Paper Title\* **(**use style: ***paper title***)

Subtitle as needed ***(paper subtitle)***

Chi Feng Lina and Chunshien Lib

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

aEmail: j8888888871@yahoo.com.tw

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

*Abstract*— This paper mainly discusses the prediction performance of different algorithms in time series, and conducts performance through simulated investment strategies. Time series prediction is a wide and important issue. This paper proposes a sphere complex neuro-fuzzy system (SCNFS) to predict the time series. A complex neuro-fuzzy system (CNFS) has a complex-valued output, the real part and the imaginary part can predict different target. In this paper, we improve the complex fuzzy sets (CFSs) in CNFS and called it sphere complex fuzzy sets (SCFSs), the membership degrees are also complex-valued, but SCNFS can have multiple targets, which makes the model can predict more than two targets simultaneously. In the part of model design, Gaussian-type SCFSs are used in the premise part, and Takagi-Sugeno linear functions are used in the consequence part, both use the IF-THEN rule type link. In addition, in order to optimize the model output, this paper uses the concept of divide-and-conquer, the premise part respectively uses particle swarm optimization (PSO), artificial bee colony optimization (ABCO) to optimize the parameters, the consequence part uses recursive least squares estimator (RLSE) to optimize the parameters. Finally, three experiments are used to testify performance of different algorithms and combine with the investment strategy proposed by this paper to simulate the investment effect.

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

# Introduction

Artificial intelligence has many application fields in reality. From the 18th and 19th centuries, there are many related researches, but due to hardware equipment problems, the results have not been applied to ordinary life. Nowadays, with the development of hardware devices, artificial intelligence has already brought a lot of help to human beings, and research is even more numerous, like financial and economic forecasts. In 2017, Jardin et al. [7] proposed a quantitative approach to defining a company's health status to predict whether a company is bankrupt; Ravi et al. [26] used chaos theory, multi-layer perceptron and multi-target optimization algorithms to predict exchange rates; Heaton et al. [12] proposed a deep learning model for the financial portfolio. In the picture classification, Feng et al. proposed Discriminative Locality Alignment Network (DLANet) for image recognition [9], which improves the ability to recognize scenes; Howard et al. [14] proposed MobileNets to give neural network weights through a streamlined structure, which is specifically provided for mobile devices for image recognition; Zbontar et al. [38] proposed a method for applying the convolutional neural network to the image patch to make pairing of the stereograms easier. In terms of energy forecasting, Fumo et al. [10] used regression analysis to predict the residential energy consumption; Jovanović et al. [16] used back-propagation neural networks to predict heating energy, with good precision.

From the above, it can be known that the application of artificial intelligence is very extensive. Among them, financial finance is the most widely studied because it is a multi-factor problem involving global trends, local culture, political influence, etc. Therefore, exchange rate and stock investment are the most difficult to predict. The data of these problems are sorted according to time, and can be regarded as the prediction problem of time series. In the past, many scholars have proposed different methods to predict the prediction problem of time series, like ARIMA[21, 22], fuzzy theory, neural network, neuro-fuzzy systems (NFSs), etc. Among them, the most commonly proposed is the NFSs [13, 28, 30, 40], which has the best performance for time series prediction.

NFSs are based on the extension of the Neural Network, a network in which multiple neuron layers are connected, transmitted through layers between layers to get the outputs, originated in 1943, McCulloch et al. [24] used threshold logic to simulate neurons in the human brain, meaning that when a neuron receives information, it determines whether a mechanism for excitatory responses is to be produced. In 1956, Rochester [27] further created the perceptron, but still could not solve the nonlinear problem. It was not until 1975 that Werbos [36] proposed a back-propagation algorithm that allowed multiple layers of neural networks to be trained, which not only solved the nonlinear problem, but also created the concept of training in the neural network in the future. Today's deep learning is a kind of neural network, and there have been many studies in this field [9, 11, 38].

NFSs are composed of fuzzy systems and neural networks. The fuzzy systems have so-called IF-THEN rules. These rules are like human experience, which are easy to be understood for human. Nowadays, the widely used NFSs are hybrid NFSs [34], which means that the elements in the fuzzy system are used in the neural network to make the overall architecture more flexible. As mentioned above, the characteristics of the neuro-fuzzy system make him have a good effect on the prediction of time series. Therefore, most of the research on time prediction now uses a neural network as the model architecture.

This paper uses a NFS as the framework, which contains the IF-THEN rules to build neurons of multiple layers. In term of model implementation, IF-part uses the Gaussian sphere complex fuzzy sets, which make the model can have multiple outputs to predict multiple targets, and the THEN-part use Takagi-Sugeno linear function [31]. The two neurons are combined by the IF-THEN rule concept, which is a one-to-one relationship. Through this model and machine learning, we expect that the prediction of time series can be excellent.

Regarding fuzzy sets, in 1965, Zadeh first proposed the concept of fuzzy sets [39], so that data can obtain a membership degree of a set between 0 and 1 through a function. Then in 2002, another study proposed the concept of complex fuzzy sets (CFSs) [25]. The complex-valued membership degree can be obtained by a function, which allows the membership degree to be presented in a complex unit disk of radius 1. This concept makes the original membership degree more abundant. We can get the complex-valued outputs by a complex neuro-fuzzy system (CNFS) [19, 20], the real and imaginary part can predict for two different targets, respectively. There are already many research outputs in the prediction of the two targets [19, 20, 22]. In order to predict more targets at the same time, this paper improves the original CNFS and changes the CFSs into sphere complex fuzzy sets (SCFSs), which membership degree is also complex-valued but can present in a 3-D space. SCFSs can have more complex-valued outputs, means can predict more targets at the same time.

