuzzy model identification based on cluster estimation,” *Journal of Intelligent and Fuzzy Systems*, vol. 2, iss. 3, pp. 267-278, June 1994.
- [26] K. Huarng and H. K. Yu, “A Type 2 fuzzy time series model for stock index forecasting,” *Physica A: Statistical Mechanics and its Applications*, vol. 353, pp. 445-462, August 2005.
- [27] C. H. Cheng and G. W. Cheng and J. W. Wang, “Multi-attribute fuzzy time series method based on fuzzy clustering,” *Expert Systems with Applications*, vol. 34, iss. 2, pp. 1235-1242, February 2008.
- [28] S. M. Chen, “Forecasting Enrollments Based On High-order Fuzzy Time Series,”

- Cybernetics and Systems, vol. 33, iss. 1, pp. 1-16, 2002.
- [29] L. W. Lee and L. H. Wang and S. M. Chen and Y.H. Leu, "Handling forecasting problems based on two-factors high-order fuzzy time series," IEEE Transactions on Fuzzy Systems, vol. 14, iss. 3, pp. 468-477, June 2006.
- [30] E. Egrioglu, C. H. Aladag, U. Yoclu, V. R. Uslu and N. A. Erilli, "Fuzzy time series forecasting method based on Gustafson–Kessel fuzzy clustering," Expert Systems with Applications, vol. 38, iss. 8, pp. 10355-10357, August 2011.
- [31] L. Wang, X. Liu and W. Pedrycz, "Effective intervals determined by information granules to improve forecasting in fuzzy time series," Expert Systems with Applications, vol. 40, iss. 14, pp. 5673-5679, October 2013.
- [32] E. Bas, U. Yoclu, E. Egrioglu and C. H. Aladag, "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.
- [33] O. C. Yoclu, U. Yoclu, E. Egrioglu and C. H. Aladag, "High order fuzzy time series forecasting method based on an intersection operation," Applied Mathematical Modeling, vol. 40, iss. 19-20, pp. 8750-8765, October 2016.
- [34] W. Zhang, S. Zhang, S. Zhang, D. Yu and N. N. Huand, "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, May 2017.
- [35] C. Li and T. W. Chiang, "Complex Neurofuzzy ARIMA Forecasting—A New Approach Using Complex Fuzzy Sets," IEEE Transactions on Fuzzy Systems, vol. 21, iss. 3, pp. 567-584, November 2012.
- [36] K. H. Huarng, T. H. K. Yu and Y. W. Hsu, "A multivariate heuristic model for fuzzy time-series forecasting," IEEE Transactions on Systems, Man, and Cybernetics., vol. 37, iss. 4, pp. 836–846, July 2007.
- [37] S. M. Chen, "Forecasting enrollments based on fuzzy time series," Fuzzy Sets System, vol. 81, iss. 3, pp. 311–319, August 1996.
- [38] T. H. K. Yu and K. H. Huarng, "A bivariate fuzzy time series model to forecast the TAIEX," Expert System with Applications, vol. 34, iss. 4, pp. 2945–2952, May 2008.
- [39] T. H. K. Yu and K. H. Huarng, "Corrigendum to A bivariate fuzzy time series model to forecast the TAIEX," Expert Systems with Applications, vol. 37, no. 7, p. 5529, July 2010.
- [40] S. M. Chen and C.D. Chen, "TAIEX forecasting based on fuzzy time series and fuzzy variation groups." IEEE Transactions on Fuzzy System, vol. 19, iss. 1, pp. 1–12, February 2011.
- [41] H. J. Sadaei, R. Enayatifar, F. G. Guimarães, M. Mahmud and Z. A. Alzamil, "Combining ARFIMA models and fuzzy time series for the forecast of long memory time series," Neurocomputing, vol.175, pp. 782-796, January 2016.

- [42] T. T. Hsieh, H. F. Hsiao and W. C. Yeh, "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, March 2011.
- [43] 國立中央大學資訊管理所李俊賢教授, 研究生訓練課程內容2016-2018, 紀錄筆記。(未發表)