行政院國家科學委員會專題研究計畫 成果報告

使用人工智慧的方法來探討金融衍生性市場之認購權證的動態變化
研究成果報告(精簡版)

計 畫 類 別 ： 個別型
計 畫 編 號 ： NSC 99-2221-E-216-052-
執 行 期 間 ： 99 年 08 月 01 日至 100 年 07 月 31 日
執 行 單 位 ： 中華大學資訊管理學系

計 畫 主 持 人 ： 邱登裕

計畫參與人員： 碩士班研究生-兼任助理人員：荊至德
碩士班研究生-兼任助理人員：徐榮生
博士班研究生-兼任助理人員：Agus Andri
博士班研究生-兼任助理人員：鍾典村
博士班研究生-兼任助理人員：翁永健

報 告 附 件 ： 出席國際會議研究心得報告及發表論文

處 理 方 式 ： 本計畫可公開查詢

中 華 民 國 100 年 10 月 27 日

國科會計畫 精簡報告書

計畫編號: NSC 99-2221-E-216-052

計畫名稱:使用人工智慧的方法來探討金融衍生性市場之認購權證的動態變化

目錄

一 報告內容

1. 前言

由於認購權證的波動幅度比現股來的劇烈,不只下跌期間可能造成的損失,標的股盤整期間隨著時間價值的流逝或股價劇烈的振盪情形,都不是單靠支持向量機預測模型或是傳統的選擇權定價模式,可以有效處理的問題。因此,要投入認購權證這個新興的衍生性金融市場,必定要對其標的股之漲跌時機有精準的掌握程度,如此才能獲得比投資現股更高的利潤。

2. 研究目的

本計畫的目的是結合不同的方法來探討認購權證的內部機制,試著找出一個較佳的探討認購權證動態的方法。我們藉由 Black-Scholes 選擇權定價模式來判斷價格合理性,降低預測誤差。然後利用灰色模型來降低輸入資料之雜訊對於預測模型的影響程度。利用三點成一直線的最小平方法來改善灰色模型常發生剩餘誤差的超過現象。最後我們利用基因演算法中之參數最佳化的特性來挑選出不同支持向量機(SVM)的輸入變數組合和 SVM 的參數,然後從訓練資料中取出這些變數組合所對應的值來形成 SVM 的輸入向量,利用 SVM 可在非線性且高維度中有不錯分類效果的特性來進行訓練,經由演化的方式試著找出最佳的結果。

3. 文獻探討

在此介紹與本計畫有關的文獻。

(1) 「灰色系統」理論(Grey System Theory)是在 1982 年時，由華中理工大學的鄧聚龍教授所提出，當初主要是應用在「控制領域」。不過發展到今日，已經擴展到社會及自然科學系統的各層面。主要就是因為灰色系統模式為一連續型的微分方程式，利用此灰色系統模式，可針對研究對象的演變作一個綜觀的分析與預測。且此灰色系統主要的特點就是可以應用在訊息不完全、行為模式不明確、運作機制不清楚之情況下，做系統的關聯分析、模型建立、預測、決策與控制，且灰色系統可以針對研究對象的不確定性、多變量性輸入與離散的數據等，都能做有效處理。

(2)支持向量機(Support Vector Machines, SVM)是一種分類的技術，其植基於統計學習理論中的結構化風險最小誤差法(Structural Risk Minimization principle, SRM)的機器學習系統，在 1995 年由 Vapnik 和其研究團隊於 AT&T 實驗室所發展出來的。SVM 最先被提出來的簡單模式稱之為最大邊際分類器(maximal margin classifier)，其作用僅用於特徵空間(feature space)中線性可分割(linearly separable)的資料，然現實情況中，很多狀況並不適用，爾後植基於此發展出複雜的支持向量機。支援向量機主要是利用分隔超平面(Separating Hyperplane)的方法，尋找最大的邊界(Margin)，進而將資料分隔成兩類或多類的類別(Class)。

(3) Black-Scholes 定價模式 Fischer Black &Myron Scholes （1973）於 1973 年提出的 B-S 模型主要概念是建構一個無風險的避險投資組合，利用賣出一單位買權，同時買進 delta (= dC/dS)單位的股票，由於此組合乃是在很短的時間內建立的，不存在套利機會，因此這無風險資產組合的報酬率應等於無風險利率，但是股價的變動呈現連續性，所以 dC(買權價格微量變動)與 dS(標的股股價微量變動)的關係時常改變，為使得該組合不具風險，必須不斷調整(rebalance)資產