In this study, in order to apply the data effectively, in the data preprocessing, we calculate the information contribution of the individual features to the target through Shannon Entropy [29]. In addition, we through the concept of multi-target feature selection [23], the effective information amount of each feature to the target is calculated as the basis for selecting training data. Extracting the most effective data from the data can not only effectively improve the performance of the prediction but also reduce the computational burden of the model. Finally, the parameter learning part, we use the well-known particle swarm optimization (PSO)[8], artificial bee colony optimization (ABCO)[17] two algorithm to combine with recursive least squares estimator (RLSE)[15] for parameter optimization, which are called PSO-RLSE [18] and ABCO-RLSE. We use different algorithm to train the premise and consequence parameters, expect reduce the search dimension through the divide-and-conquer concept and make the model find the best solution easier.

# Methodology

本章節將說明本研究一系列的研究方法以集模型設計。本研究使用資料驅動方式決定模型結構的大小，模型實作使用球式複數模糊集，且分別利用不同的演算法(PSO、ABC)優化其參數，遞迴式最小平方演算法則最佳化後鑑部參數。在資料進入模型之前，透過多目標特徵挑選[23]，挑選出對所有目標較為有效之特徵資料集合，減少龐大資料對模型的負擔。最後將結果配合投資策略做不同演算法的比較。

## Sphere Complex Fuzzy Sets

1965年提出的模糊集合概念，可以導出元素對集合一對一的歸屬程度。隨後Buckley提出了模糊複述的概念[1-3]，直到2002年，Ramot et al.[25]提出了複數歸屬度型態的模糊集合，可以擁有更豐富的資訊。為了使日後應用更加廣泛，我們希望可以透過一個觀念使得歸屬程度比以往模糊集合更加豐富。

球式複數模糊集合是本論文所提出的原創概念，透過此概念可以將一筆資料轉換成多個複數型態的歸屬程度，以便之後模型可以一次預測多個目標。首先，將原高斯函數得到的歸屬程度置放於半徑為1的球式複數模糊集合內(Fig. 1)，即可得到一組空間向量，其成分表示如下。

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

其中，為原高斯函數(25)的歸屬程度;;。藉由的拆解，可得到至少四組的複數型態歸屬程度，包含了以下的歸屬程度。

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

其中，。

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[29]概念，並參考多目標特徵選取方法[23]，實作特徵挑選，最後從挑選後的特徵中取得訓練資料。

在1948年，學者Shannon則提出了資訊熵[29]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。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) [23]，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[23]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[23]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

透過資料驅動的概念，模型的結構會透過資料自動產生。本論文藉由機器學習，使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[5]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。

本研究採用傳統的if-then 規則建構模型，因此前鑑部的神經元個數會與後鑑部個數一樣，透過上述區塊挑選的方法，可以得出個第2層神經元，最後，會以相同個數建構出個第4層神經元。後鑑部神經元為T-S神經元，由T-S 函式構成，T-S公式如下。

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

其中，{}是第個T–S神經元的參數，是第個輸入。

透過結構學習的過程，本論文以客觀的方法建構模型架構，模型運算過程以及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。

模型是透過條T-S模糊規則組成，每條規則是由前鑑部(If part)與後鑑部(Then part)所結合，其規則形式如下：

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

規則數；為模糊系統的輸入變數；為第條規則的複數模糊集；則是輸入的語意變數。複數模糊集的參數即是前鑑部(If part)參數，為後鑑部(Then part)參數。此複數類神經糢糊模型可轉為一個六層架構的類神經網路模型。

**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:** 此層為正規化層，會將個輸入向量中的各個元素正規化，之後將結果以向量型態輸出，輸出如下。

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

其中，為第個神經元中第個元素正規化後的值，。本研究中使用的是複數模糊集，因此輸入是複數型態，故輸出也是複數型態。

**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 5:** 此層為輸出層，將上一層得到的個模型輸出加總，即為模型的輸出。

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

## 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。前鑑部的參數優化會使用PSO以及ABCO，後鑑部的參數優化會使用RLSE，期待透過混合型演算法PSO-RLSE以及ABCO-RLSE帶來優異的效能表現。

1. Particle Swarm Optimization

粒子群演算法是由J. Kennedy et al. [8] 1995年所提出的優化演算法。其原理類似鳥群尋找食物，除了粒子自身提供的資訊，慣性以及群體智慧中全群最佳位置也會被使用在演算法中，用以調節粒子的速度，此演算法特性為收斂快速，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，粒子的位置代表前鑑部的參數，其中包含了每個維度的分群中心、標準差以及相位頻率參數。

1. Artificial Bee Colony Optimization

人工蜂群演算法為Karaboga學者所提出 [17]，原理類似蜜蜂尋找食物來源的概念，其特性包含群體智慧與隨機性，如蜜蜂傳達食物位置時會彼此透過搖擺舞溝通，而搖擺舞所指示食物位置會有偏差帶有隨機性。此演算法中，總共有三種蜜蜂，包括工蜂 (Employed bee)、觀察蜂 (Onlooker bee)以及偵查蜂 (Scout bee)。其中，工蜂負責尋找食物源的位置，然後以跳舞傳達食物源的收益訊息，每次在傳達時會帶有隨機性，代表著整體蜜蜂大維度的搜尋；觀察蜂負責在其中一個食物源附近搜尋，首先會觀察各個食物收益度，並以輪盤法選擇一個食物源，到食物源的鄰近周遭進行搜尋，代表著食物源附近小維度的搜尋；偵查蜂作用為當食物源經過一定的開發次數後，若食物源的收益度沒有改善時，則會派出偵查蜂取代食物源，意味著隨機的探索一個全新的食物源，可以避免演算法陷入區域最佳解中。其步驟與真實的蜜蜂找尋食物不盡相同，演算法步驟如下:

