**國 立 中 央 大 學**

資訊管理研究所

碩 士 論 文

複數型模糊類神經系統及連續型態之多蟻群演化在時間序列預測之研究

Complex Neuro-Fuzzy System with Multi-Group Continuous Ant Colony Optimization for Time Series Forecasting

|  |  |  |
| --- | --- | --- |
|  | 研 究 生：姚宇謙  指導教授：李俊賢 博士 |  |

中 華 民 國 106 年 6 月

複數型模糊類神經系統及連續型態之多蟻群演化在時間序列預測之研究

|  |  |
| --- | --- |
| 研究生：姚宇謙 | 指導教授：李俊賢博士 |

國立中央大學資訊管理學系碩士班

**摘要**

隨著資料的快速增加，如何有效分析隱藏在大量資料中的價值日益重要。在資料分析的領域中，時間序列的分析與預測是一主要的研究方向。本研究提出一個複數模糊類神經模型(Complex neuro-fuzzy model)，其結合複數模糊集(Complex fuzzy set)、T-S模糊系統(T-S fuzzy system)形成該模型。在參數學習，以多群連續型蟻群演算法(Multi-group continuous ant colony optimization, MGCACO)與遞迴最小平方演算法(Recursive least squares estimator, RLSE)結合，成為MGCACO-RLSE複合型演算法，進行參數的搜尋與最佳化。多群連續型蟻群演算法在演化過程中，加入資料流通、淘汰、繼承等特性，能夠減少演算法落入區域最佳解以及加速進行參數的最佳化。在資料進入模型前，利用特徵選取(Feature selection)的方式，擷取其中較為有影響力的資料進行預測，減少模型負擔。此外，模型的歸屬程度(Membership degree)是複數型態，並且能拆解成多個不同的歸屬程度，使模型達到預測多目標的效果。本研究以三個實驗來驗證模型的效能與研究理論。實驗一的單目標證實本研究的理論，實驗二與實驗三的多目標實驗，個別證明模型的複數形態輸出以及利用多個歸屬程度達到的多目標輸出方法。個別實驗結果皆與過往文獻比較，實驗顯示本研究模型在時間序列預測上有良好效能，更加證實本研究的可行性。

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

Complex Neuro-Fuzzy System with Multi-Group Continuous Ant Colony Optimization for Time Series Forecasting

|  |  |
| --- | --- |
| Graduate student: Yu-Chien Yao | Advisor: Dr. Chunshien Li |

Department of Information Management, National Central University, Taiwan

**Abstract**

In the age of information, it is increasingly important to deal with big data effectively for both scientific researches and applications of data science. In this study, we have proposed a complex neuro-fuzzy system to data prediction, using complex fuzzy sets and logic and T-S neural fuzzy modeling. Complex fuzzy set is an advance fuzzy set whose membership degrees are complex-valued and defined with the unit disc of the complex plane, in contrast to regular fuzzy set whose membership degrees are real-valued and defined within [0,1]. For model construction, we have used feature selection to get influential data for the proposed model. For the optimization of the proposed model, we have developed a hybrid method for machine learning, denoted as MGCACO-RLSE, integrating the proposed multi-group continuous ant colony optimization (MGCACO) and the well-known recursive least squares estimator (RLSE) method. The MGCACO-RLSE has been applied to optimize the parameters of the proposed model. For MGCACO, we have added some properties, such as data exchange, elimination, and inheritance to increase its search efficacy, in the sense of reducing the possibility of being trapped at a local optimum and thus increasing the chance of finding the optimization solution. In this study, we have conducted three experiments to verify the effectiveness and rationale of the proposed approach. In experiment one, the proposed approach was tested for data prediction with single target to see the feasibility of the research thought. In experiments two and three, the proposed approach was tested for multi-target prediction. With the experimental results, the proposed approach has shown good performance, through performance comparison to other methods in literature.

|  |  |
| --- | --- |
| **Keywords:** | Feature selection; Complex fuzzy set; Complex neuro-fuzzy system; Ant colony optimization; RLSE; Multi-target forecasting. |

# 致謝

# 目錄

|  |  |
| --- | --- |
| 章節 | 頁次 |

[複數型模糊類神經系統及連續型態之多蟻群演化在時間序列預測之研究 i](#_Toc490765897)

[Complex Neuro-Fuzzy System with Multi-Group Continuous Ant Colony Optimization for Time Series Forecasting ii](#_Toc490765898)

[致謝 iii](#_Toc490765899)

[目錄 iv](#_Toc490765900)

[圖目錄 vii](#_Toc490765901)

[表目錄 ix](#_Toc490765902)

[第一章 緒論 1](#_Toc490765903)

[1.1 研究背景 1](#_Toc490765904)

[1.2 研究動機與目的 1](#_Toc490765905)

[1.3 研究方法概述 3](#_Toc490765906)

[1.4 論文架構 4](#_Toc490765907)

[第二章 文獻探討 6](#_Toc490765908)

[2.1 特徵選取 6](#_Toc490765909)

[2.1.1 過濾法 7](#_Toc490765910)

[2.1.1 打包法 7](#_Toc490765911)

[2.1.1 嵌入法 8](#_Toc490765912)

[2.1.2 夏農資訊熵 9](#_Toc490765913)

[2.2 複數模糊集合 9](#_Toc490765914)

[2.2.1. 模糊集合緣起 9](#_Toc490765915)

[2.2.2. 模糊集合 10](#_Toc490765916)

[2.2.3. 複數模糊集 11](#_Toc490765917)

[2.3 類神經模型 12](#_Toc490765918)

[2.4 蟻群演算法 13](#_Toc490765919)

[2.4.1 連續型蟻群演算法 15](#_Toc490765920)

[2.5 多群演算法 19](#_Toc490765921)

[第三章 系統設計與架構 21](#_Toc490765922)

[3.1 特徵選取 21](#_Toc490765923)

[3.1.1 單目標特徵選取策略 27](#_Toc490765924)

[3.1.2 多目標特徵選取策略 28](#_Toc490765925)

[3.2 複數模糊類神經模型 29](#_Toc490765926)

[3.3 多群連續型蟻群演算法 34](#_Toc490765927)

[3.3.1 資訊流通 34](#_Toc490765928)

[3.3.2 淘汰 35](#_Toc490765929)

[3.3.3 繼承 35](#_Toc490765930)

[3.4 遞迴式最小平方演算法 37](#_Toc490765931)

[3.5 MGCACO-RLSE複合演算法 39](#_Toc490765932)

[第四章 實驗 43](#_Toc490765933)

[4.1 實驗一：道瓊工業指數時間序列預測 43](#_Toc490765934)

[4.2 實驗二：高盛與微軟股價資料預測 49](#_Toc490765935)

[4.3 實驗三：巴西股市指數、日經平均指數、道瓊工業指數時間序列預測 57](#_Toc490765936)

[第五章 討論 69](#_Toc490765937)

[5.1 利用複數類神經網路模型針對單目標資料進行預測 70](#_Toc490765938)

[5.2 利用複數型態輸出針對雙目標進行預測 70](#_Toc490765939)

[5.3 解構歸屬程度值進行多目標預測 71](#_Toc490765940)

[5.4 特徵選取之應用 71](#_Toc490765941)

[5.5 MGCACO-RLSE複合式演算法效能分析 72](#_Toc490765942)

[第六章 結論與未來研究方向 73](#_Toc490765943)

[6.1 結論 73](#_Toc490765944)

[6.2 未來研究方向 74](#_Toc490765945)

[參考文獻 76](#_Toc490765946)

# 圖目錄

|  |  |
| --- | --- |
| 圖 | 頁次 |

圖 1 預測方法流程圖 4

圖 2 Dash對於特徵選取的流程圖 6

圖 3 Yu對於特徵選取改良後的流程圖 7

圖 4 過濾法流程圖 7

圖 5 打包法流程圖 8

圖 6 嵌入法流程圖 8

圖 7 類神經網路模型架構 13

圖 8 連續型蟻群演算法費洛蒙表單 16

圖 9 連續型蟻群演算法流程圖 19

圖 10 25

圖 11 特徵選取流程圖 26

圖 12 單目標特徵選取流程 28

圖 13 多目標特徵選取流程 29

圖 14 複數模糊類神經系統模型 32

圖 15 多群連續型蟻群演算法各群取代較差蟻群解流程 35

圖 16 多群連續型蟻群演算法淘汰最差蟻群流程 35

圖 17 多群連續型蟻群演算法演化流程 37

圖 18 遞迴式最小平方演算法流程圖 39

圖 19 複合式演算法學習流程圖 42

圖 20 輸入特徵數對於目標總選取增益量變化 46

圖 21預測結果-以道瓊工業指數為目標: 藍色實線為實際目標值，紅色虛線為模型預測值 47

圖 22 模型預測誤差值-以道瓊工業指數為目標 47

圖 23 實驗一模糊類神經模型學習曲線 48

圖 24輸入特徵數對於目標總選取增益量變化-以高盛集團為目標 51

圖 25輸入特徵數對於目標總選取增益量變化-以微軟集團為目標 52

圖 26預測結果-以高盛集團股價為目標: 藍色實線為實際目標值，紅色虛線為模型預測值 54

圖 27模型預測誤差值-以高盛集團股價為目標 54

圖 28預測結果-以微軟股價為目標: 藍色實線為實際目標值，紅色虛線為模型預測值 55

圖 29模型預測誤差值-以微軟股價為目標 55

圖 30 實驗二模糊類神經模型學習曲線 56

圖 31輸入特徵數對於目標總選取增益量變化-以巴西股市指數為目標 60

圖 32輸入特徵數對於目標總選取增益量變化-以道瓊工業指數為目標 61

圖 33輸入特徵數對於目標總選取增益量變化-以日經平均指數為目標 62

圖 34預測結果-以巴西股價指數為目標: 藍色實線為實際目標值，紅色虛線為模型預測值 63

圖 35模型預測誤差值-以巴西股價指數為目標 63

圖 36預測結果-以道瓊工業指數為目標: 藍色實線為實際目標值，紅色虛線為模型預測值 64

圖 37模型預測誤差值-以道瓊工業指數為目標 64

圖 38預測結果-以日經平均指數為目標: 藍色實線為實際目標值，紅色虛線為模型預測值 65

圖 39模型預測誤差值-以日經平均指數為目標 65

圖 40實驗三模糊類神經模型學習曲線 66

# 表目錄

|  |  |
| --- | --- |
| 表 | 頁次 |

表 1 實驗一之特徵選取次序、特徵變數、選取增益及總選取增益- 以道瓊工業指數為目標(節錄前十選擇次序) 45

表 2 實驗一複數類神經模型設定 46

表 3 實驗一MGCACO-RLSE複合式演算法參數設定 46

表 4 實驗一學習後前鑑部參數值 48

表 5 實驗一學習後後鑑部參數值 48

表 6 實驗一效能比較-以道瓊工業指數為目標[ 34] 49

表 7實驗二之特徵選取次序、特徵變數、選取增益及總選取增益- 以高盛集團為目標(節錄前十選擇次序) 51

表 8實驗二之特徵選取次序、特徵變數、選取增益及總選取增益- 以微軟為目標(節錄前十選擇次序) 52

表 9實驗二複數類神經模型設定 53

表 10實驗二MGCACO-RLSE複合式演算法參數設定 53

表 11實驗二學習後前鑑部參數值 56

表 12實驗二學習後後鑑部參數值 56

表 13實驗二效能比較-以高盛集團為目標 [ 35] 57

表 14實驗二效能比較-以微軟為目標 [ 35] 57

表 15實驗三之特徵選取次序、特徵變數、選取增益及總選取增益- 以巴西股市指數為目標(節錄前十選擇次序) 59

表 16實驗三之特徵選取次序、特徵變數、選取增益及總選取增益- 以道瓊工業指數為目標(節錄前十選擇次序) 60

表 17實驗三之特徵選取次序、特徵變數、選取增益及總選取增益- 以日經平均指數為目標(節錄前十選擇次序) 61

表 18實驗三複數類神經模型設定 62

表 19實驗三MGCACO-RLSE複合式演算法參數設定 62

表 20實驗三學習後前鑑部參數值 66

表 21實驗三學習後後鑑部參數值 67

表 22實驗三效能比較- 以巴西股市指數為目標 [ 36] 68

表 23實驗三效能比較- 以道瓊工業指數為目標 [ 36] 68

表 24實驗三效能比較- 以日經平均指數為目標 [ 36] 68

# 緒論

## 研究背景

時間序列資料是利用時間作為排序依據的一組隨機數值的變量集合，包括股市、匯率皆是一種時間序列資料。利用電腦，加以數理方法或是統計模型對其進行研究，是現今社會進行時間序列資料分析常使用的方式。隨著網路快速發展，以及科技資訊的進步，導致資訊量的增加相當快速。利用機器學習或是資料探勘等方式，可以找出隱藏於資料背後所隱藏的關聯性或是規則，而這些知識是無法僅依靠人類智慧就可找出的。

時間序列資料的預測與分析被使用於許多產業領域上，如交通或是財政狀況；或是短期的演化上，如醫療或是天氣學；以及預測情勢方面，包括商業決策、犯罪或是疾病趨勢。但是當中以財務經濟等資料的時間資料預測最為困難，因為當中考量因素眾多，包括全球情勢、當地文化等，以至於其時間序列變化最為廣泛。因此若是能夠以一訓練良好之模型，搭配關鍵因素做為輸入資料，以此正確了解未來趨勢，便能協助企業或是投資者做出正確的決定。

## 研究動機與目的

早期時間序列預測大多用於預測股票市場，在傳統的計量經濟學方式，對於時間序列的變異是固定的，與現實狀況不符合。因為股票收益的波動幅度是隨時間而變化的，並非常數。在1982年，Engle提出了ARCH模型[ 1] (Autoregressive Conditional Heteroscedasticity) 解決了時間序列的波動性（Volatility）問題，並且被大量用來進行股票市場的財經分析。之後，Bollerslev改進了ARCH模型，針對誤差的方差進行了進一步的建模，提出GARCH(Generalized ARCH)模型來分析[ 2]。但是在實際的股票市場中，股票的資料分布並不會像以往提出的假說相似，因此更需要新的方法來良好的分析與解釋股票市場。而且過往的模型都只能接受單一變數的資料輸入，但是股票市場是會被許多因素所影響，如世界經濟環境或是其他國家的股票市場等等。因此在往後的股票預測，許多研究開始導入人工智慧作為新型態的時間序列分析工具，並且以多資料輸入的方法提出改良。1990年，Kimoto, Asakawa, Yoda和Takeoka提出利用類神經網路的方式來預測股票市場[ 3]。接著在2000年Kim與Han將基因演算法以及類神經網路結合來預測股市[ 4]。2007年Roh合併類神經網路與時間序列模型來預測股市價格的變化[ 5]。

但是在進行預測研究的過程中，經常發生預測失準或是時間過長等狀況而延誤判斷。因此在相關研究中，研究的方向主要是針對如何將資料做有效的分類與挑選。進行分析預測的過程中，模型需要經過學習以及最佳化的過程來加強分析精確度，目前文獻中關於尋找參數最佳解的機器學習方法有粒子群最佳化演算法 (Particle swarm optimization, PSO) [ 6] [ 7] 以及蟻群演算法 (Ant colony optimization, ACO) [ 8]等方式，但是目前研究大多是以T-S模糊系統建立架構，而對於演算法需要去學習的模糊系統模型之前鑑部參數僅以一組粒子群進行搜尋最佳解。會讓ACO等演算法面臨求解的參數維度過大時，會影響粒子的活動性。以至於發生效果不如預期，收斂過於緩慢以及落入區域最佳解的問題。

