A Novel Intelligent Time Series Forecasting Approach Using Sphere Complex Fuzzy Sets

Chi Feng Lina and Chunshien Lib

Information Management. National Central University, National Central University, Taoyuan, TAIWAN

aEmail: j8888888871@yahoo.tw

bEmail: [jamesli@mgt.ncu.edu.tw](mailto:jamesli@mgt.ncu.edu.tw) (corresponding author)

*Abstract*—The prediction of time series is a widely used and important research topic. The artificial intelligence models are currently being widely used in this topic, such as neuro-fuzzy systems. This paper proposes a new type of fuzzy sets, namely Sphere Complex Fuzzy Sets (SCFSs), which are applied to neuro-fuzzy system model and multi-objective time series prediction. Using SCFSs, the neuro-fuzzy model proposed in this paper has the capability of multiple complex-valued outputs. Every output can have real and imaginary parts for two different real-valued targets, respectively. With regard to feature selection, this study used multi-target feature selection to screen out favorable features for all objectives and use these features as a model input to reduce the overall computational loading on the model and improve data utilization efficiency. In terms of models, a multi-layer neural network architecture is used and includes an input layer, an SCFS neural layer, a premise neural layer, an aim object neural layer, a T-S neural layer (Takagi-Sugeno neural) and output layer, where, the aim object neural layer is an innovative model construction of this paper. Its purpose is to make the data driven model construction have structural adaptability and application flexibility. In machine learning, we use the divide-and-conquer principle when training the model. The parameters of the SCFS neural layer are optimized using Particle swarm optimization; the parameters of the T-S neural layer are optimized using recursive least squares estimation (RLSE); other neural layers not need to be optimized. In the multi-target time series forecasting, we designed three experiments to test the experimental results and compare performance with different methods. Although most of the literature methods aim at single target or a few methods for double-target time series prediction, the multi-target prediction method proposed in this paper shows excellent performance through performance comparison.

Keywords— Sphere complex neuro-fuzzy system (SCNFS); complex fuzzy set (CFS); complex neuro-fuzzy system (CNFS); time series forecasting.

# Introduction

Time series data is a collection that uses time series as the sorting basis, such as stock market, exchange rate, energy consumption and so on. The time series can be regarded as data observed in a discrete time sequence and can be analyzed by mathematical methods or statistical models. With the development of present information technology, both the Internet of Things and the information brought by the Internet, these data have grown explosively, resulting in a quantitative change in the amount of data. These data can no longer be dealt with by human intelligence. We need to use machine learning or data mining to find out the hidden relationships or rules hidden in the data in order to discover the value.

Time series is a very important research topic. The prediction of time series data is widely used in many fields such as medicine, energy planning, financial forecast and so on. In the medical field, Wei et al. [28] predicted the number of daily patients through a mixed AR-EMD hybrid model for resource allocation, Osthus et al. [31] predicted seasonal infectious diseases and helped in public health planning and outbreak response, S. S. Jones et al. [16] used multivariate analysis to fit the time series model to predict the number of emergency patients so that hospitals could better allocate hospital resources, with regard to the number of emergency patients, Aboagye-Sarfo et al. [3] also analyzed and proposed a vector-autoregressive moving average (vector-ARMA) model to compare the differences between multivariate and univariate; in the area of energy prediction, Camelo et al. [9] used autoregressive integrated moving average with explanatory variable (ARIMEX) and Holt-Winters (HW) to combine with neural networks to forecast wind power generation, besides this national level application, Alobaidi et al. [1] developed a new framework specifically for forecasting household electricity consumption; in financial forecasting, in 2012, Li et al. [27] used the autoregressive integrated moving average (ARIMA) model and the neuro-fuzzy system (NFS) to predict the closing price of stocks, in the next year, Li et al. [23] used a neuro-fuzzy system to predict two targets simultaneously; in 2017, Koijen et al. [20] predicted each other's data through stocks and bonds; Pan et al. [32] compare short-term and long-term leverage by predicting the volatility of stocks.

In all fields, it is most difficult to predict financial time series, because it contains many factors, such as the company's operating conditions, international situation and so on. In financial predictions, many methods have been proposed [14][21][23][27][35][37], such as ARIMA[23][27], fuzzy theory, neural network computation, neuro-fuzzy hybrid systems and so on. Among them, neuro-fuzzy systems (NFSs) [14][21][35][37] are the most commonly proposed. Neuro-fuzzy systems are fuzzy systems that use learning algorithms derived from or inspired by the neural network theory [30] to determine their parameters (fuzzy sets and fuzzy rules) by processing data samples. The neural network was created by McCulloch et al. [30] in 1943 based on mathematics and an algorithm called threshold logic, a neural network model in which neurons can receive information to calculate whether or not to produce excitatory response. In 1956, Rochester et al. [33] combined this mathematical model with the Hebb's Law [15] to create a perceptron to simulate the human brain, but fewer perceptron neurons, and its transmitted signals with a weight of 0 or 1, This method still cannot deal with the XOR problem (non-linear problem), and at that time the computer did not have enough ability to calculate. Until 1975, Werbos [42] proposed a back-propagation (BP) algorithm. This algorithm effectively solved the XOR problem and the problem of training multi-layer neural networks. Afterwards, many other methods of artificial intelligence were proposed, such as support vector machines [10], Bayes classifier [38] and so on. Until recently, due to the rapid development of information technology, the computation speed has increased, it leads to artificial intelligence in certain areas has surpassed the human level, and neural networks have been heavily studied again.

Neuro-fuzzy systems are a combination of neural networks and fuzzy theory [44]. Neuro-fuzzy systems are usually represented as multi-layer feed-forward neural networks, such as ANFIS [18]. The characteristics of neuro-fuzzy systems make it have a good effect on the prediction of time series. Therefore, most of the current research in the area of time series prediction uses a neuro-fuzzy system as model structure.

This study adopts a neuro-fuzzy system and follows the IF-THEN rules of the expert system to construct a multi-layer neuron architecture. In order to increase the flexibility of the model, it is different from the traditional IF–THEN rule method, in this study, the aim object neural layer is used to connect with the neural network layer, which makes the number of neurons of each layer can be different. In terms of the model, we combine fuzzy theory with the neural network to form neuro-fuzzy system, the sphere complex neuro-fuzzy sets and Takagi-Sugeno linear function are used, and the two are combined through the aim object layer. Through this model and machine learning we expect that the prediction of time series can be more accurate than other literature.

In 1965, the concept of fuzzy sets was proposed [44] by Zadeh. The purpose was to make human ideas become mathematic formula. Through the concept of fuzzy sets, elements can belong to a certain set and the membership degree is between 0 and 1. Afterwards, Buckley proposed the concept of fuzzy complex numbers [5] – [7], In 2002, Ramot et al. [34] proposed complex fuzzy sets (CFSs), in which the complex-valued membership degree can be obtained by the function, which allows the membership degree to present in a unit disk of the complex plane (UDCP) with a radius of 1. This concept makes membership degree more abundant. We can obtain the complex-valued output through the complex neural fuzzy set system (CNFS) [24][26]. The real part and the imaginary part can be used to predict different targets respectively, so we can predict two targets simultaneously. At present, there are a lot of research of dual-target prediction [23][24][26]. In order to forecast for more targets at the same time, this paper improves the original neuro-fuzzy set system and changes the complex fuzzy sets (CFSs) to sphere complex fuzzy sets (SCFSs). The membership degree is presented in the three–dimensional space. Through projecting numbers in different dimensions and combining multiple complex-valued, the membership degree is more abundant, and it is also possible to predict multiple targets at one time.

In order to make the data more effective for use, many scholars have proposed methods to deal with the data in data preprocessing. In time series data processing, Mikalsen et al. [29] deal with missing values in data through clustering and machine learning. Zhang et al. [43] proposed refined composite multiscale weighted-permutation entropy (RCMWPE) to improve the original multiscale weighted-permutation entropy (MWPE) to enable more accurate use of data. Alves et al. [2] proposed a completely new chaos description through theories in periodic, chaotic, and stochastic systems. In this study, we use the concept of Shannon entropy [39] to calculate feature’s individual contribution to targets. In addition, we adopted the concept of multi-target feature selection [22] to calculate the effective information for each feature on the target as a basis for selecting training data. Extracting the most effective data from the raw data can not only reduce the model computational burden, but also effectively increase the performance of the prediction. In the machine learning section, we use the popular particle swarm optimization (PSO) [19] and the well-known recursive least square estimation (RLSE) [17] to optimize parameters, and integrate the two algorithm called the PSO–RLSE method [25]. In this study, different parts of the parameters are trained through different algorithms. The divide-and-conquer principle is used to reduce the search dimension of the algorithm, which makes it easier for the model to find the best solution and improve overall performance.