1. 隨機尋找其中一隻工蜂，並透過公式形成新位置，此位置為該次迭代的食物源位置，公式如下。

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

其中，為第個食物源的第個維度;為第隻工蜂的第個維度;為其他隨機工蜂的第個維度的值。

1. 觀察蜂使用輪盤法 (Roulette method)挑選一食物源，其中食物源收益越好越容易被選中，輪盤機率公式如下。

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

其中，為第個食物源被選中的機率;為收益程度，本研究將成本函數的倒數視為收益程度;為食物源的總數目。

1. 派出每隻觀察蜂在剛剛被選中的食物源位置附近搜尋，公式如下。

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

其中，為第隻觀察蜂的第個維度;為被選中的食物源第個維度;為被其他隨機的食物源第個維度的值，若觀察蜂位置比被選中食物源位置好，則取代食物源位置。

1. 判斷每隻工蜂是否已經達到限制回合都未更新，若為真則派出偵查蜂取代，偵查蜂位置產生公式如下。

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

其中，為第個食物源的第個維度;為所有食物源中第個維度的最大值;為所有食物源中第個維度的最小值。

1. 重複Steps 2~4，直到反覆運算結束。
2. Recursive Lest squares estimation

在本論文使用遞迴式最小平方演算法 (Recursive least squares estimation, RLSE)[15]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。

混合型演算法中各部分參數優化會交由不同演算法負責，本研究使用到PSO-RLSE以及ABCO-RLSE兩種混合型演算法，PSO和ABCO負責前鑑部的參數學習，RLSE則負責線性的T-S function 參數學習。混合演算法與模型計算的流程如下:

1. 準備訓練資料及測試資料。
2. 以前鑑部演算法(PSO或ABCO)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。
3. 用RLSE更新T-S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，正規化後的啟動強度為一向量，使為一矩陣，因此在原本的公式(52)中，利用單位矩陣取代原本的常數項1，改良後公式如下。

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

1. 更新完所有參數後，計算出模型的輸出。
2. 計算成本，更新前鑑部演算法粒子的位置及相關數據。
3. 重複Steps 2~5，直到迭代結束。

## Methodology for financial application

In order to evaluate whether the model is helpful to the investment, the use of the cost function is not enough, because cost function means the matching rate, and the high matching rate does not equal high profit. Therefore, this study will predict the closing price in conjunction with the investment strategy [35], and further decide whether to buy or sell, the buying and selling formula is as follows.

|  |  |
| --- | --- |
| buying: if  , |  |
| selling: if  , |  |

where, is the threshold parameter, present the fluctuation of stocks; is the model output, means the predicted closing price of the day; is the actual closing price of the dat. If the predicted closing price is higher than the closing price, means investor has to buy, vice versa.

In order to make investment strategy more cautious, this paper proposes a new strategy as follows.

1. Use formulas (43)-(44) to decide buying or selling, and go to the second phase.
2. Calculate the mean of the past 30 days of the day as the standard of the second phase. Note that after many tests the average of the past 30 days is best.
3. If , and , then buying; if , and , then selling.
4. All trading days need to pass two phase check, if failed in any phase, then not operation.

The profit is calculated by the today closing price and the next day closing price, as follows.

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

where, is the model profit, is the total days when strategy is buying; is the total days when strategy is selling; is the actual closing price of the day.

Through the investment strategy and the profit formula, we can calculate the profits brought by the model to simulate the effect of real-world investment. This study will show the result of two strategies and the other parameters.

# Expeimentation

There are a total of three experiments in this study. The first experiment is a two-target prediction. The target is the Taiwan stock exchange capitalization weighted stock index (TAIEX) and the Hang Seng index (HSI); experiment 2 is the prediction for four targets, test whether the model is feasible by complex-valued membership degree. The complex-valued outputs are used to predict the four targets. For the first complex-valued output, the real part is responsible for the first target, and the imaginary part is responsible for the second target; For the second output, the real part is responsible for the third target, and the imaginary part is responsible for the fourth target. In this experiment, the model simultaneously predicts TAIEX, Dow Jones industrial average index (DJI), national association of securities dealers automated quotation (NASDAQ) and standard and Poor’s 500 (S&P 500). Experiment 3 is also a prediction of four goals, using the two complex-valued outputs, including the closing prices of Apple, IBM, DELL and Microsoft. The third experiment is four well-known technology companies. In addition, the other experiments are very well-known stock indicators. Among them, TAIEX is a weighted calculation index of Taiwan-listed stocks, which represents the fluctuation of Taiwan-listed stocks; HSI is an important indicator reflecting the Hong Kong stock market. The index is calculated from the market value of 50 HSI constituent stocks; DJI covers 9 major industries such as finance, which is a stock price weighted indicator; NASDAQ is a market-weighted indicator of more than 3,000 stocks which are mostly in the technology industry; The S&P500 is the market capitalization of the top 500 US companies, including 11 industries such as IT. These indicators represent the trend of a country or even global stocks, so accurate prediction can bring a lot of help to investors.

For comparison of the other literatures, we will evaluate the model through the error indicator and calculate the profit after the investment. The cost function and the evaluation use root mean square error (RMSE) as follows.

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

, is the number of data; is the error vector of the data; is the target vector; is the output vector; is the Hermitian transpose, means the matrix elements conjugate after the matrix transposing.

In the part of structure learning, the number of rules in this paper which are the number of neurons in the second layer is between 4 and 15. Because the upper limit in block selection is set to 15, and the lower limit is set to 4.

In the investment, the threshold parameter in all experiments is between 0 and 0.1, because means the fluctuation of the stocks, and the stocks ups and downs limit are 10%. For the best selection, this study set 0.001 as the iterated step, and search the threshold from 0 to 0.1. All experiments will find the best threshold through the training data, and calculate the profit in testing data, if the profit in training phase is 0, the profit in testing phase is set to 0, means not invest.