本研究將提出一種以多群的方式並且結合連續型蟻群演算法 (Continuous ant colony optimization, CACO)以及粒子群演算法其概念的新形態多群演算法。以多群來減低機器學習在尋求最佳解的過程中，落入區域最佳解的機率。並且會以動態的方式來進行群體的數量增減，使演算法可以忽略區域最佳解並且往全域最佳解的方向進行搜尋。搭配粒子群演算法中，以全域最佳解和粒子最佳解更新粒子速度的概念，讓連續型多群蟻群演算法在每次迭代更新群內各組解時，可以包含全域及該群最佳解的內容。

除此之外，本研究針對輸入模型的資料，將利用特徵選取的方式，從雜亂的資料中萃取有效的內容。目前研究缺乏一個良好的方法對於多樣化的資料進行選取，大多皆是以主觀的方式來預測目標。如[ 9 ]是當資料產生的啟動強度若是超過門檻值，便會選入進行預測。而[ 10]則是固定以要預測的目標的前幾筆作為輸入資料，以上方法皆是以資料本身所產生的數值的方法作為特徵選取的方式，可能會忽略資訊本身對於目標的影響，進而影響到預測的結果。但是如果針對資料所提供的資訊分析，就必須要將其資訊量化。當收集到一件資料，同時就獲得了一些資訊，而資訊量就是代表其資料的內涵。1949年，夏農(Shannon)提出了資訊熵的理論[ 11]，理論中將資訊的不確定性量化，若是不確定的值越低，則該資訊對於目標能夠提供的資訊量就越多，亦即對目標越有影響力。而本研究採用的特徵選取是基於夏農資訊熵理論所開發。

本研究在模型的建置上應用了類神經網路系統結合T-S (Takagi-Sugeno)的模糊規則型態組成模糊類神經模型。並且將模糊類神經模型的參數分為兩部分，分別為前鑑部參數 (If-part parameters)與後鑑部參數 (Then-part parameters)。

傳統上，在建立模糊集的設計多是以一維高斯函數建立模糊集，資料在模糊集上產生的歸屬程度值是在x軸上介於0到1的值，因此只能在一維線條上移動。Ramot在2002年提出了Complex Fuzzy Sets[ 12]，將原本實數型歸屬程度再加上一個虛部值，使歸屬程度的值從傳統的一維區間增加到實數軸x與虛數軸y兩個維度上的單位圓盤移動及調整學習。相較於原本的實數型歸屬函數，每個x軸的值都可以再有y軸的調整空間，增加歸屬程度的精準度。

## 研究方法概述

本研究提出融入粒子群演算法的概念於最佳化演算法中，並且在最佳化演算法中，結合了以下演算法：連續型蟻群演算法，遞迴式最小平方估計法 (Recursive least squares estimator, RLSE)，加上多群體的方式進行模型參數的最佳化。命為MGCACO-RLSE (Multi-group continuous ant colony optimization-RLSE)。

首先會利用特徵選取 (Feature selection, FS)對資料分析，藉由計算原始資料對目標所提供的資訊量作為標準，並且加入冗餘資訊的計算，最後選出對目標有良好影響的資料作為輸入至模型的依據，可以降低資料複雜度與計算成本。之後透過MGCACO-RLSE演算法，利用MGCACO調整前鑑部參數，RLSE更新後鑑部參數。利用RLSE更新後鑑部參數原因是因為其演算法能夠以計算方式搜尋到後鑑部線性規則參數的最佳解，使其不必用MGCACO在參數空間中緩慢搜尋最佳解。藉著本研究提出的新形態演算法對模型的參數學習，使模型能夠得到良好的預測結果。圖 1是本研究的預測方法流程圖。

原始資料

特徵選取

訓練資料

T-S模糊類神經模型

訓練目標

MGCACO-RLSE

**+**

評估效能

最佳化參數

圖 1 預測方法流程圖

## 論文架構

本論文一共分為六章，第一章為緒論，介紹本研究的背景、動機及研究之目的、概述本研究所要使用的方法及流程。第二章為研究理論方法的介紹及概述，包括特徵選取的各種方式及本研究的特徵選取採用的夏農資訊熵理論；模糊集的緣起，以及如何將複數模糊集以歸屬函數表現。接著介紹類神經模型及似於本研究的倒傳遞類神經網路模型。第二章最後提及本研究的參數學習蟻群演算法，分別從離散型以及連續型等不同面向介紹，以及在研究中採用的多群演算法概念。第三章為系統設計與架構，概述如何將特徵選取用於輸入資料的篩選，如何將複數模糊集、T-S模糊系統、類神經網路結合成為複數模糊類神經模型。以及利用多群連續型蟻群演算法和遞迴式最小平方演算法個別優化前鑑部與後鑑部的參數。最後則是概述以混和兩種演算法進行參數學習與訓練。第四章是實驗及結果，本研究用了3個實驗進行時間序列預測之研究。實驗一利用道瓊工業指數進行單一目標的預測，實驗二是同時進行高盛集團及微軟的股價值預測實驗，實驗三則是一次進行三目標的預測，分別是巴西股市指數、日經平均指數、道瓊工業指數。以上實驗皆會以過去文獻所提出的方法進行校能比較，以證實本研究方法的優劣性。第五章為針對實驗的結果進行討論。第六章則是本研究的結論以及未來可以改良的研究方向。

# 文獻探討

## 特徵選取

特徵選取在機器學習或是資料探勘的領域中是個必經的過程，目的是挑選出有用以及去除掉部分多餘的資料。因為在原始的資料集，包含了許多雜亂的資料，若不經過篩選就全部進行預測或學習，會讓其中所包含的冗餘或是錯誤的資料輸入到模型中，造成預測失準或是計算量過於龐大。並且讓使用者無法了解那些資料是關鍵資料，進而不能做出良好的判斷。但是如果僅挑選預測目標的前一或兩筆資料進行預測，卻又過度主觀。特徵選取是從資料中挑出一些最為有效的子集合或是相關特徵以降低資料維度的過程。對於預測模型，有著良好的學習資料是預測分析的關鍵。Pedro Domingos曾提到機器學習成功的關鍵就是在於選取特徵的策略[ 13]。Dash[ 14]定義特徵選取的目的是在不會影響分類精準度及資料分布的前提下盡可能縮小特徵子集合。Dash對特徵選取的流程概念如圖 2。

是

形成子集合

評估子集合

停止條件

原始資料

候選子集合

否

最終子集合

圖 2 Dash對於特徵選取的流程圖

此方式當面臨高維度或是大量資料時，反而會因為其高維度的資料在遞迴的過程中造成計算上的資源消耗。因此Yu提出了新的選取流程[ 15]，免去了重複搜尋子集合的過程，減少時間資源浪費。如圖 3。

原始資料

相關性分析

冗餘分析

相關子集合

最終子集合

圖 3 Yu對於特徵選取改良後的流程圖

特徵選取的方式主要有以下三種，過濾法、打包法及嵌入法。

### 過濾法

過濾法(Filter method)是用事先定義的評估標準來決定被評估的特徵的優劣，接著再根據特徵的重要性依序得到特徵子集合，再挑選較好的子集合進行學習。其流程圖如圖 4。相關的過濾法包含了RELIEF, CFS等。RELIEF[ 16]的過濾評估準則是以特徵資料之間的距離作為區分，其優點在於可以處理連續或是離散型的資料，但是只限於解決兩個類別的分類問題。另外一種則是關聯性特徵選擇(Correlation-based Feature Selection, CFS) [ 17]，主要是利用評估函數來評價子特徵集合。判斷出那些特徵是多餘的，那些特徵具有預測能力。

特徵子集評價

特徵子集

特徵子集合

原始資料集

特徵子集選擇

評估函數

最終子集合

圖 4 過濾法流程圖

### 打包法

打包法(Wrapper method)是將學習演算法取代過濾法中的評價函數，對各個特徵子集合進行評估，其流程圖如圖 5。相較於過濾法，所消耗的時間資源較多。因此有許多研究策略的開發皆是用於降低打包法的計算成本。如啟發式搜索中的變數前向選取(Sequential forward selection, SFS)與變數後向選取(Sequential backward selection, SBS)。SFS是先令特徵子集合為空集合，每次選擇一個特徵加入子集合並評估，藉此找到最優的特徵子集合。SBS則是相反做法，先使特徵子集合為全集，每次從子集合去除一個特徵並評估。但是以上兩種方式皆容易落入區域最佳解。Yu [ 18]在2009年提出以RELIEF為基礎的RankWrapper演算法，其將特徵分為子集合後，用以訓練資料進行學習，找到最佳效果時，其所利用的特徵子集合即為最適當集合。

特徵子集評價

特徵子集

特徵子集合

原始資料集

特徵子集選擇

學習演算法

最終子集合

圖 5 打包法流程圖

### 嵌入法

嵌入法(Embedded method)則是把特徵選取的步驟放入學習模型中，圖 6為該方法的流程圖。最常使用的方式是利用正規化進行特徵選取。其會定義最佳化目標的評估標準並放入預測演算法中，減低對模型的複雜度。如Sun [ 19]的RSRank法，便是在建構學習模型的過程中先建立評估標準，再利用梯度下降法來求解最佳化的問題。

特徵子集效能

特徵子集

特徵子集合

原始資料集

特徵子集選擇

模型

圖 6 嵌入法流程圖

### 夏農資訊熵

本研究的特徵選取方式基於夏農資訊熵(Shannon information entropy)的理論，熵一詞最早是由德國物理學家Rudolph Clausius 於 1854 年提出[ 20]。是一種對物理系統之無秩序或亂度的量度，即熱力學中用於測量熱能當中無法轉換成有用的功的一種物理量。而在本研究中所採用的夏農資訊熵已非原本的物理量，夏農(Shannon)將熵定義為資訊內容其不確定性的量值。夏農資訊熵的公式如下：

|  |  |
| --- | --- |
|  | (1) |

就是隨機變數的資訊熵，則是事件的發生機率。由(1)可知，當時，。即事件的發生機率越大，熵值越小。反之，若事件出現機率趨於平均，則熵值會趨近最大值。如前所敘，進行特徵選取的目的是要找出對預測目標最有效益的特徵子集合並且排除冗餘的資料。亦即模型在經由特徵選取選出的特徵子集合後，對於預測的誤差或不確定性降低。因此本研究的特徵選取策略將應用夏農資訊熵推導特徵子集合對於目標的資訊量，資訊量即為目標原始的資訊熵量與目標在接受特徵子集合後其資訊熵的差量。

## 複數模糊集合

在真實世界中，模糊概念到處存在。許多事物的性質或是關係都是模糊的，人類在大多數語言交談及思維也都是模糊的。例如：他跑的非常快、這位明星非常美麗、這兩個人長的很相似等。這些句子中，非常快、非常美麗、很相似，都是模糊語意詞，雖然無法準確得將這些詞彙分類並定義其值域，但是人腦都可以很容易地處理並理解。

### 模糊集合緣起

傳統所採用的集合為明確集合( Crisp set)，最早關於集合論的研究起源於Cantor[ 21]。集合的概念最基礎是由一個元素與一個集合而成，若是元素屬於集合，可記為。

集合中另外一個概念為特徵函數( Characteristic function)，在此我們定義為集合的特徵函數。則可以表示如下

|  |  |
| --- | --- |
|  | (2) |

從以上可知，若是為的元素，則特徵函數；反之，若是不屬於集合，則其特徵函數。的值域可以記為，可以表示是否屬於該集合。但是此論點於Bertrand Russell提出理髮師悖論後產生矛盾，該悖論內容為理髮師只為，而且一定要為，城裡所有不為自己理髮的人理髮。但是理髮師是否該為自己理髮？如果他為自己理髮，根據論點「只為城裡所有不為自己理髮的人理髮」他不應該為自己理髮；但如果他不為自己理髮，同樣按照論點「一定要為城裡所有不為自己理髮的人理髮」他又應該為自己理髮。於是便開始有了關於模糊的概念產生。

模糊的概念最早可以由柏拉圖”理想國”中的線喻概念提及，把世界分成可見與可知兩個部分。而可知的世界是模糊的，也就如同人類的語言一般。Bertrand Russell在1923年的” Vagueness”[ 22] B. Russell, “Vagueness,” *Australasian Journal of Philosophy*, vol.1, iss. 2, pp.84 – 92, 1923。他認為語言都是模糊的，如”紅的”、”老的”，皆沒有一個明確的方式去定義是何種情況。但是在特定情況下，人類利用該詞來進行描述，卻又可以互相了解。在1965年，Zadeh在期刊上發表”Fuzzy Sets” [ 23]把傳統集合特徵函數從非0即1的二值選擇，推廣為可從0到1之間的任何值作選擇，此新型的特徵函數，稱之為歸屬函數(Membership function)，開始了模糊理論新的一頁。

### 模糊集合

如前所敘，集合其特徵函數的值只包含0與1。假設有一模糊集，其歸屬函數為。而元素屬於模糊集的程度可表示如下：

|  |  |
| --- | --- |
|  | (3) |

就是對模糊集的歸屬程度(Membership degree)，其為一個0到1之間的數值，0到1之間的地帶則是類似人類思考邏輯的模糊概念。因此Fuzzy理論就是針對人腦對於模糊的訊息或不完全的資料，其不需經過精密繁雜的計算過程，仍能做出正確判斷的特色而發展出來。

### 複數模糊集

傳統的模糊集合的歸屬程度數值是介於0到1的區間，在2002年，D. Ramot提出了有關複數模糊集的概念[ 24]，複數模糊集是將傳統在一維上的歸屬程度值拓展到二維的複數單位圓盤上。假設有一複數模糊集，其定義於一宇集上。而對於模糊集的歸屬函數則為且所有。複數模糊集可表示如下：

|  |  |
| --- | --- |
|  | (4) |

如前所敘，複數模糊集的歸屬程度是位在單位圓盤的平面上，而對於複數模糊集，其歸屬函數可表示如下：

|  |  |
| --- | --- |
|  | (5) |

為，與為實數數值，且是介於[0,1]間。由此可知，歸屬函數其值是由振幅函數以及相位函數所定。而當相位函數的情況下，，即等於。又因為是介於[0,1]間，因此若是相位函數為0，此歸屬函數會回歸原本的一維歸屬函數。

複數模糊集把元素和其集合之間的歸屬關係從一維區間0至1的範圍，擴展為複數平面上的單位圓盤範圍的歸屬關係。亦即以前所了解的實數區間的歸屬關係，變成複數單位圓盤在虛部是0時的歸屬關係。此關係拓展超過平常經驗所能理解的內涵，有助於提升模糊系統的推理能力與應用效能。2013年，Li將複數模糊集合融入高斯模糊集，使其可以進入模型並分析資料[ 25]。高斯複數模糊集(Complex Gaussian membership function, cGMF)可以表示如下：

|  |  |
| --- | --- |
|  | (6) |
|  | (7) |
|  | (8) |

分別為輸入資料、中心值以及模糊集合的延展度，值為此函數的相位頻率參數。

## 類神經模型

類神經模型，是藉由模擬生物神經元的運作行為，藉由不斷接收資訊並且由錯誤中修正神經輸出，最終達到正確輸出的一種預測模型。建構類神經模型的架構即由數個類神經元組合而成，而神經元是接收外界輸入訊號，經由處理後得到輸出值。

最早將類神經網路概念實體化於1957年，Rosenblatt發明的感知器(Perceptron)，並且發表於1958年[ 26]。但是感知器只能對於利用線性分類的目標有用，並無太大的效益。有關類神經網路模型的架構如圖 7。