The rest of this paper is as follows. In Section II, a series of research methods will be described in detail: sphere complex fuzzy sets, multi-targets feature selection, structure learning, model input and output, and machine learning. In Section III, three examples for time series forecasting are given to test the proposed approach. Experimental results are given and compared with other papers. In Section IV, we discuss the contents of the data and the experimental results. Finally, in Section V, the contributions of the research and future development will be summarized.

# Methodology

## Sphere Complex Fuzzy Sets

In the past, the concept of fuzzy sets can derive the one-to-one membership degree of elements to a set. The complex fuzzy sets [34] can have a complex-valued membership degree, but in order to make the application more widespread, we hope to make the membership degree more abundant through one idea.

The sphere complex fuzzy set is an innovative concept proposed in this paper. Through this concept, one piece of data can be converted into multiple complex-valued membership degrees, so that the model can predict multiple targets at a time. First, place the membership degree of Gaussian function in a sphere complex fuzzy set with a radius of 1 (Fig. 1), and a set of spatial vectors can be obtained , whose components are expressed as follows.

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |
|  | , |  |

where, is the membership degree of Gaussian function (25); ; . With the dismantling of , at least four groups of complex-valued membership degrees are available, including the following membership degrees.

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |
|  | , |  |
|  | , |  |

where, 。

1

1. 球式複數模糊集。為高斯函數的歸屬程度，可透過與平面的夾角及計算出在各維度上的投影。一個球式複數模糊及他的歸屬度訊息，是由一個球式空間向量所攜帶，會隨著輸入再求空間裡改變。

## Multi–Target Feature Selection

特徵挑選不僅能刪去負面的資訊來源，更有助於減輕模型的運算負擔，故是資料前處理中重要的一環。面對多個目標時的特徵挑選，更需要謹慎的處理，才能帶來正面的效果。本論文將同時預測多個目標，故使用夏農資訊熵[39]概念，並參考多目標特徵選取方法[22]，實作特徵挑選，最後從挑選後的特徵中取得訓練資料。

熵一詞最早是由德國物理學家Clausius於 1854 年提出[12]，是一種對物理系統之無秩序或亂度的量度。在1948年，學者Shannon則提出了資訊熵[39]的概念，熵定義為資訊內容其不確定性的量值，若資訊的隨機性越高，則資訊熵值會越高。對於某一個隨機變數*X*，資訊熵[39]的定義如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，是隨機變數的資訊熵是事件的發生機率密度則被視為的資訊混亂度。

但若大於，則部分會是負數，會影響到整體的期望值，所以我們對公式做了一些更改，更改後的公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |

其中，為很小的正值。

由於我們特徵的選擇是針對目標，所以我們透過資訊熵的概念，計算每個特徵與目標之間的影響資訊量(Influence information)[22]，公式如下。

|  |  |
| --- | --- |
| , |  |

其中，為隨機變數對隨機變數的影響資訊量為在事件值為正時的隨機變數為在事件值為負時的隨機變數為隨機變數以及隨機變數的互資訊為隨機變數以及隨機變數的互資訊互資訊的定義公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |

其中，為目標的資訊熵;為事件為正數時所對應的隨機變數的資訊熵;為事件為負數時所對應的隨機變數的資訊熵;條件式資訊熵公式如下。

|  |  |
| --- | --- |
| , |  |
| , |  |
| , |  |
| , |  |

其中，為事件為正的變數;為事件為負的變數;為事件為正數時的機率密度;為事件為正數時所對應的事件*y*的機率密度;為事件為負數時的機率密度;為事件為負數時所對應的事件*y*的機率密度。

透過上述影響資訊量[22]的公式，可以得到每個特徵變數對每個目標的影響資訊量，為了方便計算影響資訊量，我們可以將所有特徵變數以及目標組合成一個資料矩陣(Data matrix, DM)，如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，為第組候選特徵，，為一個特徵的總筆數，為特徵總數量; 為第個目標變數，，為一個目標的總筆數，為目標的總數量。

我們利用資料矩陣每行的特徵資料，與其他行做影響資訊量的運算，透過特徵與特徵之間以及對於第個目標的影響資訊量整理出第個目標的影響資訊矩陣(Influence information matrix, IIM)。表示如下。

|  |  |
| --- | --- |
| , |  |

其中，為第個特徵變數;為第個目標變數; 為特徵的總數量;為特徵變數對特徵變數的影響資訊量， 且。

而後可依據影響資訊矩陣裡的影響資訊量做多目標的特徵選取，步驟如下。

Step 1 : 算出第個特徵對第個目標的資訊增益量(Selection gain)標記為，其中，為第個特徵變數;為第個目標變數。資訊增益量公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，為對的影響資訊量;為第個已選特徵池(Selected pool, SP), ; 是第個已選特徵池中，第個元素； 為對中已存在特徵的冗餘資訊量(Redundancy information)。冗餘資訊量公式如下。

|  |  |
| --- | --- |
| , |  |

其中，代表第個已選特徵池內的特徵個數;為對內的第個特徵變數的影響資訊量;為內的第個特徵變數對的影響資訊量。經過上述計算若大於0，則將特徵加入第個已選特徵池 中。

Step 2 :無論重疊與否，將所有已選特徵池中出現過的特徵變數記錄下來，儲存成。其中，為目標變數個數;，是中第個特徵變數。計算每個特徵出現在所有已選特徵池的次數，標記為。

Step 3 :透過即可計算覆蓋率，公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

計算的平均，標記為。

Step 4 :累加每個已選特整池裡，特徵的資訊增益量。

|  |  |
| --- | --- |
| , |  |

計算的平均標記為。

Step 5 :根據累加後的資訊增益量和覆蓋率，計算出特徵的有效貢獻量。

|  |  |  |
| --- | --- | --- |
|  | , |  |

Step 6 :測試中所有的特徵變數，若，則將累加。

Step 7 :設定上下界，標記為和，透過上下界找出，表示最後選取的特徵數目。若介於上下界間，則將設定成;若小於下界，則將設定成;若大於上界則將設定成。

Step 8 :將排序，並選取前個特徵變數加入最後特徵池(Final pool, FP)中，當作多目標的特徵挑選結果。

## Structure Learning

結構學習是為了將訓練資料可以更有邏輯的應用到模型建造中，此外結構學習中的結果，也會成為之後參數學習的一部分。在本研究中，會將這些不同輸入維度的訓練資料，透過減數分群演算法[11]進行分群。並將分群後的群中心配合每個維度的標準差形成模糊集，各個維度的模糊集個數總和即為第1層神經元的數量，本研究採用高斯型態的模糊集合，高斯函數的公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，為輸入變數，和為群中心和標準差的參數。

基於各個輸入維度的模糊集，我們可以組成個區塊。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，為第個輸入維度的分群個數，亦即模糊集個數。以第個區塊的組成為例，如下。

|  |  |
| --- | --- |
| , |  |

，其中，為第個輸入的語意變數;為第個輸入變數，;為第個區塊中第個輸入語意變數的模糊集合，使用高斯函數建構，如(25)。



1. 模糊集合輸入空間(2維)。兩個輸入維度，各分出3群則會形成共9塊區域，其中z軸為該區的資料密度。



1. 資料密度總和。可看出部分區塊資料密度總和較高，代表用於建造模型較有利。

為了模型的運算效率，以及降低模型運算負擔，我們將會篩選出若干個較重要的區塊，成為第2層中的神經元數量。若以兩個輸入維度為例子可得到如圖(Fig. 2)的區塊，而後將透過資料密度的概念，將資料灑入區塊中，並計算資料密度量，累加資料密度量後可得到如圖(Fig. 3)，從中可挑選資料密度較高的區塊，作為第2層神經元。詳細步驟如下:

Step 1:從各個輸入維度與模糊集可以得到每個區塊的資料密度量，以第個區塊為例，資料密度量公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，為第個輸入維度的第筆資料;為第個區塊中第個輸入維度的模糊集。

Step 2:將每個區塊的資料密度量累加標記為，公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，為資料總筆數。計算平均值標記為，標準差標記為。