## Example 1—Double Time Series of Daily Taiwan Stock Exchange Capitalization Weighted Stock Index

In this example, we validate the model's performance in real-world time series data. The four targets used are TAIEX and HSI, the closing price in one year. This paper predicts the closing price of 2002 and compares performance with other literature methods. The total number of data is 248. The data of the first ten months of each year is used as training data, and the rest is used as testing data. The training data will extract 30 features, a total of 60 features, each feature data is 205, the features are sorted by TAIEX and HSI. The model proposed in this paper can have multiple complex-valued outputs at a time, so we can predict multiple targets. The real part of the first complex-valued target used in the example is the closing price of TAIEX, and the imaginary part is HSI. The parameter after structure learning are shown as Table I, the number of premises is 15, before the block selection the number of blocks is 625, means the structure learning has a great effect of model size control. Machine learning setting is shown as Table II. In order to validate the stability of the model, this paper runs the model 10 times. The performance is shown as Table III, the profit is shown as Table IV. This experiment will compare with the methods proposed in other literatures, the performance comparison is shown as Table V, the profit comparison is shown as Table VI. The learning curve is shown as Fig. 4. The target and the model output are shown as Fig. 5.

Model Setting (Experiment 1)

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Feature variable as model input | {} |
| Number of fuzzy sets of each input | {3, 3, 3, 3} |
| Number of outputs (complex-valued)\* | 2 |
| Type of premises | SCFS |
| Number of premises (after selection) | 15 |
| Number of premise parameters | 48 |
| Number of consequences | 15 |
| Number of consequence parameters | 75 |

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

Machine Learning Setting

|  |  |
| --- | --- |
| **PSO** | |
| Swarm size | 64 |
| Iterations | 100 |
|  | 0.8 2.0 2.0 |
|  | Random in [0,1] |
| Initial position | Given by SC algorithm in section II |
| Initial velocity | 0 |
| **ABCO** | |
| Swarm size | 64 |
| Iterations | 100 |
| Limit | 20 |
| **RLSE** | |
|  |  |
|  | 75x1 zero vector |
|  | **I** |
| **I** | 75x75 identity matrix |

Ten Trials Comparison (Experiment 1, RMSE)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trials | PSO | | ABCO | |
| TAIEX | HSI | TAIEX | HSI |
| 1 | 96.96 | 83.05 | **85.25** | **80.62** |
| 2 | 90.68 | 86.59 | 87.82 | 86.38 |
| 3 | 104.11 | 186.15 | 36748 | 32343 |
| 4 | 109.61 | 112.37 | 4352.6 | 2340.3 |
| 5 | 87.71 | 93.62 | 88.69 | 84.88 |
| 6 | 211.39 | 313.15 | 84.30 | 85.56 |
| 7 | 94.29 | 89.77 | 2093.3 | 1365.1 |
| 8 | **89.77** | **85.72** | 90.01 | 85.14 |
| 9 | 183.06 | 221.61 | 90.07 | 85.68 |
| 10 | 83.27 | 103.40 | 91.43 | 89.47 |

Investment Profit (Experiment 1)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Trials | PSO-RLSE | | ABCO-RLSE | | PSO-RLSE\* | | ABCO-RLSE\* | |
| Profit | | Profit | | Profit | | Profit | |
| Best | 0.07 | | 0.05 | | 0.047 | | 0.052 | |
| 1 | -1786.5 |  | 643.46 |  | 129.68 |  | 25.08 |  |
| 2 | 88.62 |  | 806.72 |  | 682.75 |  | 157.62 |  |
| 3 | -463.39 |  | -36.57 |  | 475.26 |  | 202.11 |  |
| 4 | -914.08 |  | 1664.9 |  | 64.19 |  | 82.35 |  |
| 5 | -936.58 |  | -355.81 |  | 173.98 |  | **961.18** |  |
| 6 | -173.21 |  | -689.96 |  | 14.83 |  | -3.451 |  |
| 7 | -1070.4 |  | -911.59 |  | **767.09** |  | 59.93 |  |
| 8 | **2524** |  | -1640.9 |  | 64.19 |  | -30.80 |  |
| 9 | -376.84 |  | 1620.8 |  | 119.86 |  | 514.65 |  |
| 10 | 124.96 |  | **2469.8** |  | 364.09 |  | 478.03 |  |
| Average | -298.34 | | 357.09 | | 350.48 | | 244.67 | |
| Std. | 1151.83 | | 1308.77 | | 345.16 | | 315.57 | |
| Maximum | 2524 | | 2469.8 | | 767.09 | | 961.18 | |
| Minimum | -1786.5 | | -1640.9 | | 14.83 | | -30.8 | |

\*The result of the investment strategy proposed in this paper

Performance Comparison (RMSE)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **RMSE** | | | | |
| Training phase | |  | Testing phase | |
| TAIEX | HSI |  | TAIEX | HSI |
| Chen [35] | - | - |  | 75 | 187 |
| Yu [35] | - | - |  | 101 | 170 |
| AR(1) [35] | - | - |  | 66 | 105 |
| SVR [35] | - | - |  | 66 | 107 |
| ANFIS [35] | - | - |  | 65 | 106 |
| ANFIS (EMD) [35] | - | - |  | 52 | 97 |
| PSO-RLSE (proposed) | **99.16** | **98.19** |  | **89.77** | **85.72** |
| ABCO-RLSE (proposed) | **98.77** | **97.12** |  | **85.25** | **80.62** |