輸入資料

輸入層

隱藏層

輸出層

輸出資料

圖 7 類神經網路模型架構

由圖 7可見，類神經網路可分為三類，第一類為輸入層，是用以處理輸入變數，而輸入層的神經元數目視輸入資料而定。第二類為隱藏層，是表示輸入層神經元之間的交互作用，隱藏層的層數以及神經元數的設置並沒有固定方式，理論而言，越多的隱藏層以及神經元，可以讓預測效果更加完美。但是過度複雜的模型容易造成過度學習的問題產生。直至1986年，Rumelhart 和McClelland開發倒傳遞類神經網路(Back propagation neural network)。

倒傳遞演算法的預測過程，包括了正向傳遞與反向傳遞兩部分。正向傳遞是輸入資料經過輸入層抵達輸出層，並且利用成本函數(Cost function)計算與實際目標之間的誤差的過程，而後演算法再利用反向傳遞修正神經元的參數。透過不斷的正向與反向傳遞，最佳化神經元的參數，讓成本函數達到最小值。

## 蟻群演算法

自然界中有許多群居的生物，為了覓食或是演化，發展出一套能夠適應生存環境的法則。而科學家透過觀察這些方法，推導出一套相似的演算法。如基因演算法[ 28]，便是根據生物演化而來。透過模擬基因染色體的複製、選擇、變異後，逐漸變化，而較好的染色體便會被保留下來。粒子群演算法則是一種模擬鳥群的飛行以及魚類覓食現象的演算法，以單一粒子來做為鳥類族群之中的單一個體，於演算法中賦予該粒子擁有記憶性與共享的能力，並能夠透過與粒子群體中的其他粒子之間的互動而尋求到最佳解。記憶性是粒子會將搜尋到的該粒子最佳解儲存，並作為未來搜尋使用。而共享的能力則是粒子與粒子之間會互相分享粒子最佳解，提供其他粒子在搜尋解時做為參考使用。在本研究中是採用 ACO 蟻群演算法，模仿螞蟻在尋找食物的過程。

蟻群演算法是由Dorigo 於1996年所提出[ 29]，是為了解決旅行推銷員問題 (Travelling salesman problem)，假設有一系列城市和城市彼此之間的連通橋，求出以不重複的方式走過所有城市的最短路徑。相較於其他啟發式演算法可以有較為優異的表現，在許多組合最佳化的問題都能有良好的效果。其靈感是出於螞蟻的行為模式，當螞蟻在搜尋食物時會在路徑上釋放費洛蒙，費洛蒙是一種利用化學分子傳遞訊息的溝通方式，其他螞蟻便藉由路徑上的費洛蒙濃度調整移動方向。當某條路徑會引領至食物點時，螞蟻便會經由此道路前往覓食。該路徑的費洛蒙濃度便會越濃，而濃度越高越能吸引螞蟻走此路徑。而其他較少螞蟻或沒有螞蟻行走的路徑，路徑上的費洛蒙分子會隨時間蒸發。到最後螞蟻皆會採取同一種最佳路徑覓食。Dorigo利用蟻群演算法解決旅行 推銷員問題的原則步驟如下：

首先先設置座城市，而在城市與城市之間的距離為。螞蟻在時間，城市中的數量為。為所有螞蟻數量。為時間時，連接城市與城市之間的路徑強度，亦即前述之路徑費洛蒙濃度。螞蟻在時間走完後會在選擇下一次迭代所要前往的城市，而每次迭代結束，路徑上的費洛蒙濃度都會更新。費洛蒙更新公式如下：

|  |  |
| --- | --- |
|  | (9) |

是每次迭代更新完成後的費洛蒙濃度，是為了模擬現實中，路徑若是沒有新的螞蟻走過並留下費洛蒙，其濃度會隨迭代時間蒸散，所以亦稱費洛蒙蒸發係數，。因此，若是該條路徑在迭代過程中沒有新的螞蟻經過，其路徑費洛蒙會隨時間乘上蒸發係數而逐漸變小。是在時間中，新的螞蟻經過留下的費洛蒙濃度量，其公式如下：

|  |  |
| --- | --- |
|  | (10) |

是模擬第隻螞蟻在經過城市與城市的路徑時留下的費洛蒙強度，其域值如下：

|  |  |
| --- | --- |
|  | (11) |

是路徑常數，是第隻螞蟻走完的總路徑長度。費洛蒙的更新會影響下一次迭代每隻螞蟻的路徑選擇。因為螞蟻會因為不同路徑其不同的費洛蒙濃度而有不同的路徑選擇。

每次迭代結束後，螞蟻會選擇下次前往的城市，選擇的方式是利用前往不同城市路徑上的費洛蒙濃度，進行輪盤式選擇。首先依不同城市路徑的費洛蒙濃度切割輪盤上的位置，濃度值越大則該路徑在輪盤上佔有的面積也越大。每條路徑在輪盤上所佔有的面積比例也就代表該路徑被螞蟻挑選為下次前往的城市的機率，然後隨機選取輪盤上的一點，其所對應的路徑即被選中成為下次前往城市。每個可前往的城市路徑輪盤機率公式如下：

|  |  |
| --- | --- |
|  | (12) |

即為目前位於城市的第隻螞蟻，會選擇城市作為下個拜訪的城市的機率。為城市到城市其距離的倒數。與分別為對費洛蒙濃度以及城市距離的控制參數，若是越大則控制此機率公式偏向費洛蒙導向；反之，若是值越大，則控制此機率公式偏向距離導向。指的是第隻螞蟻當下所有可以前往的城市點。公式(12)涵義即為對於第隻螞蟻所有可以抵達的城市中，個別城市其費洛蒙與距離倒數的乘積值所佔之比例。而採用距離倒數是因為若是距離越長，螞蟻在此路徑上所消耗的時間成本越大，因此若是距離越短，其倒數數值會越大。讓選擇該路徑的機率越高。

### 連續型蟻群演算法

原本Dorigo所提出的蟻群演算法為離散型演算法，所要求得的最佳解區域為離散狀態。但是若要進行連續型數值搜尋，其空間是在一連續數值範圍中。過往的作法是將連續型數值進行離散化之後，在加以利用蟻群演算法進行分析。但是此方法的結果會因為對連續型數值進行離散的精細程度而有所差別。於是 Dorigo在2008年針對其蟻群演算法進行改良，並使其能夠適用於連續型範圍[ 30]。

連續型蟻群演算法相較於離散型蟻群演算法，其最大的差異是將利用離散機率分布選擇城市位置的方式(12)改為連續型，亦即採用機率密度分布的方式進行計算。

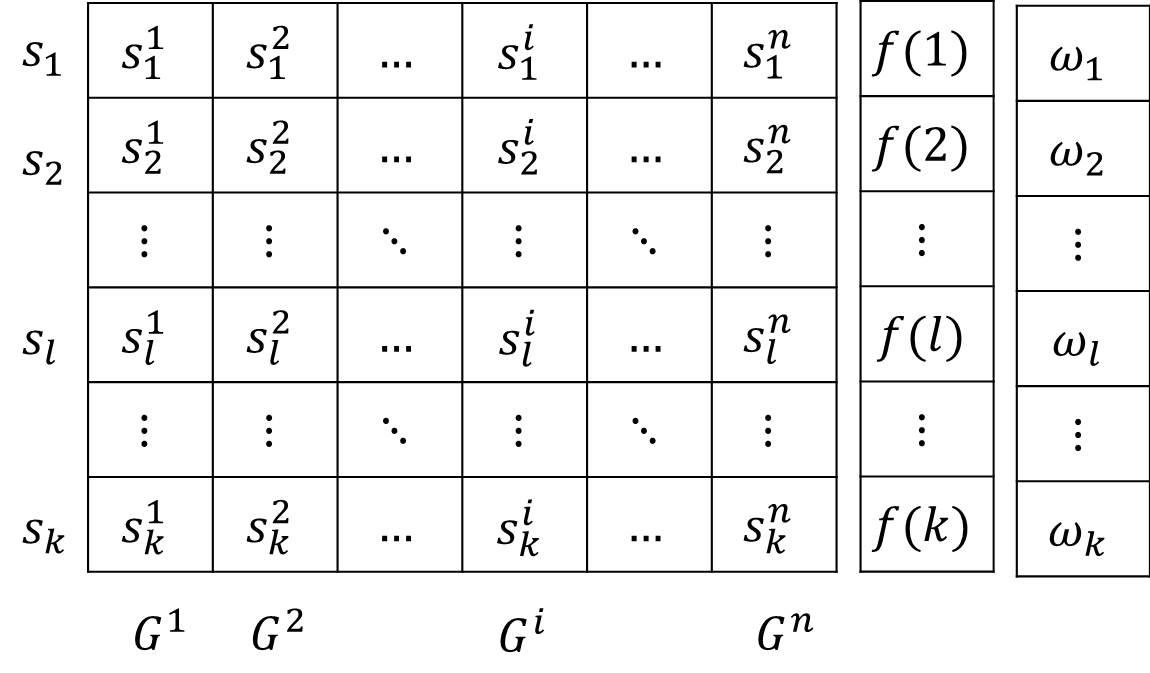


圖 8 連續型蟻群演算法費洛蒙表單

機率密度函數是指任意函數

|  |  |
| --- | --- |
|  | (13) |

機率密度函數的函數可以是任意型態，但是最常使用的為高斯函數，優勢在於可以利用一個簡易的方法取樣。Dorigo在此將機率密度函數利用於費洛蒙選取路徑的方式上。是以一種權重常態分佈密度機率函數，命為，其定義如下：

|  |  |
| --- | --- |
|  | (14) |

在(14)中包含了三個參數，分別為權重值，平均值以及標準差。這些參數皆由蟻群中的解經過處理後產生。每群蟻群中共有組蟻群解，並且儲存於一費洛蒙表單中，如圖 8。圖中每個蟻群解為，一個蟻群解代表一組前鑑部參數。每個解當中的元素記為，即為第個解的第個元素。第個蟻群解的目標成本函數為，此研究中將成本函數訂為均方根誤差(Root Mean Squared Error, RMSE)，以及各解的權重值。流程圖如圖 9，以下為執行連續型蟻群演算法的步驟：

1. 利用隨機亂數產生初始的個蟻群解，並且計算每個解的目標成本函數。
2. 依照目標成本函數的值將費洛蒙表單重新排序，排序順序是依照目標成本函數的效能表現。
3. 計算每個解的權重，公式如下：

|  |  |
| --- | --- |
|  | (15) |

即為儲存於費洛蒙表單的大小，代表有組蟻群解。為每個解經由目標成本函數排序後的順序。則為學習參數，介於0到1之間。從(15)可見，其公式類似於以1為中心值，為標準差的常態分佈機率密度函數。亦即蟻群解的順序若是越優先，則其權重值也會越高，因此，而權重值的大小會影響建構下一迭代的機率。

1. Dorigo在離散型蟻群演算法中，尋求最佳解的方式是利用蟻群不斷更新前往的節點路徑，但是連續型搜尋空間無法採用此方式，因此將其改為不斷更新搜尋解，利用新建構的解取代原本蟻群中表現較差的蟻群解。建構新的解是利用密度機率分布函數取樣的方式產生，並且僅利用一組蟻群解產生分布函數。選取製作機率密度函數的蟻群解是採用輪盤機率式的方式選取，計算權重的輪盤機率。代表費洛蒙表單中蟻群解被選入建立新蟻群解的機率。公式如下：

|  |  |
| --- | --- |
|  | (16) |

1. 利用前步驟選出的第個解的第個元素作為均值，以及標準差。其公式如下：

|  |  |
| --- | --- |
|  | (17) |

值類似於離散型蟻群演算法中的蒸發系數，其值越大，收斂速度越慢。則為先前選出的第個解。為儲存於費洛蒙表單的大小。

產生新的蟻群解方式是以各元素逐一產生的方法，針對每個元素皆會有一個機率密度函數。以本步驟產生的參數作為公式(14)所需的參數數值，以正規隨機亂數取樣的方式逐一產生新蟻群解。

1. 建構出多個新蟻群解後，取代原始費洛蒙表單中，相同數量但表現較差的蟻群解，讓迭代進行下去。

形成個蟻群解

計算蟻群解效能並排序

計算蟻群解的權重值

利用權重值計算輪盤機率

利用選中解產生參數與

建立權重常態分佈密度機率函數並取樣產生新蟻群解

取代原始較差之蟻群解

圖 9 連續型蟻群演算法流程圖

## 多群演算法

雖然現今許多演算法都擁有結構簡單、容易使用以及快速收斂的特性，如粒子群演算法。但是隨著資料維度的複雜度上升，其計算時間亦會相對上升，除此之外還會有收斂速度不穩定以及容易落入區域最佳解的問題產生。因此許多研究便針對許多演算法開發出共同演化的架構，以避免演算法結果過早收斂，並且能同時兼顧粒子的多樣性。

利用協同合作的方式開發的演算法可以推回至1991年，由Husbands所開發利用基因演算法所開發共同進化演算法[ 32]。在2004年，Engelbrecht 與Bergh將此概念應用於粒子群演算法上[ 33]。其方式是產生多個子粒子群演算法，各自演化後從每子群挑出代表性粒子進入主粒子群進行二次演化。

但是綜觀而言，Engelbrecht的多粒子群演算法對於運算時間相當冗長，並且各個子粒子群無法了解自己所蒐尋粒子解的優劣，亦即沒有資訊共享的機制。若是子粒子群落入區域最佳解，亦無法藉由參考其他子粒子群的最佳解逃離區域最佳解。

為了解決此問題，本研究在多群演算法的開發架構上，引進了學習參考機制。讓各子群體在搜尋最佳解時，可以參考其他群體的最佳解以及自身群體的最佳解，讓子群體可以互相參照。除此之外，亦有淘汰與繼承的機制。產生新的子群體以取代表現最差的子群體，但是保留被取代的子群體中的最佳解。讓取代的子群體可以有多樣性並且保留前一子群體的優點，增加演化的效益。

# 系統設計與架構

本章節將逐一說明運用於研究中之方法設計與架構。本研究利用複數模糊類神經模型結合複數模糊集、T-S模糊系統形成該模型。並利用多群連續型蟻群演算法優化模型前鑑部參數，遞迴式最小平方演算法最佳化後鑑部參數。將資料輸入模型前，亦會對資料進行特徵選取，挑選出對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。

## 特徵選取

本論文的特徵選取方式是採用過濾法，並利用夏農資訊熵(Shannon Entropy)[ 11]計算資訊的亂度。根據目標在加入特徵後目標資訊熵的變化程度，計算特徵提供給目標的資訊量。最後依照特徵的資訊量大小作為挑選特徵的依據。

夏農在1949年提出的資訊熵理論，定義了資訊內容的不確定性的量值。公式如(1)，而事件資訊的資訊熵則為。如果再加入另外一個資訊，則兩個資訊與資訊內容共同發生的資訊熵則是以來表示。公式如下：

|  |  |
| --- | --- |
|  | (18) |

是和資訊發生共同事件的機率，是事件資訊的宇集合，，是事件資訊的宇集合，。因為共同事件的發生是指同時產生，並沒有先後次序之分，因此共同事件跟發生機率會等同於，亦會等同於。及則可以改為以下公式：

|  |  |
| --- | --- |
|  | (19) |
|  | (20) |

特徵選取的目的是要了解特徵資訊所提供的資訊量，也就是目標資訊在獲得特徵資訊後資訊熵所下降的量。為了達到這個目標，首先必須先知道在事件的發生之後，目標事件的資訊熵量。此資訊熵可以命名為條件式資訊熵，也就是在有了事件發生的條件之下目標事件的資訊熵量。公式如下：

|  |  |
| --- | --- |
|  | (21) |

就是在了解事件的發生機率下，事件的發生機率，也就是。

以上關於資訊熵的定義皆為離散型，但若事件與事件為連續型事件，則離散型機率等值會改為利用機率密度函數 (Probability density function, PDF)形式表示之。