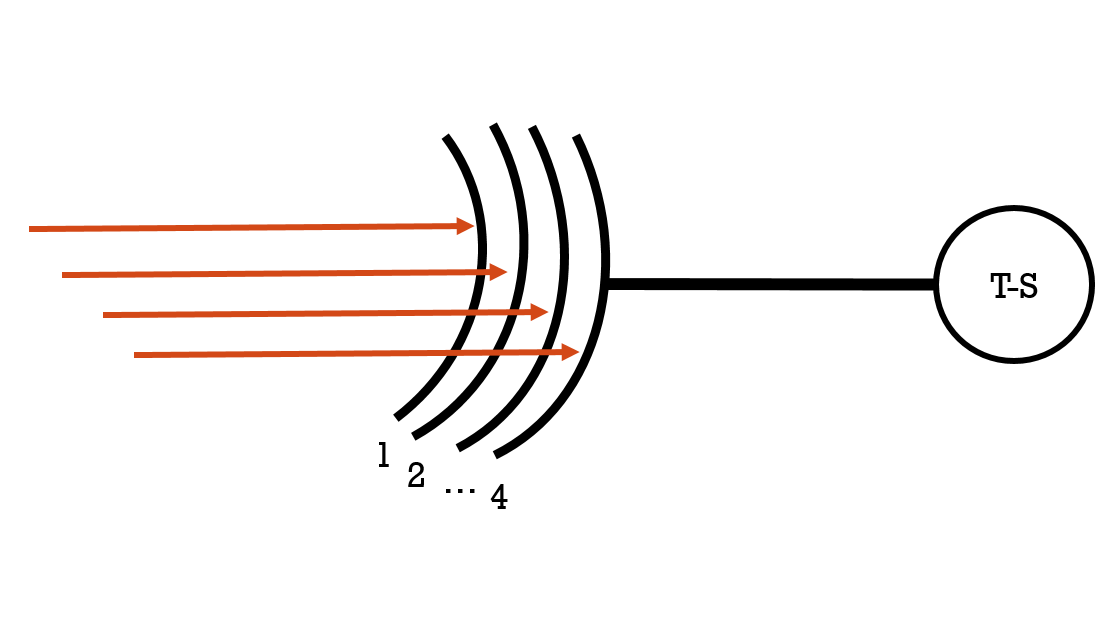
Step 3:查看每個區塊，若，則將累加。設定上下界，標記為和，透過上下界找出，表示最後選取的區塊數目。本研究中所有實驗皆設定為15，皆設定為4。若介於上下界間，則將設定成;若小於下界，則將設定成;若大於上界則將設定成。

Step 4:將排序，並保留前個區塊，當作之後模型第2層的神經元。

第3層的神經元數目將由訓練資料集合進行減數分群[11]決定，決定神經元數量後，將透過Fuzzy C–Mean分群方法[36]決定第3層神經元–箭靶神經元的相關數值，步驟如下:

Step 1:將輸入資料集合進行減數分群[11]，決定箭靶神經元個數為，即為第3層神經元的數目。

Step 2: 決定群數之後，透過Fuzzy C–Mean對目標集合進行分群，可以得到個群中心以及標準差。



1. 多層式箭靶。箭靶可一次接收多個輸入，在每個箭靶之後與T-S神經元相接。

Step 3: 由於本研究採用球式複數模糊集合，箭靶又會承接上一層的輸入，故每個箭靶都會有許多層，以接收輸入向量中的每個值，如圖(Fig. 4.)所示。其中，第層的第群中心標記為}，第層的第群標準差標記為{, }。用得到的群中心以及標準差製作箭靶，箭靶的製作以及詳細公式將會在下一小節討論。每個箭靶神經元後面連接著一個T–S神經元，T–S神經元為T–S function構成，T–S function公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，{}是第個T–S神經元的參數，是第個輸入。

在結構學習後，我們可以透過分群確認第1層的神經元個數，藉由區塊挑選得到個第2層神經元，最後透過分群可得到個第3層神經元和第4層神經元，藉此創建模型，模型詳細說明將在下個小節探討。

## Model Structure and I/O Relationship

本研究的模型為一個六層的類神經網路。訓練資料集合標記為，為資料總筆數，是*-*by*-*1的輸入向量，為輸入維度數量;為-by-1的目標向量，為複數型態目標的數量。透過模型可以得到輸出。

**Layer 0:** 此層為輸入層，是原始資料透過多目標特徵選取後，將最後挑出的特徵當作訓練資料，我們將時間序列第個點的輸入向量標記如下。

|  |  |  |
| --- | --- | --- |
|  |  |  |

**Layer 1:** 此層為SCFS神經層，透過前面結構學習的分群，可在不同維度上建構數個模糊集，每個不同維度的輸入都可經由模糊集得到歸屬程度。透過球式複數模糊集合可得到多組複數型態的歸屬程度，不同的歸屬程度可以給不同的模型輸出做應用，以達到多目標預測的效果，透過球式複數模糊集的公式(1)-(7)，從中可得到歸屬度向量，如下。

|  |  |  |
| --- | --- | --- |
|  |  |  |

**Layer 2:** 此層為前提式神經層，經過前面的結構學習，我們可以篩選出個對模型較有用的區塊並建構神經元，由於神經元的輸入為上一層的歸屬程度，且輸出為每個輸入維度的歸屬程度相乘結果，故稱之為神經元，每個神經元的輸出為該神經元的啟動強度，由於本研究採用球式複數模糊集合，因此每個神經元的輸入會是向量型態，輸出亦然如此。

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |

，為的乘積，為第個神經元中第個維度歸屬度向量的第項歸屬程度，。

**Layer 3:** 箭靶是用以承接上一層的輸出，為向量的型態，本層的輸出亦是向量的型態，如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，為第個神經元射在第個箭靶第層的值，。本研究中使用的是球式複數模糊集，因此輸入會是複數型態，故輸出也要在複數單位圓盤中，因此箭靶需進行轉換，以確保箭靶層輸出也是複數型態，如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |
|  | , |  |

其中，。為第個箭靶轉換後的中心，為第個箭靶轉換後的靶寬，公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |
|  | , |  |
|  | , |  |
|  | , |  |
|  | , |  |

其中，使用作為輸入;使用作為輸入;為第個目標數據的平均;為第個目標數據的標準差。

**Layer 4:** 此層為T–S層，經過此層的運算可以得到個模型輸出，公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |

其中，為第個輸出；{}是第個T–S神經元的參數。

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |
|  | , |  |
|  | , |  |
|  | , |  |
|  |  |  |

**Layer**此層為輸出層，將上一層得到的個模型輸出加總，即為我們的模型輸出。

|  |  |  |
| --- | --- | --- |
|  |  |  |

## Parameter Learning

根據分治法(Divide-and-conquer)的概念，我們將使用不同的機器學習演算法，對各層的參數優化，以便更容易找到最佳解。對於第1層模糊集的參數優化，我們使用知名的粒子群演算法(Particle swarm optimization, PSO)演算法[19]學習，其原理為模擬鳥群在尋找食物，每回合透過自身的最佳位置和全群最佳位置調節速度，特性為收斂快速，演算法公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |

其中，為第回合時第個粒子的位置，為第回合時第個粒子的速度，為第回合時第個粒子的最好位置，為第回合時全部粒子中最好的位置，為PSO的參數，、為介於0到1的隨機數。在本研究中，粒子的位置代表模糊集中的參數，其中包含了每個維度的分群中心、標準差以及相位頻率函數、。

在本論文使用遞迴式最小平方演算法(Recursive least square estimation, RLSE)[17]更新T–S神經元參數，RLSE方法在更新參數時是利用每筆資料，不斷的更新比起一次接收所有資料的LSE方法更加有效，一般來說LSE問題可以被視為一個線性的問題，如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，y是目標;u是模型的輸出;{}是u已知的方程式;{, =1,2,…,m}是我們估計的未知參數，則是整個模型的誤差LSE的問題也可以被寫成矩陣的方式表達，如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |
|  | , |  |
|  | , |  |

是輸入的矩陣，是我們估計的未知參數矩陣，是目標矩陣，是誤差的向量。要最佳化，可透過RLSE的等式[17]運算。

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |

其中，是遞迴次數，{}，為資料總筆數，是的第行，再開始RLSE演算法時，會設定為0，則設定為，為一極大整數，為單位矩陣。

PSO–RLSE混合演算法的流程如下:

Step 1: 準備訓練資料及測試資料。

Step 2: 以PSO粒子位置作為模糊集參數，將訓練資料帶入模型，並計算每個神經元啟動強度。

Step 3: 用RLSE更新神經元的參數，RLSE算式中的和向量如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |
|  | , |  |
|  | , |  |

其中。因為多目標預測中，啟動強度為一向量，且為一矩陣，因此在原本的公式(62)中，利用單位矩陣取代原本的常數項1，改良後公式如下。

|  |  |  |
| --- | --- | --- |
|  |  |  |

Step 4: 更新完所有參數後，計算出模型的輸出。

Step 5: 計算成本，更新PSO粒子自身的最佳位置和全群最佳位置。本研究成本函數(Cost function)使用均方根誤差(Root mean square error, RMSE)，定義如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，為模型第筆資料的誤差;為埃爾米特共軛(Hermitian transpose)，意即轉置矩陣後，並對矩陣元素做共軛運算。

Step 6: 對所有PSO粒子重複Step 2~Step 5，直到PSO迭代結束。

# Experimentation

本研究總共有3個實驗，每個實驗目標相關性高，如實驗一開盤價及收盤價的關係；實驗二為指標型指數，其中，台股加權指數(The Taiwan Stock Exchange Capitalization Weighted Stock Index, TAIEX)為台灣上市的股票中經過加權計算出的指標，代表著台灣上市股票的波動，道瓊工業指數(Dow Jones Industrial Average Index, DJI)涵蓋著財務等9大產業，為一股價加權指標，納斯達克(National Association of Securities Dealers Automated Quotation, NASDAQ)為超過三千檔股票所組合成的市值加權指標，大多以科技產業為例，標準普爾500(Standard and Poor’s ,S&P 500)為美國前500大公司的市值加權，當中包含IT等11個產業。上述皆為非常著名的股票指標，故能精準預測可帶來一定的效用。第三個實驗使用的目標多為大規模的科技公司，如IBM、APPLE、DELL、Microsoft等。為了與其他文獻比較模型的好壞，我們將透過兩種指標對模型評估。前兩個實驗使用RMSE做為評估指標，公式如(69)，第3個實驗，基於其他文獻所提供的數據資料，我們將以平均絕對誤差百分比(Mean average percent errors, MAPE)做為評估指標，MAPE公式如下。

|  |  |  |
| --- | --- | --- |
|  | , |  |

其中，為資料總筆數，為第筆真實數據值，為第筆模型輸出。

在結構學系部分，本研究中所有實驗的上限皆設定為4，下限皆設定為2，故模型輸入維度會介於2~4之間。

## Example 1—Quadruple Time Series of Daily National Association of Securities Dealers Automated Quotation Composite Index

在這個實驗中，我們使用真實世界的時間序列數據來驗證模型的效能。使用的資料為納斯達克(National Association of Securities Dealers Automated Quotation, NASDAQ) 2007年1月3號至2010年12月20號每天的開盤和收盤價以及標準普爾500(Standard and Poor’s ,S&P 500) 2007年1月3號至2010年12月20號每天的開盤和收盤價。此實驗的原始資料為1029筆，經過一次差分得到1028筆，並以4組差分資料取出30個特徵，共有120個特徵，每個特徵會有998筆資料，前500筆資料為訓練資料，剩下的為測試資料。第1至30個特徵為NASDAQ開盤價，第31至60個特徵為NASDAQ收盤價，第61至90個特徵為S&P 500 開盤價，第91至120個特徵為S&P 500收盤價，這120個特徵與目標形成資料矩陣，資料矩陣中以S&P500距離目標最近，目標排序為NASDAQ開盤價、NASDAQ收盤價、S&P500開盤價、S&P500收盤價。資料矩陣經過多目標特徵挑選[22]後，選出特徵作為模型輸入，複數型態目標有兩個，第一組複數型態的目標實部部分為NASDAQ開盤價，虛部部分為NASDAQ收盤價，第二組複數型態的目標實部部分為S&P500開盤價，虛部部分為S&P500收盤價。結構學習部分，則將每個被挑選的特徵，用減數分群演算法[11]分群。並透過第二章所介紹的神經元挑選方法，從原本的54個神經元篩選到剩下8個神經元。整體模型在結構學習後的參數，如表I所示。PSO–RLSE混合方法的機器學習參數設定，如表II所示。本篇論文所提出的模型可以一次有四個複數型態的輸出，故可以預測四個複數型態目標，在此實驗中只同時預測兩個目標。

實驗一模型設定

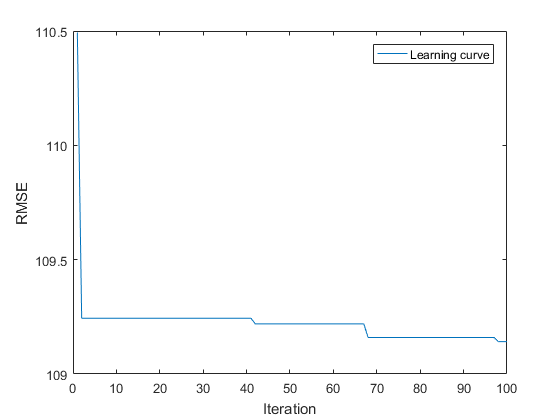
|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Feature variables as model inputs | {} |
| Number of input fuzzy sets | {2, 3, 3, 3} |
| Type of fuzzy sets | SCFS |
| Number of complex–valued targets\* | 2 |
| Number of neurons | 8 |
| Number of parameters in the SCFS layer | 44 |
| Number of aim object neurons | 5 |
| Number of T–S neurons | 5 |
| Number of parameters in the T–S layer | 25 |

\* Each complex-valued target whose real and imaginary parts contain two real-valued targets, respectively.

機器學習參數設定

|  |  |
| --- | --- |
| **PSO** | |
| Swarm size | 64 |
| Iterations | 100 |
|  | {0.8, 2.0, 2.0} |
|  | Random in [0,1] |
| Initial particle positions | By SC algorithm in section II-C |
| Initial particle velocities | 0 |
| **RLSE** | |
|  |  |
|  | 25-by-1 zero vector |
|  | **I** |
| **I** | 25-by-25 identify matrix |

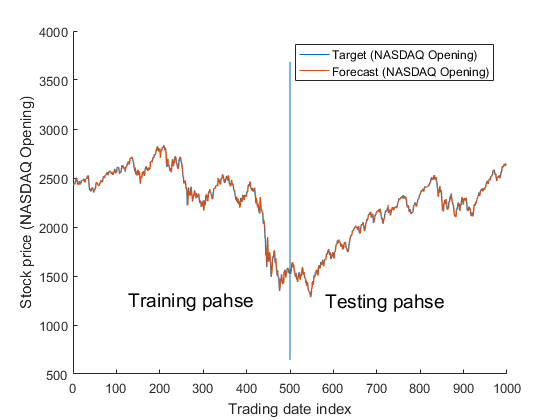
本次實驗的結果將與其他文獻[23]所提的方法做比較。其中除了SVR之外的模型，皆可以同時預測兩個實數目標。所以我們將使用模型的第一組複數型態的輸出與其他文獻做比較，結果如表III所示，十次重複實驗效能統計如表IV所示。模型的學習曲線，如Fig. 5所示，可看出在前10個迭代時，學習逐漸穩定；目標與模型輸出的結果，如Fig. 6所示，有著不錯的效果；預測誤差如Fig. 7所示，誤差呈現亂數狀態，範圍介於-50至50，代表模型預測能力穩定。



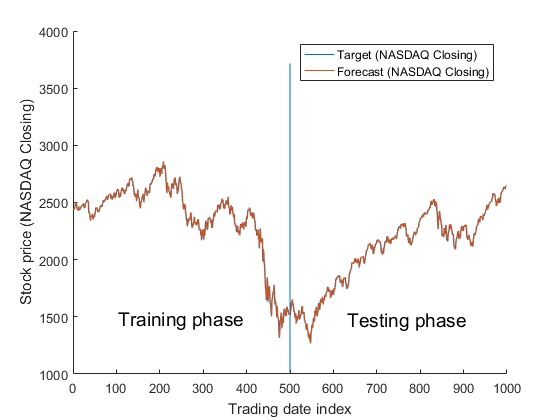
1. 學習曲線。可從此曲線看出RMSE的變化量，在前10迭代時，學習逐漸穩定。(實驗一)