組合，才能成為無風險資產組合，因此，Black & Scholes 利用上述的假設與連續調整資產組合推演出買權與賣權的定價模型(Black & Scholes，1973)。

(4)基因演算法為一種模仿自然界物競天擇、適者生存的原理，透過模擬生物基因的交配(Crossover)、 複製(Selection)、自然淘汰(Eliminate)與突變(Mutation)等演化特性，來解決一些複雜且難解的問題。其基本的精神在於由樣本中選擇具有較好特性的母代，採用隨機的方式相互交換彼此的位元資訊，以產生更加優秀的子代。所以遺傳演算法能夠同時搜尋不同的區域，是一種最佳化的搜尋機制，適合用在非線性以及解答範圍很大或是很複雜的問題上面。

(5)最小平方法是經由計算、分析求得一條 最佳逼近直線，使所有的實驗數據到此直線的鉛垂方向的距離的平方總和為最小。 一次累加生成的累積三點最小平方預測模型(Cumulate 3-point Least Square Prediction Model, C3LSP)，是將原始數列累積生成一組非負的數列，從而在運用最小平方法，取的最佳的預測趨勢線。認購權證的日成交價格線，如果成不規則的週期性變化時，根據灰色理論之檢測原則是無法建立比較正確的灰色預測模型，因此，增加累積三點最小平方法，來減小預測誤差。

4 研究方法

在此介紹本計畫的研究方法。

(1) 灰色模型(Grey model)

此模型產生了 14 個灰色指標變數，以作為基因式支持向量機的一部份輸入的參考數據。灰色理論假設一個訊息是包含信息的性質元(不可度量)與信息的數據元(可度量)兩部分。

我們利用 GM(1,1)來計算 14 個灰色指標變數，其灰色指標模型公式如下：

$$\hat{x}^{(0)}(n+1)=\frac{b-ax^{(1)}(n)}{1+0.5a}$$

由於灰色指標模型所採用的方法，會隨著日數的增加而增加誤差，所以我們採用了「滾動模型」來解決此問題，此方法是來自於時間軸移動視窗的概念，在滾動模型中建立完第 k 日的 GM(1,1)的預測模型後，將可以根據此模型求得明日的預測值，然後再把此一預測值納入我們的原始序列中，並且把序列中最舊的一個數值移除，以保持原始序列中的建模數據不斷更新。

(2) 累積三點最小平方預測模型(Cumulate 3-point Least Square Prediction Model, C3LSP)

累積三點最小平方預測模型也同樣採取累加生成的方式，待生成新數列後，再採取最小平方法計算最佳化的直線方程式的係數。再以所得的累加生成數列，來預測下一個累加的預測值。

在建立第 k 日的 C3LSP 的預測模型後，我們可以根據此模型求得明日的預測值(即第 k+1 日之預測值)，承襲灰色滾動模型的作法，我們把此一預測值納入我們的原始序列中，並且把序列中最舊的一個數值移除，以保持原始序列中的建模數據不斷更新。接下來再以這三個建模數據來建立第 k+2 日的累積最小平方預測預測模型，反覆進行上述的建模預測流程，直到所有實驗日期之數據皆預測完畢。

(3)基因演算法(Genetic algorithm)

我們利用基因演算法的參數最佳化的特性來挑選出較具影響力的變數，期望能夠藉由輸入變數的篩選來提高支持向量機的效能並提高準確性。此演算法透過模擬生物基因的複製(Selection)、交配(Crossover)與突變(Mutation)等演化特性，可有效解決一些複雜且難解的問題並避免陷入局部最佳解的問題。

染色體的組成總共有 48 個基因，前面 28 個基因代表 28 個輸入變數，其後的 10 個基因代表 SVM 的參數 C (cost)，其數值以二進位表示，最後的 10 個基因代表參數 γ (Gamma)。

基因演算法中最重要的部分在於適應函數的設計,我們的設計上是以獲得最大的預測報酬與預測的精準程度為主要考量依據。另外若我們刻意提高精確率,則勢必會導致捕捉率的下降,反之亦然。因此我們加入 F-measure 的評估方式,希望同時兼顧模型的精確率與補捉率兩種指標。適應函數如下所示。 實驗過程中,我們以試誤法則決定 α 、 β 的數值。

適應值 = α * 報酬率 + β * F_Measure

其中 α 、 β 為常數

(4) 支持向量機(Support Vector Machine, SVM)