機率密度函數可以使用常態分部函數(Normal distribution function)等方式表示，本論文則是採用核心密度(Kernel Density Estimation)[ 11]估計的方式計算機率密度函數。核心密度估計公式如下：

|  |  |
| --- | --- |
| ; | (22) |

如公式(20)的，以連續型機率密度表示如下：

|  |  |
| --- | --- |
|  | (23) |

是連續型隨機事件之變數，為事件的資訊熵，為連續型事件的機率密度函數。

事件發生後事件的條件式資訊熵，以連續型態表示，如(21)，則如下表示：

|  |  |
| --- | --- |
|  | (24) |

為已知的條件下*的*資訊熵，為已知事件發生的條件下事件發生的機率。

在得知事件發生後事件的條件式資訊熵後，便可以利用(23)減去(24)後得出與的互資訊 (Mutual information)[ 37]，互資訊表示如下：

|  |  |
| --- | --- |
|  | (25) |

根據夏農資訊熵的理論，資訊熵為一非負數值。但是因為資料集的分布因素，導致在計算以及時可能產生負數值。因此對於公式(23)、(24)會加入一調整參數[ 37]，使資訊熵為一非負數。值的定義如下：

|  |  |
| --- | --- |
|  | (27) |

將加入公式(23)、(24)更新後如下：

|  |  |
| --- | --- |
|  | (28) |
|  | (29) |

由於本研究是針對時間序列資料進行特徵選取以及預測分析，預測的方式是利用時間序列的差量資料進行目標差量的預測。對於前述之事件將事件分為兩類：第一類為時間序列差量資料為正值對於目標差量的事件；第二類則為時間序列差量資料為負值對於目標差量的事件。個別表示如下：

|  |  |
| --- | --- |
|  | (25) |
|  | (26) |

如2.1.2所敘，在本研究中所要知道的資訊量即為目標原始的資訊熵量與目標在接受特徵後其資訊熵的差量。因此在計算出與後，個別計算互資訊量，最後再將這兩個互資訊量個別乘上資料機率密度分布的面積比例後在加總。該值即為特徵變數對目標變數所提供的資訊，稱為影響資訊 (Influence information)[ 37]。影響資訊可表示如下：

|  |  |
| --- | --- |
|  | (31) |
|  | (32) |
|  | (33) |

是特徵資料是正值的情況下，特徵變數對目標變數的互資訊量。則是特徵資料是負值的情況下，特徵變數對目標變數的互資訊量。為目標資料本身的資訊熵量。 是目標資料在特徵資料是正值的情況下加入特徵資訊後的資訊熵量。為目標資料在特徵資料是負值的情況下加入特徵資訊後的資訊熵量。

以上為計算特徵對於目標所提供之資訊量，進行特徵選取是為了瞭解那些特徵提供給目標的資訊量較多，進而選出有利特徵進行學習預測。而後可以利用特徵與特徵之間以及對於目標的影響資訊量整理出影響資訊矩陣(Influence information matrix)。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | Target |
|  | 0 | 0.29 | 0.18 | 0.37 |
|  | 0.48 | 0 | 0.34 | 0.25 |
|  | 0.57 | 0.28 | 0 | 0.14 |
| **Target** | 0.32 | 0.26 | 0.44 | 0 |

圖 10 資訊影響矩陣範例。  
0.25為特徵變數提供給目標Target的影響資訊量，0.28為特徵變數提供給特徵變數的影響資訊量。

特徵變數除了會對目標產生資訊量外，同時亦會對其他的特徵變數產生資訊量。因此若是要選取多個特徵進入模型學習，除了要考慮對目標的影響資訊外，亦要將特徵彼此之間產生的資訊列入考量。對於已選取的特徵對目標所帶來的資訊量已經產生效益，若是新選取的特徵其能夠給目標的資訊已經有部分由以選取的特徵提供過，則那些資訊量為冗餘資訊。在本研究中特徵對於目標資料產生的資訊量必須扣除該特徵對於已選取的特徵所產生的冗餘資訊。所以本研究對於特徵最後能夠提供給目標的資訊量定義為選取增益(Selection gain, SG)，而特徵對於已經選取的特徵所產生的資訊量定義為冗餘資訊(Redundant information, RI)。亦可利用選取增益量找出每個特徵對目標所提供增益量的排序組合，該組合定義為增益表(Gain Table)[ 37]。選取增益定義如下：

|  |  |
| --- | --- |
|  | (32) |

是候選特徵對於目標所提供的影響資訊量，則是候選特徵對於已經選取的特徵所產生的冗餘資訊量

冗餘資訊的計算是特徵對於已選特徵彼此之間的資訊量平均，其定義如下：

|  |  |
| --- | --- |
|  | (33) |

在此為即將被選入之特徵，則是已經被挑選的特徵，為已經選取的特徵中的第個。流程圖如圖 11，以下對於此特徵選取流程細部說明：

1. 定義一已選特徵池(Selected feature pool, SFP)以及一候選特徵池(Candidate feature pool, CFP)。並將所有計算完資訊量之特徵放入候選特徵池中。
2. 由於目前尚未有特徵被挑選，因此。第一個挑選的特徵為該特徵對於目標有最大資訊量的特徵，將該特徵放入已選特徵池並且將其從候選特徵池中移除。
3. 針對尚留在候選特徵池中的特徵，逐一計算其選取增益，並從中選擇選取增益最大之特徵。將其放入已選特徵池並從候選特徵池中移除。
4. 回到Step3直到所有候選特徵池中的特徵皆計算完選取增益且放入已選特徵池。而特徵進入已選特徵池的順序即為選取特徵進入模型的次序。

對於每個位於已選特徵池中的特徵皆有一選擇增益，其值代表在已有先前的特徵的情況下，對於目標的資訊量值。若選擇增益為正值，代表該特徵對於目標是帶有有效增益，反之則給目標帶來負向增益。

開始特徵選取

初始化候選特徵池(CFP)與已選特徵池(SFP)

從CFP中挑出最大的特徵放入SFP

從CFP中移除該特徵

從CFP中挑出最大的特徵放入SFP

從CFP中移除該特徵

是否檢查所有特徵

結束此特徵選取

是

否

圖 11 特徵選取流程圖

經由前次特徵選取，可以將所有對於目標帶有正向效益的特徵輸入模型進行學習，但是一旦進入模型的特徵數過多，會給模型帶來運算上的負擔。於是在本研究將針對單一目標的特徵選取以及多目標的特徵選取會在進行一次對於特徵的篩選。

### 單目標特徵選取策略

此節之單目標特徵選取將採門檻篩選的方式進行特徵選取，門檻值的設立是利用特徵的選取增益的平均值以及標準差，並挑選出選取增益大於門檻值的特徵。單目標特徵選取流程如圖 12，其詳細說明如下：

1. 定義一特徵選取池(Feature selection pool, FSP)以及一最終特徵池(Final feature pool, FFP)。並且將前次特徵選取時，位於SFP中選取增益為正數的特徵放入FSP中。
2. 設立特徵選取數目的上界以及下界，若是最後選取的特徵數目超過上界，則選取與上界數目相同的特徵數。反之，則選取與下界數目相同的特徵。
3. 計算位於FSP中的特徵其選取增益的平均值以及標準差，並且以一平均值加一標準差作為門檻值。
4. 挑選FSP中選取增益有大於門檻值的特徵，檢查是否低於下界。若選擇特徵低於下界，則依下界數量依序按照選取增益大小放入FFP。反之，則將選擇的特徵依序放入FFP中，直到達到上界。

開始單目標特徵選取

初始化特徵選取池(FSP)與最終特徵池(FFP)

將SFP中選取增益大於0的特徵放入FSP

設定特徵數量上下界

計算FSP中特徵選取增益平均值及標準差

設一平均值加一標準差為門檻值

依照門檻值及上下界選擇特徵進入FFP中

結束單目標特徵選取

圖 12 單目標特徵選取流程

### 多目標特徵選取策略

因為每次前一階段的特徵選取是將每個目標產生的所有特徵集合，再個別對不同目標值進行特徵選取，所以會有多組已選特徵池的結果。而本研究對於多目標的特徵選取策略，是採用同一特徵出現在不同已選特徵池並且其選取增益為正值的次數作為選取特徵的數目，再以對不同目標的選取增益總和大小作為選取的次序。其流程圖如圖 13，詳細說明如下：

1. 初始設立特徵選取數目的上界以及下界，若是最後選取的特徵數目超過上界，則選取與上界數目相同的特徵數。反之，則選取與下界數目相同的特徵。
2. 定義一FSP以及一FFP。選出每個SFP中的其SG是正數的特徵，並且放入FSP中。

，為已選特徵池數目，亦即目標個數。為特徵。

1. 列出所有FSP中選出的特徵其涵蓋的目標數目，並以涵蓋數目最大的作為選取特徵數(Number of feature selected, NFS)。若涵蓋數目太小或太大，則以上下界為NFS。
2. 計算FSP中的特徵個別對每個目標的SG總和，記為。即為欲計算的特徵，表示如下：

|  |  |
| --- | --- |
|  | (34) |

是該特徵對於目標的SG。

1. 將由大至小順序排列，並且挑選出NFS個特徵放入FFP中。

開始多目標特徵選取

設定特徵數量上下界

將各個SFP中選取增益大於0的特徵放入FSP

初始化FSP與FFP

列出所有FSP中選出的特徵其涵蓋的目標數目並設定NFS

計算FSP中的特徵個別對每個目標的SG總和

挑選出NFS個特徵放入FFP

圖 13 多目標特徵選取流程

## 複數模糊類神經模型

1965年，Zadeh發表了模糊集合的概念，將傳統集合特徵函數「不屬於」與「屬於」的二值選擇，推廣為可從0到1之間的任何值作選擇，作為一個元素屬於一個集合的程度，稱之為歸屬程度。歸屬程度越高，代表該元素屬於該集合的程度越高，反之則越低。而 Ramot在2002年將原本一維的實數歸屬程度拓展到二維單位圓盤上，使歸屬程度能夠有實數部與虛數部的值域，增加歸屬程度的描述力與自由度。

由於複數模糊集相較於實數模糊集有更佳的描述能力，因此更能處理複雜問題。而其歸屬程度的輸出有實數部以及虛數部，本研究將利用此歸屬函數進行多目標之預測，假設現有一歸屬程度如下

|  |  |
| --- | --- |
|  | (35) |

在此為虛數，。因此可以將解構出多個歸屬程度輸出，表示如下：

|  |  |
| --- | --- |
|  | (36) |

是擷取該值的實數部位置數值，則是擷取該數值的虛數部位置數值。由(36)可見，利用其原本的歸屬程度值以及各自擷取歸屬程度的實數部以及虛數部，可以將原本的歸屬程度分成三個不同的歸屬程度。而使用一組輸入資料並且產生多組歸屬程度，亦即代表可以進行多個不同的目標輸出，增加模型的預測能力。

本研究中，對於模糊系統所建立的方式為利用Takagi-Sugeno (T-S) fuzzy 模型作為模糊系統的依據。T-S模糊模型最早是由Takagi與Sugeno於1985年提出[ 31]，以一個複合式非線性系統並藉由一系列的If-Then模糊規則組合而成。If-Then模糊規則類似於模擬人類的經驗，因而更容易被人類所理解。此外，類神經網路所採用的分散式架構能夠有效處理大量資料，並從其中找到隱藏的資訊。而且不會因為資料的增加而擴大類神經網路模型的構造。因此，本研究將複數模糊集、模糊系統以及類神經網路模型結合，設計出可用於多目標預測之複數類神經網路模型。該模型說明如下：

假設有一輸入資料，為輸入資料向量，則是預測目標值。並且由條T-S模糊規則所組成的預測模型，其規則形式如下：

|  |  |
| --- | --- |
|  | (37) |

規則數，為模糊系統的輸入資料，為第條規則的複數模糊集，則是輸入的語意變數。即為後鑑部(Then part)參數，而複數模糊集的參數即是前鑑部(If part)參數。此複數模糊類神經模型可轉為一個六層架構的類神經網路模型，如圖 14，以下將對各層進行說明。

第1層

輸入層

第2層

複數模糊集層

第3層

啟動強度層

第4層

正規化層

第5層

後鑑部層

第6層

輸出層

前鑑部參數

後鑑部參數

圖 14 複數模糊類神經系統模型

第一層 (Layer 1)：此層為輸入層，目的是將資料輸入系統模型並傳遞給下一層。輸入資料是從時間序列歷史資料而來，該層在迭代次數時的輸入資料向量表示如下：

|  |  |
| --- | --- |
|  | (38) |

第二層 (Layer 2)：此層為複數模糊集層，此層中每一神經元代表一複數模糊集合。本研究在複數模糊集合的歸屬函數採用高斯型複數歸屬函數，公式如(6)。經由歸屬函數計算可得知輸入資料屬於該複數模糊集合的歸屬程度，歸屬程度又可以經由分析其值，得到三個不同的歸屬程度，進行多目標運算。在此層中，所有高斯型複數模糊集合的中心值、延展度以及相位頻率函數，合為此模型系統的前鑑部參數。

第三層 (Layer 3)：此層為啟動強度層，目的是計算每條規則的啟動強度。本研究對於前鑑部的歸屬程度邏輯運算採用積運算，每條模糊規則啟動強度計算如下：

|  |  |
| --- | --- |
|  | (39) |

代表第條規則的第個複數模糊集的歸屬程度，代表解構歸屬程度得到的不同歸屬程度值，。

第四層 (Layer 4)：此層為正規化層，將每條規則得到的啟動強度進行正規劃計算。目的是調整比例值，用以了解每條規則佔總輸出的比例。計算正規化啟動強度如下：

|  |  |
| --- | --- |
|  | (40) |

第五層 (Layer 5)：此層為後鑑部層，此層利用前一層所得到的正規化啟動強度計算每條規則在後鑑部的反應值。經由公式(39)與(40)，可以得知每條規則正規化後的比例輸出值。公式如下：

|  |  |
| --- | --- |
|  | (41) |

第六層 (Layer 6)：此層為輸出層，將前一層每條規則的正規化輸出值加總，即為模型最終的總輸出。

|  |  |
| --- | --- |
|  | (42) |

系統輸出可視為一向量，因為針對多目標輸出，系統輸出則為多個值。。此外，亦為一複數值，如下表示：

|  |  |
| --- | --- |
|  | (43) |

代表系統輸出的實數部，代表系統輸出的虛數部數值，。

得到系統輸出值後，會與目標資料值進行比較得到誤差值，利用誤差值計算成本函數，並利用學習演算法調整參數值，使最後誤差達到最小值。誤差值計算方式如下：

|  |  |
| --- | --- |
|  | (44) |

在下一節會介紹用於優化第二層複數模糊集層的高斯複數模糊歸屬函數的多群連續型蟻群演算法。

## 多群連續型蟻群演算法

在原本Dorigo所提出之連續型蟻群演算法，是以單群體進行演化。但若陷入區域最佳解，該演算法並沒有跳脫機制使新產生的蟻群解離開區域最佳解。因此，本研究提出一新概念，以多群的方式進行連續型蟻群演算法的演化，增加蟻群解之多樣性。並且在演化的過程中，讓各群的搜尋結果可以互相分享。並且不斷淘汰表現較差之群體，但是並非完全淘汰。仍會保留被淘汰之群其優良解，讓之前的優良結果不會消失。以下便會針對資訊流通、淘汰、繼承等三個面向對本研究之多群連續型蟻群演算法進行闡述。