Investment Profit Comparison (Experiment 1)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Chen [4] | Yu [4] | SR+ANFIS [4] | SR+SVR [4] | Elman [4] | Cheng et al. [4] | PSO-RLSE | ABCO-RLSE |
| Best | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.012 | 0.07 | 0.05 |
| Profit (TAIEX) | 0 | 0 | 0 | 0 | 0 | -231.02 | **1284.2** | 847.8 |
| Profit (HSI) | -1471 | -1368 | -602.94 | 190.71 | **2342** | 1793.12 | 1239.7 | 1622.0 |



(a)



(b)

1. Learning curve. (a) PSO-RLSE (b) ABCO-RLSE



(a)



(b)



(c)



(d)

1. Actual values and the model outputs. (a) PSO-RLSE (TAIEX) (b) PSO-RLSE (HSI) (c) ABCO-RLSE (TAIEX) (d) ABCO-RLSE (HSI)

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

In this example, the used targets are TAIX, DJI, NASDAQ and S&P500. This paper predicts the closing price of 2001 and compares performance with other literature methods. The data of the first ten months of each year is used as training data, and the rest is used as testing data. The training data is 181, and the testing data is 66. The training data will extract 30 features, a total of 120 features for four targets, the features are sorted by TAIEX, DJI, NASDAQ and S&P500. The model proposed in this paper can have multiple complex-valued outputs at a time. In this experiment has two target, the real part of the first complex-valued target is the closing price of TAIEX, and the imaginary part is DJI, the real part of the second target is NASDAQ, and the imaginary part is S&P500. In the structure learning part, the number of the premise is reduced from 81 to 9 by the block selection, which significantly reduces the size of the model and the model parameters. The model parameters are shown as Table VII. Machine learning setting is shown as Table VIII. In this experiment, except the comparison between PSO-RLSE and ABCO-RLSE, also compares with other literature methods, such as ANFIS[22], CNFS-ARIMA[22], RBF network[22] and SVR[22]. Except the SVR, the other models can predict two targets at the same time. So we use the first complex-valued output to compare with other methods, the result is shown as Table IX, the profit is shown as Table X. In order to validate the model stability, this paper runs model 10 times, result is shown as Table XI. Learning curve of machine learning is shown as Fig. 6; targets and the model output are shown as Fig. 7.

Model Setting (Experiment 2)

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Feature variable as model input | {} |
| Number of fuzzy sets of each input | {3, 3, 3, 3} |
| Number of outputs (complex-valued)\* | 2 |
| Number of premises (before selection) | 81 |
| Type of premises | SCFS |
| Number of premises (after selection) | 9 |
| Number of premise parameters | 48 |
| Number of consequences | 9 |
| Number of consequence parameters | 45 |

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

Machine Learning Setting

|  |  |
| --- | --- |
| **PSO** | |
| Swarm size | 64 |
| Iterations | 100 |
|  | 0.8 2.0 2.0 |
|  | Random in [0,1] |
| Initial position | Given by SC algorithm in section II |
| Initial velocity | 0 |
| **ABCO** | |
| Swarm size | 64 |
| Iterations | 100 |
| Limit | 20 |
| **RLSE** | |
|  |  |
|  | 45x1 zero vector |
|  | **I** |
| **I** | 45x45 identity matrix |

Performance comparison (RMSE)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **RMSE** | | | | |
| Training phase | |  | Testing phase | |
| TAIEX | DJI |  | TAIEX | DJI |
| SVR (two models, each with single output) [22] | - | - |  | 162.46 | 101.44 |
| ANFIS (two models, each with single output) [22] | - | - |  | 147.36 | 105.56 |
| ANFIS (one model with two outputs) [22] | - | - |  | 151.62 | 128.20 |
| RBF (two models, each with single output) [22] | - | - |  | 134.32 | 106.33 |
| RBF (one model with two outputs) [22] | - | - |  | 137.58 | 181.79 |
| CNFS(4)-ARIMA (one model with two outputs) [22] | - | - |  | 115.82 | 103.06 |
| PSO-RLSE (proposed) | **92.92** | **91.50** |  | **96.51** | **92.18** |
| ABCO-RLSE (proposed) | **93.69** | **91.06** |  | **87.92** | **91.95** |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Profit Comparison | | | | | | | | |
|  | Chen [35] | Yu [35] | AR(1) [35] | SVR [35] | ANFIS [35] | Wei [35] | PSO-RLSE | ABCO-RLSE |
| Best | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.047 | 0.012 |
| Profit (TAIEX) | -92 | -73 | 671 | 202 | 686 | **795** | 745.1 | 458.48 |



(a)



(b)

1. Learning curve. (a) PSO-RLSE (b) ABCO-RLSE



(a)



(b)



(c)



(d)

1. Actual values and model output. (a) PSO-RLSE (TAIEX) (b) PSO-RLSE (DJI) (c) ABCO-RLSE (TAIEX) (d) ABCO-RLSE (DJI)

Ten trials performance (Experiment 2, RMSE)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trials | PSO | | ABCO | |
| TAIEX | DJI | TAIEX | DJI |
| 1 | 98.95 | 108.39 | **87.92** | **91.95** |
| 2 | 94.15 | 94.49 | 89.21 | 94.14 |
| 3 | 93.24 | 97.72 | 88.69 | 93.54 |
| 4 | 335.68 | 464.85 | 102.07 | 92.31 |
| 5 | 114.71 | 122.97 | 86.22 | 94.15 |
| 6 | **96.51** | **92.18** | 88.19 | 94.27 |
| 7 | 87.59 | 102.31 | 88.96 | 93.62 |
| 8 | 90.24 | 95.65 | 87.44 | 93.98 |
| 9 | 151.31 | 148.51 | 88.70 | 93.91 |
| 10 | 96.29 | 121.39 | 88.47 | 93.48 |