效能比較(NASDAQ, 實驗一)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **RMSE** | | | | |
| Training phase |  |  | Testing phase |  |
| Opening index | Closing index |  | Opening index | Closing index |
| SVR (two models, each with single output) [23] | 35.18 | 35.24 |  | 37.23 | 40.24 |
| ANFIS (two models, each with single output) [23] | 37.83 | 38.66 |  | 38.80 | 42.36 |
| ANFIS (one model with two outputs) [23] | 62.75 | 71.51 |  | 72.52 | 85.08 |
| RBF (two models, each with single output) [23] | 37.59 | 33.89 |  | 37.52 | 44.08 |
| RBF (one model with two outputs) [23] | 178.57 | 179.87 |  | 261.37 | 258.89 |
| CNFS(5)-ARIMA (one model with two outputs) [23] | 21.56 | 20.81 |  | 32.52 | 33.70 |
| SCNFS(proposed) | 38.58 | 38.59 |  | 27.73 | 27.44 |

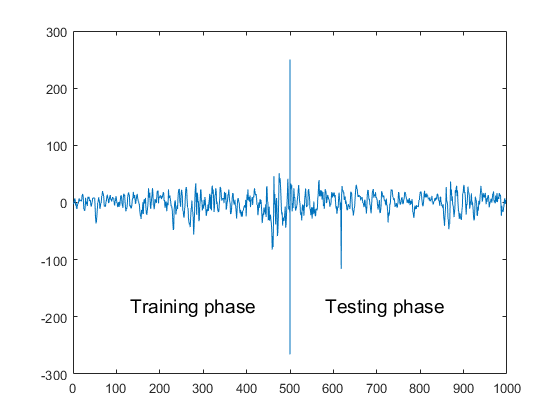


(a)



(b)

1. NASDAQ的實際數值和模型輸出(a)每日開盤價(b)每日收盤價。x軸為交易日，y軸為股票價格，可看出訓練和測試階段皆有不錯的效果。(實驗一)



1. 預測誤差。誤差呈現亂數狀態，範圍介於-50至50，代表模型預測能力穩定。(實驗一)

十次重複實驗效能統計(實驗一)

|  |  |  |
| --- | --- | --- |
|  | Performance (RMSE) | |
| Trials | Opening index | Closing index |
| 1 | 27.67 | 27.75 |
| 2 | 29.62 | 28.04 |
| 3 | 33.85 | 33.99 |
| 4 | 28.09 | 28.77 |
| **5** | **27.73** | **27.44** |
| 6 | 33.32 | 42.27 |
| 7 | 27.68 | 27.61 |
| 8 | 29.35 | 27.93 |
| 9 | 27.67 | 27.91 |
| 10 | 27.72 | 27.76 |

## Example 2—Quadruple Time Series of Daily Dow Jones Industrial Average Index

在這個實驗中，我們也是使用真實世界的時間序列數據來驗證模型的效能，與實驗一不同的是，四個目標之間不是收盤價與開盤價的關係，意即目標的曲線相似度沒有像實驗一目標曲線的相似度高。使用的資料為台股加權指數(The Taiwan Stock Exchange Capitalization Weighted Stock Index, TAIEX)、道瓊工業指數(Dow Jones Industrial Average Index, DJI)、納斯達克(National Association of Securities Dealers Automated Quotation, NASDAQ)以及標準普爾500(Standard and Poor’s ,S&P 500)，資料區間為2001年至2004年的收盤價，每年收盤價筆數如下，2001年資料為245筆，2002年資料為248筆，2003年資料為249筆，2004年資料為250筆，每年的做一次模型預測，前十個月的資料當作訓練資料，剩餘的當作測試資料。每年四組原始資料會做一次差分，並以4組差分資料取出30個特徵，共有120個特徵。第1至30個特徵為TAIEX收盤價，第31至60個特徵為DJI收盤價，第61至90個特徵為NASDAQ收盤價，第91至120個特徵為S&P 500收盤價，這120個特徵與目標形成資料矩陣，資料矩陣中以S&P500距離目標最近，目標排序為TAIEX收盤價、DJI收盤價、NASDAQ收盤價、S&P500收盤價。資料矩陣經過多目標特徵挑選[22]後，將選出的特徵作為模型輸入，複數型態目標有兩個，第一組複數型態的目標實部部分為TAIEX收盤價，虛部部分為DJI收盤價，第二組複數型態的目標實部部分為NASDAQ收盤價，虛部部分為S&P500收盤價。結構學習部分，則將每個被挑選的特徵，用減數分群演算法[11]分群。並透過第二章所介紹的神經元挑選方法，減少神經元數目。整體模型在結構學習後的參數，如表V-表VIII所示。PSO–RLSE混合方法的機器學習參數設定，如表IX所示。本篇論文所提出的模型可以一次有四個複數型態的輸出，故可以預測四個複數型態目標，但在此實驗中只同時預測兩個目標。

實驗二模型設定(2001年)

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Feature variables as model inputs | {} |
| Number of input fuzzy sets | {3, 4, 3, 4} |
| Type of fuzzy sets | SCFS |
| Number of complex–valued targets\* | 2 |
| Number of neurons | 15 |
| Number of parameters in the SCFS layer | 56 |
| Number of aim object neurons | 3 |
| Number of T–S neurons | 3 |
| Number of parameters in the T–S layer | 15 |

\* Each complex-valued target whose real and imaginary parts contain two real-valued targets, respectively.

實驗二模型設定(2002年)

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Feature variables as model inputs | {} |
| Number of input fuzzy sets | {3, 3, 3, 3} |
| Type of fuzzy sets | SCFS |
| Number of complex–valued targets\* | 2 |
| Number of neurons | 13 |
| Number of parameters in the SCFS layer | 48 |
| Number of aim object neurons | 5 |
| Number of T–S neurons | 5 |
| Number of parameters in the T–S layer | 25 |

\* Each complex-valued target whose real and imaginary parts contain two real-valued targets, respectively.

實驗二模型設定(2003年)

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Feature variables as model inputs | {} |
| Number of input fuzzy sets | {3, 3, 3, 3} |
| Type of fuzzy sets | SCFS |
| Number of complex–valued targets\* | 2 |
| Number of neurons | 11 |
| Number of parameters in the SCFS layer | 48 |
| Number of aim object neurons | 4 |
| Number of T–S neurons | 4 |
| Number of parameters in the T–S layer | 20 |

\* Each complex-valued target whose real and imaginary parts contain two real-valued targets, respectively.

實驗二模型設定(2004年)

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Feature variables as model inputs | {} |
| Number of input fuzzy sets | {5, 4, 4, 4} |
| Type of fuzzy sets | SCFS |
| Number of complex–valued targets\* | 2 |
| Number of neurons | 15 |
| Number of parameters in the SCFS layer | 68 |
| Number of aim object neurons | 8 |
| Number of T–S neurons | 8 |
| Number of parameters in the T–S layer | 40 |

\* Each complex-valued target whose real and imaginary parts contain two real-valued targets, respectively.

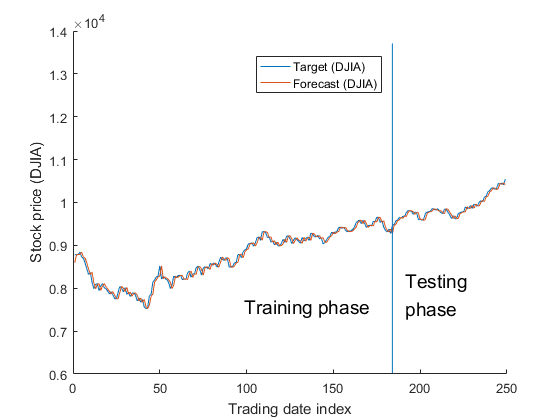
機器學習參數設定

|  |  |
| --- | --- |
| **PSO** | |
| Swarm size | 64 |
| Iterations | 100 |
|  | {0.8, 2.0, 2.0} |
|  | Random in [0,1] |
| Initial particle positions | By SC algorithm in section II-C |
| Initial particle velocities | 0 |
| **RLSE** | |
|  |  |
|  | By number of parameters in the T–S layer |
|  | -by-1 zero vector |
|  | **I** |
| **I** | -by- identify matrix |