### 資訊流通

單群連續型蟻群演算法進行選擇蟻群解作為建構新的蟻群解時，利用權重機率進行選擇，但是當該群權重較大的蟻群解為區域最佳解時，該解有較大的機率被選入建構新蟻群解。為了解決區域最佳解的問題，本研究採用多群的方式進行連續型蟻群演算法。當利用多群演化時，群體與群體彼此之間的資訊共享便顯得重要。若是各群獨立搜尋最佳解，給模型帶來的只有運算資源的消耗，效率反而不比單群演算法。因此本研究在多群的資訊共享方面，融入了類似於粒子群演算法中，兩個對於粒子位置的影響因素，個體最佳位置()與群體最佳位置()。運用於多群連續型蟻群演算法中，為利用權重機率選到的該群蟻群解，則為所有蟻群中，成本函數最低的蟻群解。其流程圖如圖 15。

設定利用及取代較差蟻群解之比例

分別利用及建立新蟻群解

取代群體中較差蟻群解

圖 15 多群連續型蟻群演算法各群取代較差蟻群解流程

於取代蟻群時，會產生多個新的蟻群解，用於取代原本蟻群解中，表現較差之蟻群。而此方法，是在產生的多個新的蟻群解中，部分蟻群解是利用產生，其餘利用產生。利用此方法，讓蟻群能夠同時針對全體最佳解以及自群體最佳解分別進行搜尋，增加找到全域最佳解的機會。

### 淘汰

利用多群演算法進行搜尋時，難免會有整個表現較差之群體，若是繼續讓該群體搜尋解，即便有機會找到最佳解，但是仍然會消耗許多不必要的計算資源。於是本研究採用淘汰機制，在每次迭代結束後會產生全新的一群蟻群解，並且根據當下進行搜尋的所有蟻群中，各蟻群表現最佳的成本函數作為評斷標準，以成本函數最差的蟻群做為被淘汰之蟻群。並將該蟻群以新蟻群完全取代之。類似於將陷入區域最佳解的群體重新初始化用以逃離區域最佳解。淘汰機制流程如圖 16。

建立新蟻群

找出表現效能最差之蟻群

取代該蟻群

圖 16 多群連續型蟻群演算法淘汰最差蟻群流程

加入淘汰機制於蟻群演算法中，可以在維持蟻群解多樣性的前提下，減少有蟻群落入區域最佳解的機會。

### 繼承

如3.3.2提及，加入淘汰機制是為了避免蟻群解落入區域最佳解。但是直接取代該群就浪費了該群先前所進行的搜尋。於是在取代蟻群時，會加入繼承的機制。本研究演算法在產生新的蟻群時，會先對該次迭代存在的各蟻群蒐集各自部分最佳解，成為繼承蟻群解。並利用繼承蟻群解的權重機率挑出部分繼承蟻群解用於產生新蟻群解，使繼承蟻群解及產生的新蟻群解總數與被取代蟻群之蟻群解數目相同。繼承機制如圖 17。讓新產生的蟻群繼承了各蟻群的最佳解，加速搜尋的進行。此機制的優勢在於被取代之蟻群在搜尋上的資源消耗不會浪費，並且可以結合各群較佳的搜尋結果，使新的蟻群表現更為優異。

從不同蟻群中，挑出部分蟻群解用於產生新蟻群

蟻群一

蟻群二

蟻群三

新產生蟻群

利用蟻群解的權重值進行輪盤式挑選，  
並用於建立新蟻群所不足之蟻群解

圖 17 繼承機制產生新蟻群方式

將以上三個機制融入多群連續型蟻群演算法，使其在演化的過程中，可以保有多樣性並且能避免落入區域最佳解。整體演算法流程如圖 17。流程概述如下：

1. 初始組蟻群，每組蟻群中包含個不同的蟻群解。
2. 個別計算每組蟻群中每個蟻群解的效能並且利用成本函數值進行蟻 群解排序。
3. 個別計算每組蟻群中各蟻群解的權重值。
4. 各組蟻群中蟻群解利用權重值建構該蟻群解的輪盤機率。
5. 各組各自利用輪盤機率選到該組蟻群中的蟻群解並且建構機率函 數。
6. 各組利用各自機率函數取樣產生新蟻群解並各自取代該組蟻群中相 同數量的較差蟻群解。
7. 各組蟻群皆取代蟻群解完畢後，擷取各組蟻群中表現較為優異的部 分蟻群解作為繼承蟻群解用於產生新蟻群。
8. 利用繼承蟻群解的權重機率挑出一繼承蟻群解並建立機率函數。
9. 根據機率函數取樣產生一組新蟻群。
10. 將各組蟻群中，以各組的最佳蟻群解作為評斷標準，以性能表現最 差蟻群解的所屬蟻群作為被淘汰的蟻群。並完全刪除該蟻群。
11. 以新產生的蟻群取代原本被刪除的蟻群，使所有蟻群組數量保持一 定。

形成組蟻群

計算各組蟻群中蟻群解效能並排序

計算蟻群解的權重值

利用權重值計算輪盤機率

利用選中解產生參數與

建立權重常態分佈密度機率函數並取樣產生新蟻群解

取代相同數量但較差之蟻群解

蒐集各群優異蟻群解

建立新蟻群

以新蟻群代替被取代之蟻群

完全刪除效能最差之蟻群

圖 18 多群連續型蟻群演算法演化流程

## 遞迴式最小平方演算法

遞迴式最小平方演算法(Recursive least squares estimation, RLSE)是由最小平方估計演算法演化而來，最小平方估計法是利用輸入資料點，尋找一線性函數，使資料點與該函數的平方誤差達到最小值，藉此找到符合資料關係的函數。假定有一輸入資料集，其線性函數模型如下表示：

|  |  |
| --- | --- |
|  | (45) |

為該模型輸入，則為的函數，則為需要尋找的參數，則是誤差值。假設共有筆模型輸入資料，記為。將其利用(45)表示，可得一組線性方程式集合，如下表示：

|  |  |
| --- | --- |
|  | (46) |

將上述方程式集合以矩陣形式表示，其結果如下：

|  |  |
| --- | --- |
|  | (47) |

為模型輸出矩陣，為輸入矩陣，為需要尋找的參數矩陣，為誤差矩陣。各自表示如下：

|  |  |
| --- | --- |
|  | (48) |
|  | (49) |
|  | (50) |
|  | (51) |

對於，利用最小平方估計演算法可得以下公式：

|  |  |
| --- | --- |
|  | (52) |

但以此方式進行運算會造成龐大的運算量。而遞迴式最小平方演算法則是以遞迴方式循序漸進計算，可以降低運算量並增加效率。在遞迴次數為時，參數矩陣為。遞迴最小平方演算法定義如下：

|  |  |
| --- | --- |
|  | (53) |
|  | (54) |

其中是遞迴次數，， 為投影運算元(Projection operator)或稱之為增益矩陣(Gain matrix)。是RLSE在第次遞迴計算後得出的估計參數值。在(48)、(49)中，第組，即為。進行遞迴演算法前須先將與初始化。初始成所需估計參數量之零向量，而的初始化定義如下：

|  |  |
| --- | --- |
|  | (55) |

設為一極大整數，則為單位矩陣。RLSE運行的原理為利用前一次估計出的結果值，加入當次觀測的輸入以及輸出值，進而計算當次的估計結果值。其流程如圖 18。

RLSE演算法

及

估計值及

觀測值與

成為下次遞迴之及

圖 19 遞迴式最小平方演算法流程圖

## MGCACO-RLSE複合演算法

建構完複數模糊類神經網路模型系統後，利用輸入資料可得系統輸出值，並與目標相比獲得誤差值，系統便根據誤差值的大小調整前鑑部與後鑑部參數。本研究結合多群連續型蟻群演算法(Multi-group continuous ant colony optimization, MGCACO) 以及遞迴式最小平方演算法(Recursive least squares estimation, RLSE)，提出MGCACO-RLSE的複合型演算法。將模型中前鑑部的高斯複數模糊集和參數以MGCACO優化，後鑑部的規則參數以RLSE尋找參數解。利用此方式可以將系統學習時間降低，增加效率。該複合型演算法流程如圖 19 複合式演算法學習流程圖，詳細的演算法流程如下：

1. 將蒐集的時間序列資料分為兩部分，一部分是訓練資料，另一部分則為測試資料。
2. 初始多群連續型蟻群演算法以及遞迴最小平方演算法的參數。由於蟻群演算法是用於優化前鑑部複數高斯模糊集的中心值、延展度以及相位頻率函數。為了讓高斯模糊集對輸入資料有良好的覆蓋性，則以輸入資料的平均值、標準差以及常數1乘上正規隨機亂數的絕對值作為初始蟻群解的數值。建構第個蟻群解的方式如下：

|  |  |
| --- | --- |
|  | (56) |
|  | (57) |

1. 利用訓練資料輸入模型並計算每條規則的正規化啟動強度。
2. 利用遞迴式最小平方演算法更新模型後鑑部參數。公式(53)中的定義如下：

|  |  |
| --- | --- |
|  | (58) |
|  | (59) |

當中代表規則數指標，是每條規則正規化之後的啟動強度，根據(54)可得後鑑部參數集。因為多目標預測中，正規化啟動強度為一向量，，且為一矩陣，因此在原本的公式(53)中，利用單位矩陣取代原本的常數項1，改良後公式如下：

|  |  |
| --- | --- |
|  | (60) |
|  | (61) |
|  | (62) |

，為第組資料進入RLSE後產生對於第條規則的後鑑部規則參數。

1. 計算模型輸出，並與預測目標的差量計算誤差值。方式如下：

|  |  |
| --- | --- |
|  | (63) |

是在時間的預測目標，則是在時間的類神經模型預測值。

1. 利用誤差值計算均方根誤差(Root-mean-square error, RMSE)，RMSE計算如下。為的共軛轉置矩陣。

|  |  |
| --- | --- |
|  | (64) |

1. 回到Step3直到完成該群所有蟻群解的計算。
2. 更新以及，並且更新該群蟻群解。
3. 回到Step3直到完成所有蟻群的蟻群解計算。
4. 產生新的蟻群並且用以取代表現最差之蟻群，各蟻群效能表現以 RMSE作為指標，表現最差即為該群RMSE值最高。
5. 檢查是否達到停止條件，例如已達最高迭代次數。若達到停止條件， 即為利用複數模糊類神經模型所優化之前鑑部參數。否則回 到Step3再次計算。

RLSE

Model

MGCACO

初始化蟻群解

將蟻群解作為前鑑部參數

計算蟻群解RMSE

更新與

該群所有蟻群解皆計算

所有族群皆計算

更新蟻群解

建立新蟻群取代表現較差之蟻群

達到迭代停止條件

第一層

輸入層

第二層

複數模糊集層

第三層

啟動強度層

第四層

正規化層

第五層

後鑑部層

第六層

輸出層

初始,

輸入觀測值

與

更新

更新

設置為後鑑部參數

結束學習

是

是

是

否

否

否

圖 20 複合式演算法學習流程圖

# 實驗

本研究採用三個實驗驗證MGCACO-RLSE測試優化複數模糊類神經模型的可行性以及利用複數型態的歸屬函數讓模型有能力預測多目標和預測的準確度。於時間序列的預測中，測試資料的誤差是用於判斷該模型效能的一個重要指標。測試的誤差越小，越能顯現該模型的優異處。於實驗中，將會與部分文獻[ 34][ 35][ 36]進行效能比較，用以證實本模型之優劣。使用到的效能指標有絕對誤差百分比(Mean Absolute Percentage Error, MAPE)以及均方根誤差(RMSE)，效能指標公式如下：

|  |  |
| --- | --- |
|  | (65) |
|  | (66) |

於實驗一，進行類神經模糊模型對於單一目標的預測。目的是為了獲得初步的證實，驗證MGCACO-RLSE優化模型參數的可行性。實驗二則是進行雙目標的預測，因為複數模糊類神經模型的輸出可以為一複數，利用該複數的實數部以及虛數部個別對不同目標預測，驗證複數輸出預測不同目標的可行性。實驗三是進行三個不同目標的預測實驗，如3.2所提及，一複數歸屬程度可以解構出不同的歸屬程度值，並且利用不同的歸屬程度值對不同目標進行預測。實驗三採用兩個不同歸屬程度值，讓模型輸出兩個複數數值。以第一個數值的實數、虛數以及第二個數值的實數部，個別對不同目標預測。證實模型多目標輸出預測的可行性。

## 實驗一：道瓊工業指數時間序列預測

本實驗是進行單目標的時間序列預測，實驗目標是建立模型，訓練後並且預測每日股價指數該股價指數為道瓊工業平均指數(Dow Jones Industrial Average Index)。模型輸出為一複數值，取其實數部為模型預測道瓊指數的數值。

道瓊工業平均指數是由華爾街日報(The Wall Street Journal)以及Dow於1896年開始公布。該指數包含美國30間最為知名的上市公司，足以反映總體經濟市場狀況。本實驗擷取1993年一月至2010年十二月作為實驗所使用的資料，最後60筆為測試資料，其餘的作為訓練資料。在實驗的學習過程中，每個蟻群解的效能優劣以RMSE作為性能指標。

訓練資料至模型前會先將資料進行前處理後再執行特徵選取，選取目標交易月前30個交易月的指數漲跌變化作為候選特徵，此30個交易月是從最遠離目標交易日漲跌依序靠近，記為。候選特徵的選擇增益則是基於對目標資料提供的資訊量及對於已選擇的候選特徵彼此之間的冗餘資訊量，此實驗的候選特徵對於目標的候選特徵選擇順序於表 1，候選特徵對於目標的總選擇增益量變化如圖 20。該圖為隨著選擇特徵的數量增加，對於目標的總選擇增益量變化，從圖中可見，約選擇到第14個特徵時，總選擇增益量最大。但是若採用14個特徵集合進入模型學習，會造成模型龐大的計算負擔。於是本實驗會採用單目標特徵選取策略針對特徵再次進行篩選。

經過特徵選取後，為被選入模型的候選特徵變數，為第30個以及第二個特徵，此二特徵個別為目標交易月第30個月前以及兩個月前的指數漲跌變化量。預測模型的輸入資料為**，**是前述輸入特徵資料的向量，則是模型所要預測的道瓊工業指數目標資料，則為資料筆數。定義如下：

|  |  |
| --- | --- |
|  | (67) |

每個輸入在複數模糊集層皆有三個複數高斯複數模糊集，所以總共劃分出九個前鑑部區域。後鑑部則是採用T-S規則的方式，根據高斯模糊集數量總共建立了九條規則。複數模糊類神經模型的設定如表 2。

本實驗是採用MGCACO-RLSE的複合型演算法訓練資料集，其設定如表 3，在MGCACO中的蟻群解，其位置初始方式如3.5所提及，以輸入資料的平均值、標準差以及常數1乘上正規隨機亂數的絕對值作為初始蟻群解的數值。以此方式便能良好覆蓋資料集，增加效率。

為了測試此模型與演算法的穩定性，本實驗一共進行十次試驗，呈現的結果為學習過程中表現最佳的第2次試驗。該次試驗的學習後前鑑部以及後鑑部參數值如表 4、表 5。在優化參數的過程中，最佳蟻群解效能表現亦為學習曲線(Learning curve)，表示如圖 23。模型對於道瓊工業指數的預測值以及實際指數值如圖 21，實際值以及預測值之間的誤差量如圖 22。相較於比較文獻[ 34]與本研究的實驗結果於表 6。