我們使用台灣大學資訊工程學系林智仁教授所開發的 LIBSVM(Chang, C., Lin, C.,2001)軟體來進行實驗,LIBSVM 軟體有一般常用的五種分類方法可以選擇使用,有 C-SVC、nu-SVC、epsilon-SVR、nu-SVR 和 one-class SVM,我們選用較適合此實驗資料的 C-SVC 分類方法。

5 結果與討論

本計畫的研究對象為群益 I8(04953)、永豐 98(04973)及群益 06(06235) 這三檔認購權證,其標的股分別為 9904/寶成、3380/明泰及 3009 奇美電。

5.1 結果 1- 群益 I8(04953)

本實驗採用 14 個基本變數加上灰色預測值與累積三點最小平方預測值的平均數做為另外 14 項輸入變數,然後進行基因演算法與支持向量機的運算處理,(簡稱本模型為 GM-C3LSP-GA-SVM),經過 100 代的演化後停止。

我們另外建立三個模型,一個是完全沒有灰色預測值與累積三點最小平方方法

預測值的 GA-SVM 來做為對照組(簡稱 Pure GA-SVM 模型)，其模型建構方法與評估效能方式皆與我們原本的實驗設計相同,唯一的差別僅僅在於對照組完全不選擇灰色預測值與累積三點最小平方法預測值做為輸入變數。因此,在 Pure GA-SVM 架構中，總共只有 14 個輸入變數可供基因演算法做基因演化的工作。

我們發現 GM-C3LSP-GA-SVM 整合式的預測系統在準確率、F_Measure 以及適應函數值均優於 Pure GA-SVM、C3LSP-GA-SVM 與 GM-GA-SVM，這是因為認購權證的價格波動劇烈，容易產生雜訊而影響到 GA-SVM 的效能，因此驗證了本研究所提出的假設:加入灰色預測值與累積三點最小平方法的平均值作前期的資料整理，可以有效降低波動性數據對於預測模型所產生的雜訊，提昇其準確性。

5.2 結果 2-永豐 98(04973)

我們清楚發現前四名的 GM-C3LSP-GA-SVM 模型中，全部包含 Black – Scholes 定價值(W_B-S/GC_W_B-S) 為其輸入變數中，由此可見，在預測「永豐 98(04973)」之走勢時，Black – Scholes 定價值是一個有效的輸入變數，它比起採用標的股的技術指標值來說，有著更佳的適應函數值。

而適應值前五名之 GM-C3LSP-GA-SVM 模型均有包含 GM-C3LSP 的平均值為輸入變數中，而且佔整體輸入變數不少的比重。因此我們可以知道，GM-C3LSP 的平均值做為輸入變數，對於 GA-SVM 的效果有重要的助益。

我們發現 GM-C3LSP-GA-SVM 整合式的預測系統在準確率、F_Measure 以及適應函數值均優於 Pure GA-SVM、C3LSP-GA-SVM 與 GM-GA-SVM，這是因為認購權證的價格波動劇烈，容易產生雜訊而影響到 GA-SVM 的效能，因此驗證了本研究所提出的假設:加入灰色預測值與累積三點最小平方法的平均值作前期的資料整理，可以有效降低波動性數據對於預測模型所產生的雜訊，提昇其準確性。

5.3 結果 3-群益 06(06235)

我們發現前五名的 GM-C3LSP-GA-SVM 模型中，全部包含 Black – Scholes 定價值(W_B-S/GC_W_B-S)為其輸入變數中，由此可見，在預測「群益 06(06235)」之走勢時，Black – Scholes 定價值是一個有效的輸入變數，它比起採用標的股的技術指標值來說，有著更佳的適應函數值。

而適應值前五名之 GM-C3LSP-GA-SVM 模型均有包含 GM-C3LSP 的平均值為輸入變數中，而且佔整體輸入變數不少的比重。因此我們可以知道，GM-C3LSP 的平均值做為輸入變數，對於 GA-SVM 的效果有重要的助益。

我們發現 GM-C3LSP-GA-SVM 整合式的預測系統在準確率、F_Measure 以及適應函數值均優於 Pure GA-SVM、GM-GA-SVM 與 C3LSP-GA-SVM，這是因為認購權證的價格波動劇烈，容易產生雜訊而影響到 GA-SVM 的效能，因此驗證了本研究所提出的假設:加入灰色預測值與累積三點最小平方法的平均值作前期的資料整理，可以有效降低波動性數據對於預測模型所產生的雜訊，提昇其準確性。