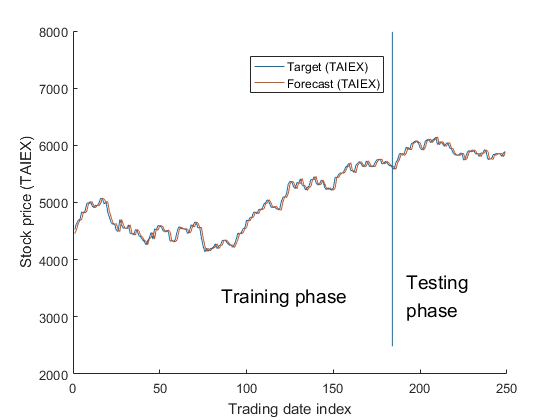
本次實驗的結果將與其他文獻[23]所提的方法做比較。其中除了SVR之外的模型，皆可以同時預測兩個實數目標。所以我們將使用模型的第一組複數型態的輸出與其他論文做比較，結果如表XI和表XII所示，十次重複實驗效能統計如表X所示。目標與模型輸出的結果，如Fig. 8所示，有著不錯的效果；預測誤差如Fig. 9所示，誤差呈現亂數狀態，範圍介於-100至100，代表模型預測能力穩定。

十次重複實驗效能統計(實驗二)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Performance (RMSE) | | | |
| Trials | 2001 | 2002 | 2003 | 2004 |
| 1 | 259.74 | 279.93 | **196.91** | 282.82 |
| 2 | **259.27** | 286.02 | 198.10 | 283.94 |
| 3 | 260.99 | 280.34 | 199.06 | 286.83 |
| 4 | 260.44 | 280.56 | 199.54 | 284.57 |
| 5 | 263.29 | 281.85 | 196.94 | 278.60 |
| 6 | 263.57 | 282.00 | 197.85 | 288.52 |
| 7 | 259.86 | **279.46** | 200.05 | **275.31** |
| 8 | 262.02 | 282.35 | 198.33 | 285.69 |
| 9 | 261.23 | 282.11 | 197.53 | 288.15 |
| 10 | 260.59 | 280.93 | 197.62 | 286.68 |

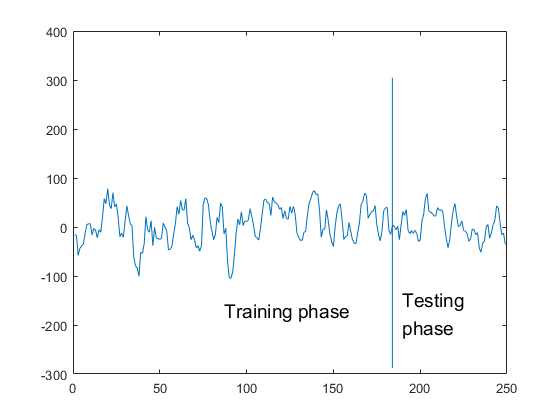


(a)



(b)

1. 實際數值和模型輸出(a)DJIA (2003年) (b)TAIEX (2003年)。x軸為交易日，y軸為股票價格，可看出訓練和測試階段皆有著不錯的效果。(實驗二)



1. 預測誤差。誤差呈現亂數狀態，範圍介於-100至100，代表模型預測能力穩定。(實驗二)

效能比較(DJIA, 實驗二)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **RMSE** | | | |
| **Method** | **2001** | **2002** | **2003** | **2004** |
| SVR (two models, each with single output) [23] | 101.44 | 117.95 | 82.76 | 71.49 |
| ANFIS (two models, each with single output) [23] | 105.56 | 111.69 | 72.09 | 68.00 |
| ANFIS (one model with two outputs) [23] | 128.20 | 142.05 | 90.37 | 83.69 |
| RBF (two models, each with single output) [23] | 106.33 | 131.24 | 97.58 | 81.79 |
| RBF (one model with two outputs) [23] | 181.79 | 136.28 | 154.14 | 148.11 |
| CNFS(5)-ARIMA (one model with two outputs) [23] | 103.06 | 103.42 | 70.70 | 66.55 |
| SCNFS(proposed) training phase | 91.95 | 98.69 | 69.66 | 97.99 |
| SCNFS(proposed) testing phase | 94.31 | 85.52 | 56.26 | 61.64 |

效能比較(TAIEX, 實驗二)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **RMSE** | | | |
| **Method** | **2001** | **2002** | **2003** | **2004** |
| SVR (two models, each with single output) [23] | 162.46 | 67.72 | 59.47 | 58.81 |
| ANFIS (two models, each with single output) [23] | 147.36 | 70.17 | 72.61 | 65.33 |
| ANFIS (one model with two outputs) [23] | 151.62 | 78.27 | 81.69 | 70.54 |
| RBF (two models, each with single output) [23] | 134.32 | 65.15 | 60.41 | 102.86 |
| RBF (one model with two outputs) [23] | 137.58 | 78.54 | 115.92 | 126.48 |
| CNFS(5)-ARIMA (one model with two outputs) [23] | 115.82 | 64.34 | 57.69 | 55.56 |
| SCNFS(proposed) training phase | 92.03 | 100.26 | 69.96 | 99.11 |
| SCNFS(proposed) testing phase | 89.59 | 86.81 | 55.34 | 60.30 |

## Example 3—Quadruple Time Series of Daily Taiwan Stock Exchange Capitalization Weighted Stock Index

在這個實驗中，我們使用真實世界的時間序列數據來驗證模型的效能。使用的資料為蘋果電腦(Apple computer inc.)、IBM(International business machines corporation)、戴爾(Dell inc.)、微軟(Microsoft inc.) 四個股票的收盤價，期間為2003年2月10號至2005年1月21號。此實驗的原始資料為492筆，經過一次差分得到491筆，並以4組差分資料取出30個特徵，共有120個特徵，每個特徵會有460筆資料，前433筆資料為訓練資料，剩下的為測試資料。第1至30個特徵為APPLE收盤價，第31至60個特徵為IBM收盤價，第61至90個特徵為Dell收盤價，第91至120個特徵為Microsoft收盤價，這120個特徵與目標形成資料矩陣，資料矩陣中以Microsoft距離目標最近，目標排序為APPLE、IBM、Dell和Microsoft。資料矩陣經過多目標特徵挑選[22]後，選出特徵作為模型輸入，複數型態目標有兩個，第一組複數型態的目標實部部分為APPLE收盤價，虛部部分為IBM收盤價，第二組複數型態的目標實部部分為Dell收盤價，虛部部分為Microsoft收盤價。結構學習部分，則將每個被挑選的特徵，用減數分群演算法[11]分群。並透過第二章所介紹的π神經元挑選方法，從原本的81個π神經元篩選到剩下13個π神經元。整體模型在結構學習後的參數，如表XIII所示。PSO–RLSE混合方法的機器學習參數設定，如表XIV所示。本篇論文所提出的模型可以一次有四個複數型態的輸出，故可以預測四個複數型態目標，在此實驗中只同時預測兩個目標。

實驗三模型設定

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Feature variables as model inputs | {} |
| Number of input fuzzy sets | {3, 3, 3, 3} |
| Type of fuzzy sets | SCFS |
| Number of complex–valued targets\* | 2 |
| Number of neurons | 13 |
| Number of parameters in the SCFS layer | 48 |
| Number of aim object neurons | 3 |
| Number of T–S neurons | 3 |
| Number of parameters in the T–S layer | 15 |

\* Each complex-valued target whose real and imaginary parts contain two real-valued targets, respectively.