表 1 實驗一之特徵選取次序、特徵變數、選取增益及總選取增益-  
以道瓊工業指數為目標(節錄前十選擇次序)

|  |  |  |  |
| --- | --- | --- | --- |
| 選擇次序 | 特徵 | 選取增益 | 總選取增益 |
| 1 |  | 0.0237 | 0.0237 |
| 2 |  | 0.0281 | 0.0518 |
| 3 |  | 0.0160 | 0.0677 |
| 4 |  | 0.0156 | 0.0834 |
| 5 |  | 0.0144 | 0.0977 |
| 6 |  | 0.0128 | 0.1105 |
| 7 |  | 0.0105 | 0.1210 |
| 8 |  | 0.0077 | 0.1287 |
| 9 |  | 0.0054 | 0.1341 |
| 10 |  | 0.0047 | 0.1389 |



圖 21 輸入特徵數對於目標總選取增益量變化

表 2 實驗一複數類神經模型設定

|  |  |
| --- | --- |
| 類神經模型設定 |  |
| 輸入特徵 |  |
| 學習次數 | 50 |
| 輸出值數量(複數型態) | 1 |
| 前鑑部個數 | 9 |
| 前鑑部參數總個數 | 18 |
| 後鑑部規則個數 | 9 |
| 後鑑部規則參數總個數 | 27 |

表 3 實驗一MGCACO-RLSE複合式演算法參數設定

|  |  |
| --- | --- |
| MGCACO |  |
| 蟻群組數量 | 3 |
| 每蟻群組回傳蟻群解數量 | 5 |
| 每個蟻群組中蟻群解數量 | 30 |
| 每個蟻群組更新蟻群解數量 | 20 |
| 更新蟻群解比例 | 40% (取代8個蟻群解) |
| 更新蟻群解比例 | 60% (取代12個蟻群解) |
| RLSE |  |
| 後鑑部參數數量 | 27 |
|  | 109 |
|  | 27\*27 單位矩陣 |
|  |  |
|  | 27\*1 零矩陣 |



圖 22預測結果-以道瓊工業指數為目標:  
藍色實線為實際目標值，紅色虛線為模型預測值



圖 23 模型預測誤差值-以道瓊工業指數為目標



圖 24 實驗一模糊類神經模型學習曲線

表 4 實驗一學習後前鑑部參數值

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 前鑑部參數 | | | | | | | | |
|  |  | | |  |  |  | | |
| CGFS # | c | σ | λ | CGFS # | c | σ | λ |
| 1 | 72.5187 | 18.1860 | 126.2135 | 4 | -6.9879 | 78.9989 | -4.9252 |
| 2 | 42.2308 | 85.9688 | 266.2049 | 5 | 9.7063 | -69.1629 | -161.2525 |
| 3 | 78.2718 | -5.4354 | 504.5853 | 6 | -10.5809 | -77.1484 | -86.0750 |

表 5 實驗一學習後後鑑部參數值

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | | | | |
|  |  |  |  |  |  |  |
| 1 |  | 8.0025E-23+7.275E-24i |  | 1.67613E-20-2.328E-21i |  | 1.3968E-20-4.656E-21i |
| 2 |  | 3.9785E-24+1.5914E-24i |  | 6.693E-22-1.455E-22i |  | 2.037E-21+1.164E-22i |
| 3 |  | 3.9558E-23+1.364E-24i |  | 1.1523E-20-9.312E-22i |  | 7.2168E-21+6.984E-22i |
| 4 |  | 0.0004+0.0004i |  | 0.1034+0.103i |  | 0.1183+0.0905i |
| 5 |  | 0.000009+0.00002i |  | 0.0045+0.0054i |  | 0.0111+0.0119i |
| 6 |  | 0.0002+0.0001i |  | 0.0474+0.0615i |  | 0.0527+0.031i |
| 7 |  | 1.0123E-211-2.939E-212i |  | 2.8006E-209-6.688E-210i |  | 3.5112E-209+8.36E-210i |
| 8 |  | 6.3271E-213-4.082E-214i |  | 9.9275E-211+3.3962E-211i |  | 2.8737E-210-1.045E-211i |
| 9 |  | 2.7757E-212-4.082E-213i |  | 8.778E-210+6.27E-211i |  | 1.5466E-209+3.344E-210i |

表 6 實驗一效能比較-以道瓊工業指數為目標[ 34]

|  |  |  |
| --- | --- | --- |
| 方法 | RMSE | MAPE |
| ESM | 5984.2 | 47.61% |
| ARIMA | 5715.4 | 41.3% |
| BPNN | 5821.4 | 38.97% |
| EWH | 5061.4 | 34.71% |
| PHM | 4356.6 | 30.53% |
| RWM | 6178.2 | 51.29% |
| MGCACO-RLSE | 517.21 | 3.93% |

## 實驗二：高盛與微軟股價資料預測

於實驗二是進行雙目標的預測，實驗目標是建立模型，訓練後並且將其輸出之複數資料，以實部數值和虛部數值分別預測兩種不同目標的日價格。實數部的數值用於預測高盛集團(Goldman Sachs)股價資料，虛數部數值預測微軟(Microsoft)股價資料。

高盛是一間跨國銀行控股公司集團，由Marcus成立於1869年。其業務是向全球的投資者提供投資、諮詢和金融服務，目前為美國財富企業前100強，是位於全球領導地位的投資銀行；微軟是一間跨國電腦科技公司，由Gates與Allen創立於1975年，其在全球個人電腦硬體與軟體、服務與網際網路技術上居領導地位。微軟未來更朝向人工智慧，如語音辨識的技術進行開發。是科技資訊產業中極為重要的全球領導企業。

本實驗針對此二企業，皆擷取自2010年一月四日至2012年12月31日止，這段期間的股價資料。當中前80%為訓練資料，後20%為測試資料。實驗過程中以RMSE作為蟻群解的效能指標。

在進行特徵選取的前處理時，各自選取目標交易日前30個交易日的股價漲跌變化作為候選特徵，因此共有60組候選特徵，記為。至為高盛集團的特徵資料，至是微軟的特徵資料。各自依序由最遠離目標交易日的資料依序向目標交易日靠近。此60個特徵對於高盛集團做為目標時的選擇增益量如表 7，該選擇增益增減量圖如圖 24；以微軟做為目標時的選擇增益量如表 8，該選擇增益增減量圖如圖 25。由圖 24及圖 25可見，兩個目標大約各自在選取20及30個特徵時，其選取增益可以達到最大值。但若選擇如此大量的特徵進入模型會造成模型學習上的負擔。因此在本實驗會採用多目標特徵選取策略針對特徵再次進行選取。

結束特徵選取後，為被選入模型的候選特徵變數，為第60個以及第5個特徵，此二特徵個別為微軟的一天前股價變化差量以及高盛第26天前股價變化差量。預測模型的輸入資料為**，**是前述輸入特徵資料的向量，為一複數目標資料，其實數部為高盛集團目標資料，複數部為微軟股價目標資料，則為資料筆數。定義如下：

|  |  |
| --- | --- |
|  | (68) |

類似於實驗一，此研究的每個輸入在複數模糊集層皆有三個複數高斯複數模糊集，所以總共劃分出九個前鑑部區域。根據高斯模糊集數量總共建立了九條後鑑部T-S規則。複數模糊類神經模型的設定如表 9。

MGCACO-RLSE演算法用於訓練資料集，其設定如表 10，在MGCACO中的蟻群解，其位置初始方式以訓練資料的平均值、標準差以及常數1乘上正規隨機亂數的絕對值作為初始蟻群解的數值。

為了測試此模型與演算法的穩定性，本實驗一共進行十次試驗，呈現的結果為學習過程中表現最佳的第3次試驗。該次試驗的前後鑑部參數值在學習完成後的直如表 11、表 12。該次試驗的學習曲線表示如圖 30。模型預測高盛集團股價的預測值以及實際指數值如圖 26圖 21，誤差量如圖 27；預測微軟股價的模型預測值以及實際指數值如圖 28圖 21，誤差量如圖 29。實驗結果與比較文獻[ 35]的差異表現於表 13、表 14。

表 7實驗二之特徵選取次序、特徵變數、選取增益及總選取增益-  
以高盛集團為目標(節錄前十選擇次序)

|  |  |  |  |
| --- | --- | --- | --- |
| 選擇次序 | 特徵 | 選取增益 | 總選取增益 |
| 1 |  | 0.0544 | 0.0544 |
| 2 |  | 0.0113 | 0.0657 |
| 3 |  | 0.0110 | 0.0767 |
| 4 |  | 0.0107 | 0.0874 |
| 5 |  | 0.0093 | 0.0967 |
| 6 |  | 0.0087 | 0.1054 |
| 7 |  | 0.0065 | 0.1119 |
| 8 |  | 0.0062 | 0.1181 |
| 9 |  | 0.0058 | 0.1239 |
| 10 |  | 0.0054 | 0.1293 |



圖 25輸入特徵數對於目標總選取增益量變化-以高盛集團為目標

表 8實驗二之特徵選取次序、特徵變數、選取增益及總選取增益-  
以微軟為目標(節錄前十選擇次序)

|  |  |  |  |
| --- | --- | --- | --- |
| 選擇次序 | 特徵 | 選取增益 | 總選取增益 |
| 1 |  | -0.0001 | -0.0001 |
| 2 |  | 0.0101 | 0.0100 |
| 3 |  | 0.0079 | 0.0179 |
| 4 |  | 0.0053 | 0.0232 |
| 5 |  | 0.0043 | 0.0275 |
| 6 |  | 0.0048 | 0.0323 |
| 7 |  | 0.0058 | 0.0381 |
| 8 |  | 0.0053 | 0.0434 |
| 9 |  | 0.0047 | 0.0480 |
| 10 |  | 0.0043 | 0.0523 |



圖 26輸入特徵數對於目標總選取增益量變化-以微軟集團為目標

表 9實驗二複數類神經模型設定

|  |  |
| --- | --- |
| 類神經模型設定 |  |
| 輸入特徵 |  |
| 學習次數 | 50 |
| 輸出值數量(複數型態) | 1 |
| 前鑑部個數 | 9 |
| 前鑑部參數總個數 | 18 |
| 後鑑部規則個數 | 9 |
| 後鑑部規則參數總個數 | 27 |

表 10實驗二MGCACO-RLSE複合式演算法參數設定

|  |  |
| --- | --- |
| MGCACO |  |
| 蟻群組數量 | 3 |
| 每蟻群組回傳蟻群解數量 | 5 |
| 每個蟻群組中蟻群解數量 | 30 |
| 每個蟻群組更新蟻群解數量 | 20 |
| 更新蟻群解比例 | 40% (取代8個蟻群解) |
| 更新蟻群解比例 | 60% (取代12個蟻群解) |
| RLSE |  |
| 後鑑部參數數量 | 27 |
|  | 109 |
|  | 27\*27 單位矩陣 |
|  |  |
|  | 27\*1 零矩陣 |



圖 27預測結果-以高盛集團股價為目標:  
藍色實線為實際目標值，紅色虛線為模型預測值



圖 28模型預測誤差值-以高盛集團股價為目標



圖 29預測結果-以微軟股價為目標:  
藍色實線為實際目標值，紅色虛線為模型預測值



圖 30模型預測誤差值-以微軟股價為目標



圖 31 實驗二模糊類神經模型學習曲線

表 11實驗二學習後前鑑部參數值

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 前鑑部參數 | | | | | | | | |
|  |  | | |  |  |  | | |
| CGFS # | c | σ | λ | CGFS # | c | σ | λ |
| 1 | -0.0047 | 10.4316 | -2.2936 | 4 | -0.0421 | 4.0135 | -1.0310 |
| 2 | 0.0078 | 1.2615 | 0.2928 | 5 | -0.0306 | 0.1138 | 2.0798 |
| 3 | 0.0132 | -4.0759 | -0.5799 | 6 | 0.0076 | 0.6995 | -9.4627 |

表 12實驗二學習後後鑑部參數值

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | | | | |
|  |  |  |  |  |  |  |
| 1 |  | 0.000352-0.007935i |  | 0.000933-0.020958i |  | 0.000004-0.000079i |
| 2 |  | 0.008795+0.021723i |  | 0.023207+0.057332i |  | 0.000088+0.000217i |
| 3 |  | -0.002283-0.012064i |  | -0.006027-0.031851i |  | -0.000023-0.000121i |
| 4 |  | 0.000819+0.000291i |  | 0.002162+0.000768i |  | 0.000008+0.000003i |
| 5 |  | 0.000301+0.001403i |  | 0.000793+0.003705i |  | 0.000003+0.000014i |
| 6 |  | 0.000511-0.000193i |  | 0.001349-0.00051i |  | 0.000005-0.000002i |
| 7 |  | -0.000755-0.008305i |  | -0.001991-0.021925i |  | -0.000008-0.000083i |
| 8 |  | 0.00836+0.019757i |  | 0.022074+0.052163i |  | 0.000084+0.000198i |
| 9 |  | -0.002966-0.011767i |  | -0.007829-0.031062i |  | -0.00003-0.000118i |

表 13實驗二效能比較-以高盛集團為目標 [ 35]

|  |  |  |
| --- | --- | --- |
| 方法 | RMSE | MAPE |
| Bayesian regularized ANN | N/A | 1.3291% |
| MGCACO-RLSE | 1.796 | 1.2612% |

表 14實驗二效能比較-以微軟為目標 [ 35]

|  |  |  |
| --- | --- | --- |
| 方法 | RMSE | MAPE |
| Bayesian regularized ANN | N/A | 1.0561% |
| MGCACO-RLSE | 0.3871 | 1.0297% |

## 實驗三：巴西股市指數、日經平均指數、道瓊工業指數時間序列預測

本實驗會進行三個目標的預測實驗，採用3.2提及之歸屬程度解構之方法，使類神經模型輸出兩個值。第一個輸出的值，其實數部用於預測道瓊工業指數；虛數部預測日經平均指數(Nikkei 225 index)，第二個輸出的實數部預測巴西股市指數(Bovespa index)，利用此實驗驗證此研究在進行多目標的輸出預測能有不錯之效果。

日經平均指數是由日本經濟新聞社推出的股價指數，由東京證券交易所225間公司股票價格變動的平均指數，該指數始於1950年，該指數被看作日本最有影響和代表性的股價指數，通過它可以瞭解日本的股市行情變化和經濟景氣變動狀況，是一投資日本的重要參考指標。巴西股市指數是聖保羅交易所上市的公司股票指數，此股市指數常被投資者視為投資拉丁美洲的重要指標。

實驗三的預測目標可以說是結合了各大經濟體的指數指標，若能以此模糊類神經模型及演算法達到良好的預測效果，可以給經濟研究帶來助益。

此三個目標的資料擷取期間如下：巴西股市指數自2006年3月14至2010年4月1日止；日經平均指數自2006年3月3日至2010年4月1日止；道瓊工業指數自2006年4月12日至2010年4月1日止。每個目標的最後200筆用於測試資料，其餘皆為訓練資料。在進行學習參數優化過程中，以RMSE作為各蟻群解的性能指標。

進行特徵選取前的資料前處理，各自選取目標交易日前20個交易日的指數差量之變化作為特徵學習中的候選特徵，因此共有60組候選特徵，記為。至為道瓊工業指數的特徵資料，至是日經平均指數的特徵資料，至是巴西股市指數的特徵資料。各自依序由最遠離目標交易日的資料依序向目標交易日靠近。候選特徵會影響目標變數並且向其提供資訊，而每個特徵變數的選擇增益是基於對該目標變數的貢獻以及對於已選擇的特徵變數彼此的冗餘資訊。