Investment Profit (Experiment 2)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Trials | PSO-RLSE | | ABCO-RLSE | | PSO-RLSE\* | | ABCO-RLSE\* | |
| Profit | | Profit | | Profit | | Profit | |
| Best | 0.047 | | 0.012 | | 0.061 | | 0.012 | |
| 1 | -264.81 |  | -274.69 |  | 0 |  | 0 |  |
| 2 | -168.32 |  | 359.55 |  | 77.76 |  | 0 |  |
| 3 | 105.9 |  | 916.80 |  | -21.34 |  | 298.98 |  |
| 4 | -1636.8 |  | -862.40 |  | **363.20** |  | 82.62 |  |
| 5 | -378.64 |  | -274.69 |  | 277.37 |  | 0 |  |
| 6 | 532.59 |  | -274.69 |  | -34.84 |  | 0 |  |
| 7 | **987.72** |  | 359.55 |  | 10.77 |  | 0 |  |
| 8 | -235.02 |  | **916.80** |  | -10.67 |  | **298.98** |  |
| 9 | 412.35 |  | -274.69 |  | 0 |  | 0 |  |
| 10 | 199.98 |  | 359.55 |  | 114.98 |  | 0 |  |
| Average | -44.50 | | 95.11 | | 77.72 | | 68.06 | |
| Std | 702.12 | | 579.65 | | 137.28 | | 124.40 | |
| Maximum | 987.72 | | 916.80 | | 363.20 | | 298.98 | |
| Minimum | -1636.8 | | -862.40 | | -34.84 | | 0 | |

\*The result of the investment strategy proposed in this paper.

## Example 3—Quadruple Time Series of Apple, IBM, Dell and Microsoft

In this experiment, the predicted targets are the closing price of APPLE, IBM, Dell and Microsoft, the period is from 10th February 2003 to 21th January 2005, the number of data is 492. In order to compare with other methods, we use the data from 10th February 2003 to 10th September 2004 as the training data which is 400. And the rest is testing data. In this paper, 30 features were extracted from the training data, and there were 120 features for 4 targets. The features were sorted by APPLE, IBM, Dell and Microsoft. The model proposed in this paper can have multiple complex-valued outputs, so can predict multiple targets. The real part of the first target used in this experiment is closing price of IBM, the imaginary part is APPLE, the real part of the second target is DELL, and the imaginary part is Microsoft. In the part of structure learning, through block selection, 8 blocks can be selected from the original 81 blocks to reduce the number of parameters in the model. The rest of the model settings are shown in Table XIII. Machine learning setting is shown as Table XIV. The result of this experiment compares with other literature methods, such as HiMMI [32], ANN-GA-HMM-Interpolation [32], ANN-GA-HMM-WA [32] and ARIMA [32]. So we use the real and imaginary part of the first complex-valued output to compare. The comparison result is shown as Table XV. The investment profit is shown as Table XVI. The learning curve of the model is shown as Fig. 8; The targets and the mode output are shown as Fig. 9.

Model Setting (Experiment 3)

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| Feature variable as model input | {} |
| Number of fuzzy sets of each input | {2, 3, 3, 3} |
| Number of outputs (complex-valued)\* | 2 |
| Type of premises | SCFS |
| Number of premises | 8 |
| Number of premise parameters | 44 |
| Number of consequences | 8 |
| Number of consequence parameters | 40 |

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

Machine Learning Setting

|  |  |
| --- | --- |
| **PSO** | |
| Swarm size | 64 |
| Iterations | 100 |
|  | 0.8 2.0 2.0 |
|  | Random in [0,1] |
| Initial position | Given by SC algorithm in section II |
| Initial velocity | 0 |
| **ABCO** | |
| Swarm size | 64 |
| Iterations | 100 |
| Limit | 20 |
| **RLSE** | |
|  |  |
|  | 40x1 zero vector |
|  | **I** |
| **I** | 40x40 identity matrix |



(a)



(b)

1. Learning curve. (a) PSO-RLSE (b) ABCO-RLSE

Ten Trails Performance (Experiment 3, MAPE)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trials | PSO | | ABCO | |
| APPLE | IBM | APPLE | IBM |
| 1 | 1.7948 | 0.8043 | 1.7625 | 0.8032 |
| 2 | 1.7972 | 0.8039 | 1.7835 | 0.8042 |
| 3 | 1.7963 | 0.8033 | 1.8273 | 0.8002 |
| 4 | 1.7954 | 0.8043 | 1.7818 | 0.8015 |
| 5 | **1.7776** | **0.8033** | 1.7922 | 0.8033 |
| 6 | 1.7921 | 0.8044 | 2.0309 | 0.8002 |
| 7 | 1.7942 | 0.8064 | **1.7602** | **0.8036** |
| 8 | 1.7967 | 0.8043 | 1.7946 | 0.8031 |
| 9 | 1.7920 | 0.8042 | 1.8567 | 0.8046 |
| 10 | 1.7927 | 0.8041 | 1.7758 | 0.8027 |

Investment Profit (Experiment 3)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Trials | PSO-RLSE | | ABCO-RLSE | | PSO-RLSE\* | | ABCO-RLSE\* | |
| Profit | | Profit | | Profit | | Profit | |
| Best | 0.011 | | 0.012 | | 0.021 | | 0.021 | |
| 1 | -4.5715 |  | -10.08 |  | 0 |  | -0.359 |  |
| 2 | 0.94 |  | 1.361 |  | 0 |  | 0 |  |
| 3 | -14.06 |  | -8.678 |  | **1.022** |  | -2.764 |  |
| 4 | -3.37 |  | **2.061** |  | 0 |  | -1.441 |  |
| 5 | **12.22** |  | -8.678 |  | -1.163 |  | **-2.764** |  |
| 6 | 8.72 |  | 2.061 |  | 0.176 |  | -1.441 |  |
| 7 | -10.18 |  | -0.3700 |  | -0.334 |  | 0 |  |
| 8 | 0.94 |  | -8.678 |  | 0 |  | -2.764 |  |
| 9 | -14.06 |  | 2.061 |  | 1.022 |  | -1.441 |  |
| 10 | -3.37 |  | -0.37 |  | 0 |  | 0 |  |
| Average | -2.6794 | | -2.9308 | | 0.072 | | -1.2974 | |
| Std | 8.8075 | | 5.3385 | | 0.626 | | 1.1758 | |
| Maximum | 12.22 | | 2.061 | | 1.022 | | 0 | |
| Minimum | -14.06 | | -8.678 | | -1.163 | | -2.764 | |

\*The result of the investment strategy proposed in this paper.