機器學習參數設定

|  |  |
| --- | --- |
| **PSO** | |
| Swarm size | 64 |
| Iterations | 100 |
|  | {0.8, 2.0, 2.0} |
|  | Random in [0,1] |
| Initial particle positions | By SC algorithm in section II-C |
| Initial particle velocities | 0 |
| **RLSE** | |
|  |  |
|  | 15-by-1 zero vector |
|  | **I** |
| **I** | 15-by-15 identify matrix |

本次實驗的結果將與其他論文所提的方法做比較，像是HiMMI[13]、ANN-GA-HMM-Interpolation[13]、ANN-GA-HMM-WA[13]、ARIMA[41]、Bayesian ANN[41]。所以我們將使用模型的第一組輸出的實數及複數部分和第二輸出的實數部份與其他論文做比較，結果如表XVI所示，十次重複實驗效能統計如表X所示。模型的學習曲線，如Fig. 10所示，可看出於45迭代時，學習逐漸穩定；目標與模型輸出的結果，如Fig. 11所示，有著不錯的效果，APPLE股價介於5美元左右，故預測曲線看起來有波動;預測誤差如Fig. 12所示，誤差呈現亂數狀態，範圍介於-2至2，代表模型預測能力穩定。



1. 學習曲線。由此圖可看出RMSE於45迭代時，模型學習已逐漸穩定。(實驗三)

十次重複實驗效能統計(實驗三)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Performance (MAPE) | | | |
| Trials | APPLE | IBM | DELL | Microsoft |
| 1 | 1.8909 | 0.8156 | 0.6173 | 0.7007 |
| 2 | 2.3692 | 0.8286 | 0.5957 | 0.7111 |
| 3 | **1.8453** | **0.8051** | **0.6187** | **0.8591** |
| 4 | 2.0273 | 0.8172 | 0.6184 | 0.7048 |
| 5 | 2.0476 | 0.8183 | 0.6100 | 0.7099 |
| 6 | 2.2186 | 0.8005 | 0.6270 | 0.7399 |
| 7 | 2.0239 | 0.8313 | 0.6219 | 0.7128 |
| 8 | 2.2186 | 0.8005 | 0.6270 | 0.7399 |
| 9 | 2.0239 | 0.8313 | 0.6219 | 0.7128 |
| 10 | 1.8636 | 0.8159 | 0.6159 | 0.7045 |



(a)



(b)



(c)

1. DJI的實際數值和模型輸出(a) IBM收盤價 (b) APPLE收盤價 (c) DELL收盤價。x軸為交易日，y軸為股票價格，可看出訓練和測試階段皆有不錯的效果。(實驗三)



1. 預測誤差。誤差呈現亂數狀態，範圍介於-2至2，代表模型預測能力穩定。(實驗三)

效能比較 (實驗三)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **MAPE** | | | | |
| **Method** |  | **APPLE** |  | **IBM** |  | **DELL** |
| HiMMI[13] |  | 2.8373 |  | 1.2186 |  | 1.0117 |
| ANN-GA-HMM-Interpolation[13] |  | 2.1649 |  | 1.0555 |  | 0.8446 |
| ANN-GA-HMM-WA[13] |  | 1.9247 |  | 0.8487 |  | 0.6992 |
| Bayesian ANN[41] |  | 1.9688 |  | 0.7441 |  | - |
| ARIMA[41] |  | 1.8009 |  | 0.9723 |  | 0.6604 |
| SCNFS(proposed) training phase |  | 2.4175 |  | 1.1758 |  | 0.8955 |
| SCNFS(proposed) testing phase |  | 1.8453 |  | 0.8051 |  | 0.6187 |

# Discussion

本研究提出以PSO-RLSE混合演算法優化模型參數，並且在訓練資料進入模型前，事先進行特徵選取，選出對預測目標較為有利之特徵資料進行時間序列之預測。多目標特徵挑選部分，引用夏農資訊熵[39]的理論。計算出資料特徵對於各個目標資料所能提供的資訊量，並考慮與已被選取的特徵的冗餘資訊量，選取最大增益量，代表最後該特徵被選取後能夠提供給目標資訊量，此外使用者可透過第二次挑選，選擇要進入模型的特徵數量，以提升整體效率，減少模型對於太多目標所產生的預測負擔。結構學習的過程，透過資料密度量的概念，選擇出較重要的神經元並建構，使用者可透過上下界設定，從挑選後的神經元中選擇個數，此方法減少了大量的神經元數量，同時也降低了運算所需的時間，且可針對不同的輸入資料形成不同的模型大小，提升模型適應性能力。球式複數模糊類神經系統結合球式複數模糊集合以及T-S模糊系統建立。T-S模糊系統能處理較為模糊的資訊，以一種非線性的方式描述輸入資料的強度，並且以線性規則代表模型輸出，使其系統可以較為人類所理解。球式複數模糊集合，提供了最少四組複數型態歸屬程度，比起傳統模糊集合的單目標預測，可以使模型同時預測八個目標，此外亦可將複數型態值解構，同時預測更多目標，有著較高的延展性及資料豐富性。參數學習部分，利用PSO演算法結合RLSE分別對模型的兩部分參數優化，PSO具有三種特性，第一，可隨著群體智慧自動調整步伐；第二，更新速度時具有隨機參數參與其中，可增進粒子活躍性；第三，會跟著一定的正確方向活動，具備著篤定性。但有著搜尋維度較低的缺點，故使用分治法，配合RLSE進行參數優化。RLSE利用參考前次計算結果，尋找線性函數，使資料點與該函數的平方誤差達到最小值，藉由此方法不斷的遞迴，使得模型輸出近似目標，最佳化參數。

根據3個實驗的測試後，顯示本研究的研究方法對於時間序列數據擁有預測能力。透過模型參數表格可發現，神經元從近百位數下降至個位數，很顯著的控制模型大小。本論文提出的模型SCFNS確實有多目標預測的能力，並可以發現透過一次對四個目標做時間序列的預測，各個目標的效果不亞於其他論文所提出的方法，甚至更好。PSO-RLSE的混合方法，有著一定的水準。機器學習的部分可能會受限於PSO本身的特性，收斂快速、較容易掉入區域最佳解，從Fig. 2中不難發現，PSO演算法在前幾個回合就已經接近最後一回合的RMSE，故在資料量龐大的情況下，所需搜尋維度增多，PSO效能就會較差，可能會限制住模型的整體發揮。

本研究提出的一系列研究方法，可以有效預測多個目標，證明球式複數模糊集分解的成分確實增加了資訊量的豐富度，且可藉由資料數據本身配合使用者設定控制模型大小，提升模型運算速度。機器學習除了結構學習，也應用在參數學習上，PSO-RLSE方法透過分而治之的概念，使得模型遇上多個參數時，仍然有著不錯的效果。

# Concluding Remarks

本研究提出POS-RLSE複合演算法，結合PSO演算法以及RLSE，用於優化球式複數型態模糊類神經模型系統的參數集合。模型採用球式複數模糊集合、T-S系統以及類神經網路概念。資料前處理以特徵選取，減少資料對模型的過多負擔。本研究根據夏農資訊熵[39]的觀念，開創一種對於多目標特徵選取的方式。讓輸入資料在進入模型前，針對該資料所產生的候選特徵進行篩選。避免冗餘的輸入資料進入模型，耗費運算效能。球式複數模糊集合，使模型具多個複數型態輸出，其複數型態的歸屬程度，讓模型可以有複數值的輸出。有別於一般模糊集合，提升資訊量的豐富度，讓模型能夠有同時進行多目標預測的能力。三個實驗可證明此方式的貢獻。模型系統的參數決定預測結果的優劣，PSO-RLSE複合演算法，透過分治法概念，透過不同機器學習演算法學習不同區域參數，將問題最小化，分而治之，降低演算法搜尋維度，並增進模型整體效能。

PSO演算法仍然有著維度過小，早熟(premature)的缺點，亦即在短時間內落入區域最佳解，在未來，可以透過不同的機器學習演算法結合SCFNS，像是隨機優化演算法 (Random Optimization, RO)[45]、合作型粒子群演算法 (Cooperative Particle Swarm Optimization, CPSO)[4]、螞蟻演算法 (Ant Colony Optimization, ACO)[8]等，或許能克服目前PSO所遇到的問題，做出更精準的預測。近年來深度學習的發展快速，六層式神經網路相對其他文獻來說算是較小的架構，未來可擴充更多的類神經網路隱藏層，透過更複雜的運算，嘗試尋找出更優異的預測效果。

##### Acknowledgment

This work was supported by the research projects MOST 105–2221–E–008–091 and MOST 104–2221–E–008–116, Ministry of Science & Technology, Taiwan.