此60個特徵對於巴西股市指數做為目標時的選擇增益量如表 15，選擇增益增減量圖如圖 31；道瓊工業指數做為目標時的選擇增益量如表 16，該選擇增益增減量圖如圖 32；最後以日經平均指數做為目標時的選擇增益量如表 17，選擇增益增減量圖如圖 33。根據圖 31、圖 32、圖 33可見，要讓選擇特徵讓每個目標的選擇增益達到最大值，所要選取的特徵數量會非常龐大。大量的資料會對模型造成預測及學習上的困難，於是本實驗會利用多目標特徵選取策略針對所有特徵再次進行選取。

結束特徵選取後，為最後被選入模型進行參數學習預測的候選特徵變數，此三特徵個別為第13、14、20個特徵。類神經網路模型的輸入資料為**，**是輸入資料的向量，為一複數目標向量，包含了兩個值，第一個值得實數部為道瓊工業指數預測目標，虛數部為日經平均指數預測目標；第二個值實數部為巴西股市指數等以上三個預測目標。則為資料筆數。定義如下：

|  |  |
| --- | --- |
|  | (69) |

此實驗在每個輸入在複數模糊集層皆建立三個複數高斯複數模糊集，所以總共在前鑑部空間劃分出27個區域。並且建立了27條後鑑部T-S規則。複數模糊類神經模型的設定如表 18。

研究中複合演算法用於訓練資料集，其設定如表 19，在多群連續型蟻群演算法中的各群蟻群解，其位置初始方式以所有訓練資料的平均值、標準差以及常數1乘上正規隨機亂數的絕對值作為初始蟻群解的數值。

為了測試此實驗的穩定性，一共進行十次試驗，呈現的結果是學習過程中表現最佳的第4次試驗。該次試驗的參數值在學習完成後的值如表 20、表 21。而該次試驗在學習過程中，最佳蟻群解的性能指標曲線，亦即學習曲線。表示如圖 40。模型預測道瓊工業指數、日經平均指數、巴西股市指數等三個股市指數的預測值以及實際指數值如圖 36、圖 38、圖 34圖 21，誤差量如圖 37、圖 39、圖 35。實驗結果與比較文獻的差異表現於表 22、表 23、表 24。相較於比較文獻[ 36]與本研究的預測結果示於表 22、表 23、表 24。

表 15實驗三之特徵選取次序、特徵變數、選取增益及總選取增益-  
以巴西股市指數為目標(節錄前十選擇次序)

|  |  |  |  |
| --- | --- | --- | --- |
| 選擇次序 | 特徵 | 選取增益 | 總選取增益 |
| 1 |  | 0.0068 | 0.0068 |
| 2 |  | 0.0284 | 0.0353 |
| 3 |  | 0.0158 | 0.0511 |
| 4 |  | 0.0146 | 0.0657 |
| 5 |  | 0.0142 | 0.0799 |
| 6 |  | 0.0125 | 0.0924 |
| 7 |  | 0.0125 | 0.1049 |
| 8 |  | 0.0112 | 0.1161 |
| 9 |  | 0.0102 | 0.1263 |
| 10 |  | 0.0104 | 0.1367 |



圖 32輸入特徵數對於目標總選取增益量變化-以巴西股市指數為目標

表 16實驗三之特徵選取次序、特徵變數、選取增益及總選取增益-  
以道瓊工業指數為目標(節錄前十選擇次序)

|  |  |  |  |
| --- | --- | --- | --- |
| 選擇次序 | 特徵 | 選取增益 | 總選取增益 |
| 1 |  | 0.0063 | 0.0063 |
| 2 |  | 0.0058 | 0.0121 |
| 3 |  | 0.0059 | 0.0180 |
| 4 |  | 0.0092 | 0.0272 |
| 5 |  | 0.0052 | 0.0324 |
| 6 |  | 0.0042 | 0.0366 |
| 7 |  | 0.0036 | 0.0401 |
| 8 |  | 0.0032 | 0.0433 |
| 9 |  | 0.0031 | 0.0465 |
| 10 |  | 0.0021 | 0.0486 |



圖 33輸入特徵數對於目標總選取增益量變化-以道瓊工業指數為目標

表 17實驗三之特徵選取次序、特徵變數、選取增益及總選取增益-  
以日經平均指數為目標(節錄前十選擇次序)

|  |  |  |  |
| --- | --- | --- | --- |
| 選擇次序 | 特徵 | 選取增益 | 總選取增益 |
| 1 |  | 0.0063 | 0.0063 |
| 2 |  | 0.0058 | 0.0121 |
| 3 |  | 0.0059 | 0.0180 |
| 4 |  | 0.0092 | 0.0272 |
| 5 |  | 0.0052 | 0.0324 |
| 6 |  | 0.0042 | 0.0366 |
| 7 |  | 0.0036 | 0.0401 |
| 8 |  | 0.0032 | 0.0433 |
| 9 |  | 0.0031 | 0.0465 |
| 10 |  | 0.0021 | 0.0486 |



圖 34輸入特徵數對於目標總選取增益量變化-以日經平均指數為目標

表 18實驗三複數類神經模型設定

|  |  |
| --- | --- |
| 類神經模型設定 |  |
| 輸入特徵 |  |
| 學習次數 | 50 |
| 輸出值數量(複數型態) | 2 |
| 前鑑部個數 | 27 |
| 前鑑部參數總個數 | 27 |
| 後鑑部規則個數 | 27 |
| 後鑑部規則參數總個數 | 108 |

表 19實驗三MGCACO-RLSE複合式演算法參數設定

|  |  |
| --- | --- |
| MGCACO |  |
| 蟻群組數量 | 3 |
| 每蟻群組回傳蟻群解數量 | 5 |
| 每個蟻群組中蟻群解數量 | 30 |
| 每個蟻群組更新蟻群解數量 | 20 |
| 更新蟻群解比例 | 40% (取代8個蟻群解) |
| 更新蟻群解比例 | 60% (取代12個蟻群解) |
| RLSE |  |
| 後鑑部參數數量 | 108 |
|  | 109 |
|  | 108\*108 單位矩陣 |
|  |  |
|  | 108\*1 零矩陣 |



圖 35預測結果-以巴西股價指數為目標:  
藍色實線為實際目標值，紅色虛線為模型預測值



圖 36模型預測誤差值-以巴西股價指數為目標



圖 37預測結果-以道瓊工業指數為目標:  
藍色實線為實際目標值，紅色虛線為模型預測值



圖 38模型預測誤差值-以道瓊工業指數為目標



圖 39預測結果-以日經平均指數為目標:  
藍色實線為實際目標值，紅色虛線為模型預測值



圖 40模型預測誤差值-以日經平均指數為目標



圖 41實驗三模糊類神經模型學習曲線

表 20實驗三學習後前鑑部參數值

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 前鑑部參數 | | | | | | | | | | | |
|  |  | | | | |  | |  |  | | |
| CGFS # | c | σ | | λ | | CGFS # | c | σ | λ |
| 1 | -4.5914 | -469.0391 | | 0.6920 | | 4 | -1.2537 | 361.3825 | 0.1837 |
| 2 | 11.3885 | 137.0868 | | 2.1461 | | 5 | -12.5113 | 71.0397 | -0.0769 |
| 3 | 1.5044 | 313.8373 | | 1.6371 | | 6 | 0.0474 | 593.2423 | -0.2112 |
|  |  | | | | | |
| CGFS # | c | | σ | | λ | |
| 7 | -0.6686 | | -448.5904 | | 0.0017 | |
| 8 | 1.8600 | | -393.4674 | | 0.1077 | |
| 9 | -1.1526 | | 92.1930 | | -0.2425 | |

表 21實驗三學習後後鑑部參數值

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | | | | | | | | |
|  |  | |  |  |  |  |  |  |  |
| 1 |  | | 0.0066-0.0145i |  | 0.1289-0.2822i |  | -0.4353+0.9534i |  | 0.2117-0.4632i |
| 2 |  | | 0.0065-0.014i |  | 0.1271-0.2722i |  | -0.4292+0.9191i |  | 0.2091-0.4471i |
| 3 |  | | 0.0097-0.0343i |  | 0.1882-0.6657i |  | -0.636+2.2486i |  | 0.3092-1.0936i |
| 4 |  | | 0.0029+0.0022i |  | 0.0555+0.0438i |  | -0.1879-0.1423i |  | 0.0913+0.0704i |
| 5 |  | | 0.0028+0.0026i |  | 0.0544+0.0509i |  | -0.1836-0.1685i |  | 0.0892+0.0833i |
| 6 |  | | 0.0053-0.0138i |  | 0.1038-0.2674i |  | -0.3504+0.9042i |  | 0.1705-0.4396i |
| 7 |  | | 0.0062-0.0114i |  | 0.1198-0.2224i |  | -0.4041+0.7511i |  | 0.1965-0.3655i |
| 8 |  | | 0.0061-0.0109i |  | 0.118-0.2123i |  | -0.3983+0.7176i |  | 0.1939-0.3486i |
| 9 |  | | 0.0093-0.0316i |  | 0.1803-0.6145i |  | -0.609+2.0737i |  | 0.2958-1.0093i |
| 10 |  | | 0.0031+0.0064i |  | 0.061+0.1238i |  | -0.2062-0.417i |  | 0.1002+0.2032i |
| 11 |  | | 0.0031+0.0069i |  | 0.0594+0.1333i |  | -0.2006-0.4488i |  | 0.0976+0.2189i |
| 12 |  | | 0.0064-0.0147i |  | 0.1245-0.2852i |  | -0.4204+0.9622i |  | 0.2042-0.4674i |
| 13 |  | | 0.0002+0.0182i |  | 0.0035+0.3533i |  | -0.0118-1.1954i |  | 0.0057+0.5803i |
| 14 |  | | 0.0001+0.0186i |  | 0.0022+0.3621i |  | -0.0075-1.2237i |  | 0.0037+0.5936i |
| 15 |  | | 0.0028+0.0013i |  | 0.0548+0.0252i |  | -0.1851-0.0851i |  | 0.09+0.0409i |
| 16 |  | | 0.0026+0.0096i |  | 0.0513+0.1875i |  | -0.1731-0.633i |  | 0.0841+0.3064i |
| 17 |  | | 0.0025+0.0101i |  | 0.0496+0.1962i |  | -0.1674-0.6648i |  | 0.0814+0.324i |
| 18 |  | | 0.006-0.0118i |  | 0.1159-0.229i |  | -0.3916+0.7733i |  | 0.1902-0.3761i |
| 19 |  | | 0.0057-0.009i |  | 0.111-0.175i |  | -0.3746+0.5916i |  | 0.1822-0.2878i |
| 20 |  | | 0.0056-0.0085i |  | 0.1093-0.1651i |  | -0.3687+0.5571i |  | 0.1795-0.271i |
| 21 |  | | 0.0088-0.0291i |  | 0.1714-0.5657i |  | -0.5785+1.9086i |  | 0.2812-0.9291i |
| 22 |  | | 0.0022+0.0064i |  | 0.0419+0.1254i |  | -0.1413-0.4215i |  | 0.0687+0.2068i |
| 23 |  | | 0.0021+0.0068i |  | 0.0405+0.1325i |  | -0.1371-0.447i |  | 0.0666+0.218i |
| 24 |  | | 0.0047-0.0098i |  | 0.0908-0.1904i |  | -0.3068+0.6431i |  | 0.1491-0.3128i |
| 25 |  | | 0.0052-0.0059i |  | 0.1016-0.1142i |  | -0.3431+0.3855i |  | 0.1669-0.188i |
| 26 |  | | 0.0051-0.0054i |  | 0.0999-0.1047i |  | -0.3376+0.3522i |  | 0.1639-0.1713i |
| 27 |  | | 0.0084-0.0264i |  | 0.1632-0.5117i |  | -0.5513+1.7334i |  | 0.2679-0.8403i |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 表 22實驗三效能比較- 以巴西股市指數為目標 [ 36]   |  |  | | --- | --- | | 方法 | RMSE | | Wavelet-SVR | 878.7561 | | Wavelet-MARS | 881.0468 | | Waveelt-MARS-SVR | 876.9092 | | Single ARIMA | 883.9282 | | Single SVR | 880.0151 | | Single ANFIS | 882.0765 | | MGCACO-RLSE | 882.2974 | | 表 23實驗三效能比較- 以道瓊工業指數為目標 [ 36]   |  |  | | --- | --- | | 方法 | RMSE | | Wavelet-SVR | 32.2070 | | Wavelet-MARS | 33.0173 | | Waveelt-MARS-SVR | 32.2058 | | Single ARIMA | 33.6321 | | Single SVR | 32.9112 | | Single ANFIS | 33.2513 | | MGCACO-RLSE | 97.9072 | |
| 表 24實驗三效能比較- 以日經平均指數為目標 [ 36]   |  |  | | --- | --- | | 方法 | RMSE | | Wavelet-SVR | 133.2766 | | Wavelet-MARS | 138.8247 | | Waveelt-MARS-SVR | 132.3721 | | Single ARIMA | 140.3164 | | Single SVR | 137.6541 | | Single ANFIS | 135.5629 | | MGCACO-RLSE | 139.2009 | |  |

# 討論

本研究提出以複合型演算法MGCACO-RLSE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。根據前3個實驗，顯示本研究的一系列方法對於時間序列的預測擁有較佳的預測性能。

本研究在資料的特徵挑選方針，取自夏農資訊熵的理論。利用資料彼此之間的資訊熵以及資訊熵差量，計算劃分出的資料候選特徵對於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，以選取增益代表最後該特徵被選取後能夠提供給目標的資訊量。為了避免選入太多的特徵輸入至模型，針對單目標以及多目標的資料，還會進行二次的特徵篩選，選出所有有正向效益的特徵中，選取增益較為大量的部分特徵，以減少模型對於太多目標所產生的預測負擔。

複數模糊類神經模型以複數模糊集、T-S模糊系統等方式建立。T-S模糊系統能夠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。複數模糊集相較於傳統的模糊集合，其歸屬程度從一維實數空間延伸到一個二維單位圓盤的平面空間，此方式使歸屬程度能夠容納更多的資訊，有助於提升模糊系統的推理能力與應用效能。而利用複數歸屬程度可以使模型系統產生雙輸出，提供系統同時預測兩組時間序列資料。除此之外，其複數歸屬程度亦能解構，分別擷取其實數部以及虛數部，讓整體模糊集產生多個歸屬程度，使原本的雙輸出預測能夠預測更多目標，增強模型解釋及預測能力。

進行複數模糊類神經模型的參數學習，利用多群連續型蟻群演算法結合遞迴最小平方演算法分別對模型的前鑑部以及後鑑部參數優化。多群連續型蟻群演算法採用多群搜尋的方式，避免演算法結果過早收斂，並且能同時兼顧蟻群解的多樣性。而本研究在蟻群演算法的演化上，再加上資訊流通、淘汰、繼承等流程。使本研究之多群演算法可以共享每個群體的搜尋結果，並且將表現最差的群體淘汰。但是保留該群體的較佳解再加上其他群體的較佳解，產生新的群體繼續搜索。結合此三流程，加速搜尋的能力以及演化收斂的速度。

後鑑部是採用遞迴式最小平方演算法，利用輸入資料點，以及前次的計算結果，尋找一線性函數，使資料點與該函數的平方誤差達到最小值，藉此找到符合資料關係的函數。並且利用不斷遞迴計算的方式，最佳化其後鑑部參數。

本研究在實驗部分以三個不同資料的實驗呈現建構之模糊類神經網路模型，並且以MGCACO-RLSE優化其參數以及將輸入資料進行特徵選取的實驗結果。針對以上理論及方法將進行以下探討：