5.4 討論

從實驗的結果中，我們可以歸納出下列幾個結論：

(1) 選擇灰色預測值與累積三點最小平方法預測值平均數作為輸入變數的基因演算法的支持向量機預測模型(GM-C3LSP-GA- SVM)，不論是準確率、適應函數值以及報酬率上，都比其它對照組的預測模型來的佳。

(2) 從我們做的實驗對照組中可以清楚發現，在相同的時間點及相同的進出策略上，投資認購權證所獲得的報酬率遠勝過投資該檔認購權證之標的股的獲利率。

(3) 實驗結果證實，當輸入變數加入累積三點最小平方預測值，預測的最佳報酬與僅使用灰色預測值的實驗相比，其報酬率較佳。

三 參考文獻

1. Black, F. and Scholes, M, “The pricing of options and corporate liabilities”, *Journal of Political Economy*, pp. 637-659, 1973.
2. Chang, B. R., “Hybrid BPNN-Weighted Grey-CLMS Forecasting,” *Journal of InformationScience and Engineering*, Vol. 21, No. 1, pp. 209-221, 2005.
3. Chang, B. R. and Tsai, H. F., “Quantum-minimized BWGC/NGARCH approach to financial time series forecast,” *Neurocomputing*, 2008.
4. Cristianini, N. and Taylor, J. S., *An Introduction to Support Vector Machine*s, Cambridge University, New York, 2000.
5. Wu, C. H., Tzeng, G. H., Goo, Y. J., and Fang, W. C.. “A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy,” *Journal of Expert Systems with Applications*, Vol.32, pp.397-408, 2007.
6. Yu, L., Wang, S., and Lai, K. K., “Mining stock market tendency using GA-based support vector machines”, *Lecture Notes in Computer Science*, Vol. 3828, pp. 336-345, 2005.
7. 林詠智，「應用遺傳演化類神經網路模型建立衍生性金融商品投資策略之研究」，國立台北大學資訊管理研究所碩士論文，2007。
8. 詹沐仁，「國防預算規模預測之研究-時間數列與灰色預測模型之應用」，國防大學管理學院國防財務資源管理研究所，2007。
9. 謝宗融，「股市報酬率之灰色類神經 GJR-GARCH 模型」，國立雲林科技大學財務金融所碩士論文，2008
10. 鄭家豪，「灰關聯雙移動平均線之應用研究-以股票市場及金屬期貨為例」，國立成功大學資源工程研究所，2007。

四 計畫成果自評

本計畫實際的內容與原計畫的規劃完全相符，而預計達成的目標與研究價值亦已達成。

本計畫的研究成果很適合在學術期刊發表，我們目前正在撰寫論文中，已接近完稿階段，將於近期選擇一個相關的 SCI 國際期刊投稿。

Title: NCM2011 Paper Acceptance Notification

NCM2011 Registration Guideline disable_http://www.aicit.org/NCM/reg_deadline.html
NCM2011 Paper Format Download disable_http://www.aicit.org/NCM/paper_type_format.html

Dear Chiu, Deng-Yiv,

Congratulations! It is my great pleasure to announce you that your paper is accepted by the NCM2011: The 7th International Conference on Networked Computing and Advanced Information Management which will be held from June 21 to 23, 2011 in Gyeongju, Republic of Korea. Please check your information below. With the assigned ID and PW, you can check or revise your personal and paper information from web system of the NCM2011.

Title of the Paper: U.S.A. S&P 500 Stock Market Dynamism Exploration with Moving Window and Artificial Intelligence Approach
Number of the paper: NCMWSWS2-129004
Log-in ID: chiu6531
Log-in PW: cdy51427

Once again, we are welcoming you to NCM 2011 which is the one of most popular and competitive international conferences and only accepting outstanding high quality papers to maintain the quality of NCM 2011. This year, we have accepted a large number of various papers from more than 25 countries and it wasn't easy work for us to select the most innovative and well-written papers among them.

For inclusion of your paper in the conference proceedings, your registration form should be submitted with your final camera-ready paper by the deadline (April 15 - May 2, 2011). Please sign in with your ID and PW and follow the steps to finish your registration process and up-load your final paper.

Prof. Franz I. S. Ko, Ph.D.
General Chair, NCM2011.
Honorary Director General, IBC, Cambridge, UK.
Vice-President, The World Congress of Arts, Sciences and Communications, Cambridge, UK.