##### References

1. M. H. Alobaidi, F. Chebana and M. A. Megiud, “Robust ensemble learning framework for day-ahead forecasting of household based energy consumption,” Applied Energy, vol. 212, pp. 997-1012, Feb. 2018.
2. P. R. L. Alves, L. G. S. Duarte and L.A.C.P. da Mota, “A new characterization of chaos from aa time series,” Chaos, Solitons and Fractals, vol. 104, pp. 323-326, Aug. 2017.
3. P. Aboagye–Sarfo, Q. Mai, F. M. Sanfilippo, D. B. Preen, L. M. Stewart and D. M. Fatovich, “A comparison of multivariate and univariate time series approaches to modeling and forecasting emergency department demand in Western Australia,” Journal of Biomedical Informatics, vol. 57, pp. 62–73, Oct. 2015.
4. F. van den Bergh and A. P. Engelbrecht, “A cooperative approach to particle swarm optimization,” IEEE Transactions on Evolutionary Computation, vol. 8, no. 3, pp. 225–239, Jun. 2004.
5. J. J. Buckley and Y. Qu, “Fuzzy complex analysis–II: Integration,” Fuzzy Sets System, vol. 49, no. 2, pp. 171–179, 1992.
6. J. J. Buckley and Y. Qu, “Fuzzy complex analysis–I: Differentiation,” Fuzzy Sets System, vol. 41, no. 3, pp. 269–284, 1991.
7. J. J. Buckley, “Fuzzy complex numbers,” Fuzzy Sets System, vol. 33, no. 3, pp. 333–345, 1989.
8. A. Colorni, M. Dorigo and V. Maniezzo, “Distributed optimization by ant colonies,” Proceedings of the 1st European Conference on Artificial Life, pp. 134-142, 1991.
9. H. d. N. Camelo, P. S. Lucio, J. B. V. L. Junior, P. C. M. d. Carvalho and D. v. G. d. Santos, “Innovative hybrid models for forecasting time series applied in wind generation based on the combination of time series models with artificial neural networks,” Energy, in press.
10. C. Cortes and V. Vapnik, “Support-Vector Networks,” Machine Learning, vol. 20, pp. 273-297, 1995.
11. S. Chiu, “Fuzzy model identification based on cluster estimation,” Journal of Intelligent and Fuzzy Systems, pp. 267–278 Jan. 1994.
12. R. Clausius, “Ueber eine veränderte Form des zweiten Hauptsatzes der mechanischen Wärmetheorie,” Annalen der Physik und Chemie, vol. 93, no. 12, pp. 481–506, 1854.
13. R. Hassan, B. Nath and M. Kirley, “A fusion model of HMM, ANN and GA for stock market forecasting,” ELSEVIER. Expert Systems with Applications, vol. 33, no. 1, pp. 171-180, Jul. 2007.
14. L. J. Herrera, H. Pomares, I. Rojas, A. Guillen, J. Gonzalez, M. Awad and A. Herrera, “Multigrid–based fuzzy systems for time series prediction:CATS competition,” Neurocomputing, vol. 70, no. 13–15, pp. 2410–2425, Aug. 2007.
15. D. Hebb: *The Organization of Behavior*, 1949.
16. S. S. Jones, R. S. Evans, T. L. Allen, A. Thomas, P. J. Haug, S. J. Welch and G. L. Snow, “A multivariate time series approach to modeling and forecasting demand in the emergency department,” Journal of Biomedical Informatics, vol. 42, no. 1, pp. 123–139, Feb. 2009.
17. J. S. R. Jang, C. T. Sun and E. Mizutani, “Neuro–fuzzy and soft computing: A computational approach to learning and machine intelligence,” IEEE Transactions on Automatic Control, vol. 42, no. 10, Oct. 1997.
18. J. S. R. Jang, “ANFIS: Adaptive–network–based fuzzy inference system,” IEEE Transactions on System, Man, and Cybernetics, vol. 23, no. 3, pp. 665–685, May/Jun, 1993.
19. J. Kennedy and R. Eberhart, “Particle swarm optimization,” Proceedings of IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948, Nov.–Dec. 1995.
20. R. S. J. Koijen, H. Lustig and S. V. Nieuwerburgh, “The cross-section and time series of stock and bond returns,” Journal of Monetary Economics, vol. 88, pp. 50-69, Jun. 2017.
21. M. Z.–Kermani and M. Teshnehlab, “Using adaptive neuro–fuzzy inference system for hydrological time series prediction.” Applied Soft Computing, vol. 8, no. 2, pp. 928–936, Mar. 2008.
22. C. Li, “Multi–target feature selection,” research training material for graduate students in the dept. of Information Management, National Central University, Taiwan, unpublished, 2017.
23. C. Li and T.–W. Chiang, “Complex neurofuzzy ARIMA forecasting—a new approach using complex fuzzy sets,” IEEE Transtraction on fuzzy systems. vol. 21, no. 3, pp.567–584, Jun. 2013.
24. C. Li and T.–W. Chiang, “Complex fuzzy computing to time series prediction—A multi–swarm PSO learning approach,” Lecture Notes in Computer Science, vol. 6592, pp. 242–251, 2011.
25. C. Li and T.–W. Chiang, “Complex neuro–fuzzy self–learning approach to function approximation,” Lecture Notes in Computer Science, vol. 5991, pp. 289–299, 2010.
26. C. Li, T.–W. Chiang. J.–W. Hu, and T. Wu, “Complex neuro–fuzzy intelligent approach to function approximation,” Advanced Computational Intelligence. 2010 3rd International Workshop on Advanced Computational Intelligence, pp. 151–156, 2010.
27. C. Li and J.–W. Hu, “A new ARIMA–based neuro–fuzzy approach and swarm intelligence for time series forecasting,” Engineering Applications of Artificial Intelligence, vol. 25, no. 2, pp. 295–308, Mar. 2012.
28. W. Liang-Ying, “A hybrid time series model based on AR-EMD and volatility for medical data forecasting: A case study in the emergency department,” International Journal of Management, Economics and Social Sciences, vol. 6, no. Special issue, pp. 166-184, Oct. 2017.
29. K. O. Mikalsen, F. M. Bianchi, C. Soguero-Ruiz and R. Jenssen, “Time series cluster kernel foe learning similarities between multivariate time series with missing data,” Pattern Recognition, vol. 76, pp. 569-581, Apr. 2018.
30. W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” Bulletin of Mathematical Biophysics, vol. 5 1943.
31. D. Osthus, K. S. Hickmann, P. C. Caragea, D. Higdon and S. Y. Valle, “Forecasting seasonal influenza with a state-space SIR model,” The Annals of Applied Statistics, vol. 11, no. 1, pp. 202-224, Sep. 2017.
32. Z. Pan and L. Liu, “Forecasting stock return volatility: A comparison between the roles of short-term and long-term leverage effects,” Physica A: Statistical Mechanics and its Applications, vol. 492, pp. 168-180, Feb. 2018.
33. N. Rochester, J. H. Holland, L. H. Haibt and W. L. Duda, “Test on a cell assembly theory of the action of the brain, using a large digital computer,” IRE Transactions on Information Theory, 1956.
34. D. Ramot, R. Milo, M. Friedman, and A, Kandel, “Complex fuzzy sets,” IEEE Transactions on Fuzzy Systems, vol. 10, no. 2, pp. 171–186, Apr. 2002.
35. H. J. Rong, N. Sundararajan,G. B. Huang and P. Saratchandran, “Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction,” Fuzzy Set Syst, vol.157, no. 9, pp. 1260–1275, May 2006.
36. S. Z. Selim and M. S. Kamel, “On the mathematical and numerical properties of the fuzzy c-means algorithm1,” Fuzzy Sets and Systems, vol. 49, no. 2, pp. 181-191, Jul. 1992.
37. I. Sugiarto and S. Natarajan, “Parameter estimation using least square method for MIMO Takagi–Sugeno neuro–fuzzy in time series forecasting,” Jurnal Teknik Elektro, vol. 7, no. 2, pp. 82–87, Sep. 2007.
38. R. Stuart and N. Peter: *Artificial Intelligence: A Modern Approach 2nd. Prentice Hall*, 1995.
39. E. C. Shannon, “A mathematical theory of communication,” The Bell System Technical Journal, vol. 27, pp. 379–423, Jul. 1948.
40. T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC–15, no. 1, pp. 116–132, Jan.–Feb. 1985.
41. L. Ticknor, “A Bayesian regularized artificial neural network for stock market forecasting,” Expert Systems with Applications, vol. 40, no. 14, Oct. 2013.
42. P. Werbos, “Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences,” PhD thesis, Harvard University, 1974.
43. Y. Zhang and P. Shang, “Refined composite multiscale weighted-permutation entropy of financial time series,” Physica A: Statistical Mechanics and its Application, vol. 496, pp. 189-199, Apr. 2018.
44. L. A. Zadeh, “Fuzzy sets”, Information and Control , vol. 8, no. 3, pp. 338–353, Jun. 1965.
45. Z. B. Zabinsky, “Random search algorithms,” Technical report, University of Washington, Seattle, Apr. 2009.