(a)



(b)



(c)



(d)

1. Actual values and the model output. (a) PSO-RLSE (APPLE) (b) PSO-RLSE (IBM) (c) ACO-RLSE (APPLE) (d) ACO-RLSE (IBM) (e) ABCO-RLSE (APPLE) (f) ABCO-RLSE (IBM)

Performance Comparison (MAPE)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **MAPE** | | | | |
| Training phase | |  | Testing phase | |
| APPLE | IBM |  | APPLE | IBM |
| HiMMI [32] | - | - |  | 2.8373 | 1.2186 |
| ANN-GA-HMM-Interpolation [32] | - | - |  | 2.1649 | 1.0555 |
| ANN-GA-HMM-WA [32] | - | - |  | 1.9247 | 0.8487 |
| Bayesian ANN [32] | - | - |  | 1.9688 | 0.7441 |
| ARIMA [32] | - | - |  | 1.8009 | 0.9723 |
| PSO-RLSE (proposed) | **1.7840** | **1.1960** |  | **1.7776** | **0.8033** |
| ABCO-RLSE (proposed) | **1.7654** | **1.1958** |  | **1.7602** | **0.8036** |

# Discussion

Through three experiments, we can find that the series of methods proposed in this paper have multi-target prediction ability. In the performance of the prediction, it can be found from Experiment 1 that the prediction performance of the dual target is better than that proposed by other literatures. Representing the model of this paper, even if it predicts two goals at a time, it can have excellent performance. It predicts the four goals in experiment 2 and experiment 3, and proves the prediction ability of the model again. Through multi-target feature selection, features that are beneficial to all targets can be extracted from more than 100 features, and training data can be applied efficiently. In the structure learning part, the first experiment selects 15 useful blocks from the 625 blocks as the model premise neurons, in the second and third experiments, the blocks are reduced from 81 blocks. The number of the parameters in the model, is reduced objectively, the size of the model is effectively controlled, and the burden of the model computation is reduced. This is one of the advantages of this model. In the parameter learning part, it can be seen from the performance that the hybrid algorithm has a good performance for parameters searching, and the matching rate is better than other literature methods. Among them, ABCO-RLSE is better than PSO-RLSE, showing the advantages of the ABCO algorithm. Through local search by onlooker bees and the jump out mechanism by scout bees, the ABCO performance is better than PSO. Finally, in the simulation investment part, the profit of the model proposed in this paper in some experiments is better than the past literature, and the experiment lost to other literatures is close. It can be found from the 10 trials performance tables and the investment income statement that there is no positive correlation between performance and investment profit. The investment strategy proposed in this paper improves the standard of decision-making in investment, and judges daily transactions more cautiously. From the results, it can be found that the profit earned by the investment strategy has increased, and the standard deviation has decreased in each experiment. It means better stability when investing, which can be provided to investors for decision-making.

# Conclusion

This paper proposes a hybrid algorithm ABCO-RLSE, which has better performance than PSO-RLSE, both of which have multi-target prediction capabilities. It can be seen from the experiment that both the two targets and the four targets can be predicted at the same time. Through the concept of the sphere complex fuzzy set, even more targets can be predicted at one time. Multi-target feature selection can extract useful data from the original data according to different data, and control the size of the data entering the model, which is an important process for a large amount of data. In the part of structure learning, uses the data-driven concept to automatically adjust the input data according to the data, and the model can adapt to generate different structures when faced with different data. From the 3 experiments, it can be found that different data can be effectively predicted in this model, which means two algorithms combined with RLSE has a certain level. Finally, the investment strategy proposed in this paper helps to reduce the risk of investment and provides an objective reference data for investors.

The part of machine learning may be limited by the characteristics of the algorithm. For example, PSO converges quickly and easily fall into the local minimum. Therefore, in the case of a large amount of data, the required search dimensions are increased, and the PSO performance will be poor, which may limit the overall performance of the model. In the case of ABCO, if it falls into the local minimum, it will re-search after the limit iterations arrives. Relatively speaking, the probability of jumping out the local minimum is higher for ABCO, but the formula for re-searching is through the maximum and minimum values, the initial values of all bees are very important, which will deeply affect the training of the model and make model unstable. In the future, different machine learning algorithms can be combined with SCFNS, such as random search [37], ant colony algorithm [6], etc., to overcome the problems encountered by current algorithms, and also through swarm intelligence concept, such as CPSO [33], etc. looks for the best solution. In terms of simulation investment, we can learn from experiments that each algorithm can make a profit, but the smaller the error of the model, not means the more profit it earns, the two are not positively related. In the investment strategy, the threshold parameter is the most important, so in the future, the threshold parameter can be added to the machine learning to increase the profit brought by the model.

##### ACKNOWLEDGMENT

##### REFERENCES

[1] J. Buckley, "Fuzzy complex numbers," *Fuzzy Sets and Systems,* vol. 33, no. 3, pp. 333-345, 1989.

[2] J. Buckley and Y. Qu, "Fuzzy complex analysis I: differentiation," *Fuzzy Sets and Systems,* vol. 41, no. 3, pp. 269-284, 1991.