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

100 年 10 月 23 日

附件三

報告人姓名	邱登裕	服務機構及職稱	中華大學資訊管理系 教授
會議 時間 地點	100 年 6 月 21 日－100 年 6 月 23 日	本會核定補助文號	NSC 99-2221-E-216-052
會議名稱	(中文)第七屆網路運算與進階資訊管理國際研討會 (英文) NCM2011:7th International Conference on Networked Computing and Advanced Information Management		
發表論文題目	(中文)以移動式視窗和人工智慧方法探討美國股市變化 (英文) U.S.A. S&P 500 Stock Market Dynamism Exploration with Moving Window and Artificial Intelligence Approach		

報告內容應包括下列各項：

一、參加會議經過

第一天的議程: 參與演講和兩個 session 的報告。
第二天的議程:報告自己的論文，並且於報告中和與會者討論其提出之建議。
第三天的議程: 參與兩個 session 的報告。

在會議進行中，我們本次籌辦大會的主席 Franz Ko, Ph.D，並且一同邀約明年的研討會。透過 Franz Ko, Ph.D 的介紹，我們也認識了本次 keynote 演講學者，並且彼此交換研究心得與分享研究主題。

二、與會心得

在參與會議過程中，學者們因為有不同的學術領域與專業，因此開放的在討論中彼此交換意見與心得。

透過國際會議更了解國內外對於資訊管理領域與資訊技術領域多元的應用，對於跨領域合作是一種很好的交流。

三、建議

感謝中華民國的國科會能夠提供此經費，讓我們能夠跟其他國家與跨領域學者相互交流，並且有機會討論合作的事宜，非常感謝!

四、攜回資料名稱及內容

NCM2011 Proceeding 手冊

U.S.A. S&P 500 Stock Market Dynamism Exploration with Moving Window and Artificial Intelligence Approach

Deng-Yiv Chiu

Chung Hua University

707 Sec.2, WuFu Road

Hsinchu, Taiwan, ROC

Cheng-Yi Shiu

Chung Hua University

707 Sec.2, WuFu Road

Hsinchu, Taiwan, ROC

Yu-Sheng Lin*

National Central University

300 Jhongda Road

Jhongli, Taiwan, ROC

***Abstract*- We propose an approach of artificial immune algorithm, fuzzy theorem, support vector regression, and seasonal moving window to explore stock dynamism among same seasons in continuous years for USA S&P 500 stock indexes. First, we select optimal number of trading days to calculate technical indicator values. We apply artificial immune algorithm to locate optimal combination of technical indicators as input variables. The property of nonlinearity and high dimensionality of the support vector regression is employed to explore the stock price patterns.**

I. INTRODUCTION

The price fluctuation in securities market is especially affected by many factors, such as politics, macro economics, and investors' psychology. The dynamic change of securities market is very complex. Although many experts and scholars have researched in this field, the exploration on the dynamic trend of securities market is still a difficult challenge.

In the past decade, various methods have been widely applied to explore the internal dynamism of stock market, such as linear and nonlinear mathematical models, multi-agent mechanism, and artificial neural network (ANN) of multiple layers used to simulate the potential stock market transaction mechanism [1]. Because of the advantages of arbitrary function approximation and needlessness of statistics assumption, ANN is widely applied in the simulation of potential market transaction mechanism [13]. But there exist some problems in artificial neural network, such as local optimum and over-learning. In order to avoid those problems, some researches try to combine hybrid approaches of artificial intelligence methods and artificial neural network [12].

Then, support vector machine (SVM) became a useful and popular method used by many researchers to avoid local optimum and achieve significant performance. SVM has outstanding performances in handling high dimension entry space problems. Such a feature leads to a better performance of SVM in simulating potential market transaction mechanism than other methods. Some researches adopt SVM to predict stock market dynamism with financial factors, such as

* Corresponding author. Tel: +1 3518 6531; fax: +1 3518 6543.
E-mail address: asurada1984@yahoo.com.tw (Y. S. Lin)

of noise and outlier. SVR can convert nonlinear problems into high dimensional space and obtain good performance. Some research mixes models of support vector regression and self-organizing feature map (SOFM) as well as filtering attribute screening method to predict the tertian closing price of Taiwan Stock Index Futures (FITX) [9]. Support vector regression is also used along with fuzzy theorem in some other researches to solve the two problems in prediction of financial time series: noise and non-stationarity. In the research, six data items, such as S&P500, Google, and Microsoft, from U.S. Yahoo are adopted in the prediction experiments [11].