## 利用複數類神經網路模型針對單目標資料進行預測

本節將討論本研究所提出全新的多群複合演算法用於複數模型對於單一目標的計算。目的是利用較為容易的單一目標預測，驗證本研究理論在基礎上的可行性。實驗一利用複數類神經網路模型預測效能，並且與多個模型方式進行比較。根據表 6，可以發現本研究測試結果的RMSE為517。相較於其他方式，如最佳之PHM模型，提升約八成的預測效能。比較MAPE，即誤差百分比，也從原本的三成誤差下降到只有3個百分比。該實驗結果如圖 21、圖 22所示，皆有極近似的預測效能。顯示出利用複數模糊集，其歸屬程度包含實數與虛數部分，大幅增加歸屬程度的描述能力，提升模型的預測性能。

## 利用複數型態輸出針對雙目標進行預測

本節將探討採用複數模糊集使模型達到複數型態的輸出，並且針對雙目標進行預測應用，增強模型的應用能力。於實驗二，模型藉由輸入資料，並且利用複數模糊集，在模型輸出，個別擷取其實數部與虛數部，讓兩個不同的數值個別預測不同的目標。實驗的雙目標預測結果如圖 26與圖 28所示，在測試資料集的部份，模型的預測輸出與實際資料極為接近。證實了此複數型態輸出以及模型的預測能力。此模型的輸出校能比較，本實驗的MAPE分別為1.2612%(高盛)與1.0297%(微軟)，皆優於過去的文獻。而且本實驗的預測方式是一次進行雙目標的預測，與以往一次僅進行單一目標預測，困難度更高。本實驗更再次證明加入複數模糊集合的優勢，利用其可以讓模型進行複數型態的輸出，預測兩個不同的時間序列。而且表現皆優於只進行單一目標預測的模型，讓複數模糊集的重要性不言可喻。

## 解構歸屬程度值進行多目標預測

如3.2所敘，利用複數模糊集所得到的複數型態歸屬程度，可以個別擷取其實數部及虛數部，使歸屬函數可以獲得多個不同值，以便進行多個目標的預測應用。於實驗三，利用了歸屬程度的原始值以及其實數部數值，加上複數的模型輸出型態，進行三個不同目標的輸出。利用歸屬程度原始值調整及預測一複數型態的模型輸出，預測道瓊工業指數以及日經平均指數的資料。再將歸屬程度解構，取其歸屬程度實數值用以預測調整及另一模型輸出的實數值，預測巴西股市指數的資料。相關的實驗預測結果如圖 34、圖 36、圖 38所示，在測試資料的部分皆有極為優秀的預測校能。而在校能比較方面，相較於其他的模型預測，雖然並沒有達到最佳效果，但是仍在同一校能階層，證明了利用解構歸屬程度達到多個模型輸出及多目標預測的可行性。未來在數據資料極為龐大之下，能夠以單一模型參數達到多個目標的預測會是一個極大的優勢。

## 特徵選取之應用

於實驗中，雖然利用特徵選取的優勢並無法明顯的從實驗預測結果中得知。但是根據輸入資料產生的候選特徵數量而言，如實驗一，產生了30個候選特徵。而實驗二與實驗三更是多達60個候選特徵。若是將這些候選特徵全數作為輸入資料進入模型，會給模型帶來相當大的資料負擔。但是經由特徵選取，將實驗一與實驗二的輸入資料的特徵降到兩個，實驗三只剩三個。而在預測校能仍然能與對比文獻更加或是在同一效能階層，間接證明了採用特徵選取的優勢所在。

## MGCACO-RLSE複合式演算法效能分析

本研究所採用的模型，由於其前鑑部以及後鑑部的參數量過於龐大，並無法以人工或是直觀的方式直接計算參數值，因此利用此複合式演算法優化參數。但是當求解的參數維度增加，其落入區域最佳解的可能性也增高。於是本研究在優化前鑑部的參數時，採用了多群蟻群演算法。利用其多群搜尋的能力，減少落入區域最佳解的機率。並且加入淘汰的流程，加速尋找最佳解。對於後鑑部參數的部分，採用了遞迴最小平方演算法，利用其可以迅速處理並調整大量後鑑部線性規則參數的優勢，使後鑑部參數不須利用蟻群演算法即可達到優化。

透過三個實驗的學習曲線：圖 23、圖 30、圖 40所示，在少量的時間迭代次數即可達到較佳的效能，並且於學習過程中不斷往全域最佳解搜尋。由此可證明MGCACO-RLSE複合式演算法具有快速的演化速度，並且能夠在落入區域最佳解時可以很快脫離。

# 結論與未來研究方向

## 結論

本研究提出一新型態複合演算法，結合多群連續型蟻群演算法以及遞迴最小平方法，用於優化複數型態模糊類神經模型系統的參數集合。而模型結合複數型模糊集合、T-S模型系統以及類神經網路。於資料輸入前輔以特徵選取，減少過多資料對模型的負擔。經由實驗證明以上方式並統整本研究貢獻如下：

1. 特徵選取處理輸入資料降低模型負擔  
    本研究根據夏農資訊熵的原理與方法，開發出一種新的特徵選取方式。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。
2. 採用複數模糊集使模型能夠有多個複數型態輸出  
    本研究採用的複數模糊集，其複數型態的歸屬程度，讓模型可以有複數值的輸出。而利用解構，將歸屬程度的實數部與虛數部取出各自成為另一歸屬程度，此舉使模型可產生多個輸出。讓模型能夠有同時進行多目標預測的能力。於實驗二與實驗三亦證明此方式的貢獻。
3. MGCACO-RLSE複合式演算法使系統快速學習及避免落入區域解  
    模型系統中的參數好壞可以說是決定預測結果的優劣。本研究採用了多群連續型蟻群演算法結合遞迴最小平方演算法形成一複合式演算法，用於優化模糊模型的前鑑部以及後鑑部參數。蟻群演算法用於優化前鑑部參數，遞迴最小平方演算法用於最佳化後鑑部的參數。此方式可以分配不同類別參數給不同演算法，增加優化的效率。  
    在多群連續型蟻群演算法，加入資訊流通的特性，使不同群體間的搜尋結果可以互相流通，加速各群體搜尋最佳解的速度；加入淘汰的特性，讓演化中效能表現最差之群體被淘汰。由於其效能差，淘汰該群可以減少計算資源不必要的浪費；加入繼承的特性，使前述被淘汰的群體，其較為優秀的蟻群解可以留下。使其即使遭排除，該群體中較為良好的蟻群解可以保留下來，並融入其他群體的較佳解，建立新的一群體。該群體保有各個群體的較佳解，使其表現更為良好，增加預測能力。結合以上三個特性，使預測效能收斂速度加快，並且減少落入區域最佳解的機會。  
    而在遞迴最小平方法，估計的部分將常數項更改為單位矩陣，符合多目標輸出預測的應用。讓原本只能進行單一目標預測應用的演算法，增強到進行多目標的預測及參數調整，增強模型預測及應用能力。

## 未來研究方向

本研究提出之複數模糊類神經模型，並用複合式演算法優化其參數。效能不管是在單一目標或是多目標的預測，根據實驗結果所示，皆能有良好的預測性能。但是仍然有部分細節能夠在未來有延伸。以下提出未來可延續的研究方向：

1. 模型輸入維度之模糊集建立  
    目前研究所建立各輸入維度模糊集的方式是直觀在每個維度建立三個模糊集，並以輸入資料的平均值等初始化模糊集參數，期望能有良好的覆蓋性。但是實際情況中，資料並不會在每個維度上平均分配。若能根據資料在維度上分布的情況，利用群聚演算法，如減法集群演算法(Sub clustering algorithm)或是模糊 c-means演算法(Fuzzy C-means clustering)便能明確的了解資料的分布情況，進而利用其分布情形建立模糊集，讓模糊集能夠更有效率的覆蓋資料點。
2. 前鑑部區域選擇  
    目前在前鑑部區域的劃分是採用各輸入資料維度上的模糊集數目以網格分區法劃分出，但是並非所有劃分出的前鑑部區域都有足夠的資料點分布。若是為了部分僅涵蓋極為少數的資料點而使用該前鑑部，亦會造成不必要的資源消耗。因此未來的研究方向可以朝向利用各個前鑑部當中資料點的密集程度，選取部分的前鑑部於模型中使用。
3. T-S模糊系統  
    本研究中，採用的規則是源自於T-S模糊系統。該方式是以一前鑑部建立一後鑑部。但是若前鑑部數目過多，隨之建立的後鑑部規則數也會相當龐大。加上需要優化的參數數量，對模型是龐大的預測負擔。因此未來可以朝向將前鑑部與後鑑部的建立方式分離，使後鑑部數量不會與前鑑部有太大關係，如此便能使模型建立更為彈性。

# 參考文獻

[ ] R.F. Engle, “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, vol. 50, iss. 4, p.p. 987-1007, 1982

[ ] T. Bollerslev, ”Generalized Autoregressive Conditional Heteroscedasticity,” *Journal of Econometrics*, vol.31, pp. 307-27, 1986

[ ] T. Kimoto and K. Asakawa and M. Yoda and M. Takeoka,” Stock market prediction system with modular neural networks,” *International Joint Conference* , USA, 1990

[ ] K. Kim and I. Han, “Genetic algorithms approach to feature discretization in artificial neural networks for prediction of stock index,” *Expert System with Applications*, vol. 19, pp. 125–132, 2000

[ ] T.H. Roh, “Forecasting the volatility of stock price index,” *Expert Systems with Applications*, vol.33, pp. 916–922, 2007

[ ] J. Kennedy and R.C. Eberhart, “Particle swarm optimization,” Proceedings *IEEE International Conferencnce on Neural Networks (Perth, Australia)*, vol. 4, pp. 1942-1948, 1995

[ ] R. C. Eberhart and J. Kennedy, “A new optimizer using particle swarm theory,” *Proceedings IEEE International Symposium on Micro Machine and Human Science (Nagoya, Japan)*, pp. 39-43, October 1995

[ ] A. Colorni and M. Dorigo and V. Maniezzo, “Distributed optimization by ant colonies,” *Proceedings of the 1st European Conference on Artificial Life*, pp. 134-142, Paris, 1991

[ ] C.Juang and T.Jeng and Y.Chang, “An Interpretable Fuzzy System Learned Through Online Rule Generation and Multiobjective ACO With a Mobile Robot Control Application,” *IEEE Transactions on Cybernetics*, vol. 46, iss. 12, pp.2706-2718, December 2015

[ ] H.J. Sadaei and R. Enayatifar and M.H. Lee and M. Mahmud “A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock price forecasting,” *Applied Soft Computing*, vol.40, pp. 132-149, March 2016

[ ] C.E. Shannon and W. Weaver, “*The Mathematical Theory of Communication*,” Univ of Illinois Press, 1949.

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

[ ] D.Pedro ,” A few useful things to know about machine learning,” *Communications of the ACM*, vol.55, iss. 10, pp.78-87, October 2012

[ ] M. Dash and H. Liu, “Feature selection for classification,” *Intelligent Data Analysis*, vol.1, iss. 1-4, pp. 131-156, 1997

[ ] L. Yu and H. Liu, “Efficient Feature Selection via Analysis of Relevance and Redundancy,” *The Journal of Machine Learning Research*, vol. 5, pp. 1205-1224, Oct 1 2004

[ ] K. Kira and L.Rendell, “ A practical approach to feature selection,” In: *Proc. of the Ninth International Conference on Machine Learning*, pp. 249–256, 1992

[ ] M.A. Hall, “Correlation-based Feature Selection for Machine Learning,” University of Waikato, Hamilton, New Zealand, PhD thesis, 1999

[ ] H. Yu and J. Oh and W.-S. Han, “Efficient feature weighting methods for ranking,” in:*Proc. of the 18th ACM Conference on Information and Knowledge Management* *Hong Kong, China*, pp. 1157-1166, 2009

[ ] Z. Sun and T. Qin and Q. Tao and J. Wang, ”Robust sparse rank learning for non-smooth ranking measures,” in:*Proc. of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2009) 20 Boston, MA*, pp. 259-266, 2009

[ ] R.Clausius, " Ueber eine veränderte Form des zweiten Hauptsatzes der mechanischen Wärmetheorie," *Annalen der Physik und Chemie*, vol.93, issue.12, pp. 481–506, 1854

[ ] G. Cantor, “Ueber eine Eigenschaft des Inbegriffs aller reellen algebraischen Zahlen,” *Journal für die reine und angewandte Mathematik*, no. 77, pp. 258-262, 1874

[ ] B. Russell, “Vagueness,” *Australasian Journal of Philosophy*, vol.1, iss. 2, pp.84 – 92, 1923

[ ] L.A. Zadeh, ”Fuzzy sets,” *Information and Control*, vol.8, issue.3, pp.338-353, June 1965

[ ] D. Ramot and R. Milo and M. Friedman and A. Kandel, “Complex fuzzy sets,” *IEEE Transactions on Fuzzy Systems*, vol. 10, iss. 2, pp. 171-186, April 2002

[ ] C. Li and T. Chiang, “Complex Neurofuzzy ARIMA Forecasting—A New Approach Using Complex Fuzzy Sets,” *IEEE Transactions on Fuzzy Systems*, vol. 21, iss. 3, pp. 567-584, June 2013

[ ] F. Rosenblatt, “The Perceptron: A Probabilistic Model For Information Storage And Organization in the Brain,” *Psychological Review*, vol.65, no.6, pp. 386–408, December 1958

[ ] D.E. Rumelhart and J. L. McClelland,” *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*”, January 1986

[ ] J.H. Holland, “Adaptation in Natural and Artificial Systems,” *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology*, Control and Artificial Intelligence, MIT Press Cambridge, MA, USA, ISBN:0262082136

[ ] A.Colorni and M. Dorigo and V. aniezzo,“Distributed optimization by antcolonies,”*Proceedings of the 1st European Conference on Artificial Life*,pp.134-142, Paris, 1991.

[ ] K. Socha and M. Dorigo,” Ant colony optimization for continuous domains,” *European Journal of Operational Research*, vol.185, iss. 3, pp. 1155–1173, 16 March 2008

[ ] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *IEEE Trans. Syst. Man Cybern*, vol.15, pp. 116–132, 1985

[ ] P. Husbands and F. Mill, “Simulated Co-Evolution as the Mechanism for Emergent Planning and Scheduling,” Proceedings of the 4th International Conference on Genetic Algorithms, San Diego, CA, USA, pp. 264-270, January 1991

[ ] F. van den Bergh and A.P. Engelbrecht, “A Cooperative approach to particle swarm optimization,” IEEE Transactions on Evolutionary Computation, vol.8 , iss.3, pp.225-239, June 2004

[ ] J.J. Wang and J.Z. Wang and Z.G. Zhang and S.P. Guo, “Stock index forecasting based on a hybrid model,” Omega, vol. 40, iss. 6, pp. 758-766, December 2012

[ ] JL Ticknor, “A Bayesian regularized artificial neural network for stock market forecasting, ”Expert Systems with Applications,” Expert Systems with Applications, vol. 44, iss. 14, pp. 5501-5506, 15 October 2013

[ ] L.J. Kao and C.C. Chiu and C.J. Lu and C.H. Chang, “A hybrid approach by integrating wavelet-based feature extraction with MARS and SVR for stock index forecasting ,” *Decision Support Systems*, vol. 54, iss. 3, pp. 1228-1244, February 2013

[ ] 國立中央大學資訊管理所李俊賢教授, 研究生訓練課程內容2015-2017, 紀錄筆記。 (未發表)