[3] J. J. Buckley, "Fuzzy complex analysis II: integration," *Fuzzy Sets and Systems,* vol. 49, no. 2, pp. 171-179, 1992.

[4] C.-H. Cheng and J.-H. Yang, "Fuzzy time-series model based on rough set rule induction for forecasting stock price," *Neurocomputing,* vol. 302, no. 9, pp. 33-45, 2018.

[5] S. L. Chiu, "Fuzzy model identification based on cluster estimation," *Journal of Intelligent & fuzzy systems,* vol. 2, no. 3, pp. 267-278, 1994.

[6] A. Colorni, M. Dorigo, and V. Maniezzo, "Distributed optimization by ant colonies," presented at the Proceedings of the first European conference on artificial life, 1991.

[7] P. du Jardin, "Dynamics of firm financial evolution and bankruptcy prediction," *Expert Systems with Applications,* vol. 75, no. 1, pp. 25-43, 2017.

[8] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *MHS'95., Proceedings of the Sixth International Symposium on Micro Machine and Human Science* 1995, pp. 39-43: IEEE.

[9] Z. Feng, L. Jin, D. Tao, and S. Huang, "DLANet: a manifold-learning-based discriminative feature learning network for scene classification," *Neurocomputing,* vol. 157, no. 1, pp. 11-21, 2015.

[10] N. Fumo and M. R. Biswas, "Regression analysis for prediction of residential energy consumption," *Renewable and Sustainable Energy Reviews,* vol. 47, pp. 332-343, 2015.

[11] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging,* vol. 35, no. 5, pp. 1153-1159, 2016.

[12] J. Heaton, N. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios," *Applied Stochastic Models in Business and Industry,* vol. 33, no. 1, pp. 3-12, 2017.

[13] L. J. Herrera *et al.*, "Multigrid-based fuzzy systems for time series prediction: CATS competition," *Neurocomputing,* vol. 70, no. 13-15, pp. 2410-2425, 2007.

[14] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861,* 2017.

[15] C. T. S. a. E. M. J. S. R. Jang, *Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence prentice hall*. Upper Saddle River, 1997.

[16] R. Ž. Jovanović, A. A. Sretenović, and B. D. Živković, "Ensemble of various neural networks for prediction of heating energy consumption," *Energy and Buildings,* vol. 94, no. 1, pp. 189-199, 2015.

[17] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," *Journal of global optimization,* vol. 39, no. 3, pp. 459-471, 2007.

[18] C. Li and T.-W. Chiang, "Complex neuro-fuzzy self-learning approach to function approximation," in *Asian Conference on Intelligent Information and Database Systems*, Berlin, Heidelberg, 2010, pp. 289-299: Springer.

[19] C. Li, T.-W. Chiang, J.-W. Hu, and T. Wu, "Complex neuro-fuzzy intelligent approach to function approximation," in *Advanced Computational Intelligence (IWACI), 2010 Third International Workshop on*, 2010, pp. 151-156: IEEE.

[20] C. Li and T.-W. Chiang, "Complex fuzzy computing to time series prediction—A multi-swarm PSO learning approach," in *Asian Conference on Intelligent Information and Database Systems*, 2011, pp. 242-251: Springer.

[21] 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, 2012.

[22] C. Li and T.-W. Chiang, "Complex neurofuzzy ARIMA forecasting—a new approach using complex fuzzy sets," *IEEE Transactions on Fuzzy Systems,* vol. 21, no. 3, pp. 567-584, 2013.

[23] C. Li, "Multi-target feature selection," *unpublished,* 2017.

[24] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics,* vol. 5, no. 4, pp. 115-133, 1943.

[25] D. Ramot, R. Milo, M. Friedman, and A. Kandel, "Complex fuzzy sets," *IEEE Transactions on Fuzzy Systems,* vol. 10, no. 2, pp. 171-186, 2002.

[26] V. Ravi, D. Pradeepkumar, and K. Deb, "Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms," *Swarm and Evolutionary Computation,* vol. 36, pp. 136-149, 2017.

[27] N. Rochester, J. Holland, L. Haibt, and W. Duda, "Tests on a cell assembly theory of the action of the brain, using a large digital computer," *IRE Transactions on information Theory,* vol. 2, no. 3, pp. 80-93, 1956.

[28] H.-J. Rong, N. Sundararajan, G.-B. Huang, and P. Saratchandran, "Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction," *Fuzzy sets and systems,* vol. 157, no. 9, pp. 1260-1275, 2006.

[29] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE mobile computing and communications review,* vol. 5, no. 1, pp. 3-55, 2001.

[30] 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, 2008.

[31] 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, 1985.

[32] J. L. Ticknor, "A Bayesian regularized artificial neural network for stock market forecasting," *Expert Systems with Applications,* vol. 40, no. 14, pp. 5501-5506, 2013.

[33] 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, 2004.

[34] J. Vieira, F. M. Dias, and A. Mota, "Neuro-fuzzy systems: a survey," in *5th WSEAS NNA international conference on neural networks and applications, Udine, Italia*, 2004.

[35] L.-Y. Wei, "A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting," *Applied Soft Computing,* vol. 42, pp. 368-376, 2016.

[36] P. Werbos, "Beyond regression: new fools for prediction and analysis in the behavioral sciences," *PhD thesis, Harvard University,* 1974.

[37] Z. B. Zabinsky, "Random search algorithms," *Wiley Encyclopedia of Operations Research and Management Science,* 2010.

[38] J. Zbontar and Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches," *Journal of Machine Learning Research,* vol. 17, no. 1-32, p. 2, 2016.

[39] L. A. Zedeh, "Fuzzy sets," *Information and control,* vol. 8, no. 3, pp. 338-353, 1965.

[40] M. Zounemat-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, 2008.