Artificial immune algorithm (AIA) is an intelligent problem-solving technique which has been applied to optimization, computer security, data clustering, pattern recognition or even fault tolerance [4]. AIA seeks to capture some aspects of the natural immune system in a computational framework for solving engineering problems [8].

The theory of fuzzy sets was first introduced by Loti Zadeh [14], primarily in the context of his interest in the analysis of complex systems. It introduces vagueness by eliminating the sharp boundary dividing members of the class from nonmembers. Fuzzy theory provides the forms for representing uncertainties. As to the application of fuzzy theorem in the financial and banking areas, previous study uses TSK (Taskagi-Sugeno-Kang) fuzzy rule system, and applies technical indicators as input variables to identify the meta-heuristic among them [3]. Some studies use fuzzy theorem and technical analysis to transfer technical indicators into fuzzy technical indicators. They set up fuzzy rule for membership function of each fuzzied technical indicator. When output is generated, the decisions are made under different rules. Therefore, when there is change or certain trend in share price, an investment strategy can be planned based on fuzzy logic [7].

In this paper we employ an approach of artificial immune algorithm, fuzzy theorem, support vector regression, and seasonal moving window to explore stock dynamism among same seasons in continuous years for USA S&P 500. We use fuzzy c-means, fuzzy relation composition, and defuzzication methods to select optimal number of trading days to calculate technical indicator values. We apply artificial immune algorithm to locate optimal combination of technical indicators as input variables. The corresponding values of technical indicators are taken from the training data to form the input vectors of SVR, which is trained through the property of nonlinearity and high dimensionality. Moreover, we divide a year into four seasons and use the seasonal moving window to capture the stock market movement.

II. ARTIFICIAL IMMUNE ALGORITHM

Artificial immune algorithm is inspired by theoretical immunology and observed immune functions, principles and models, which are applied to solving engineering problems [8]. The clonal selection algorithm is a part of AIA based on clonal expansion and affinity maturation [6]. In clonal selection process, when detecting an antigen, the antibodies which best recognize the detected antigen is to proliferate by cloning. This immune response is specific to each antigen.

The immune cells will be duplicated in response to a replicating antigen until it is able to recognize and fight against this antigen. Partial newly cloned cells will be differentiated into plasma cells and memory cells. The plasma cells will yield antibody and promote genetic variation when it performs mutation process. The memory cells are responsible for the immunologic response for new antigen invasion. Finally, the selection mechanism will keep the cells with the best affinity to the antigen in the next generation. The process is described as below [5]:

(1) Initialize the first population of antibody at random.

(2) Calculate fitness value of each antibody.

(3) Generate clones by cloning all cells in the population.

(4) Mutate the clone population to produce a mature clone population.

(5) Evaluate affinity values of the clones’ population.

(6) Select the best antibody to compose the new antibody population.

(7) Steps (3) to (6) are repeated until a pre-defined termination condition is reached.

III. THE PROPOSED APPROACH

Here, we introduce how to select optimal number of trading days to calculate technical indicator values, AIA design, and the details of proposed architecture.

3.1. The Selection of Best Number of Trading Days

We employ fuzzy relation composition and defuzzication methods to obtain the membership degree between a technical indicator and the transaction strategy. The results are compared with actual stock price fluctuation to assess coincidence rate. The day number with the best coincidence rate is selected as the optimal number of trading days to calculate technical indicator values.

We apply fuzzy c-means algorithm to obtain the membership degree between a technical indicator value and a cluster. Then compute the daily actual fluctuation percentage of technical indicator values clustered into same cluster with highest membership degree to obtain membership degree between a cluster and a transaction strategy. Finally, we compute the membership between a technical indicator value and a transaction strategy by employing fuzzy relation composition and defuzzication methods.

(1) Applying fuzzy relation composition method to calculate the membership degree between a technical indicator value and a transaction (buying or selling) strategy

Suppose R and S are respectively the fuzzy relations, which are defined in the cartesian product space ($U \times V$) and cartesian product ($V \times W$); where, *U*, *V*, *W* are the fuzzy sets, the composition operator of fuzzy relations (T = R ◦ S) can be used to infer the fuzzy relation between U and W, which is defined as follows:

$$\mu_T(u, w) = \max_{v \in V}\{\mu_R(u, v) \cdot \mu_S(v, w)\} \quad u \in U, v \in V, w \in W$$

In our study, μ_R (technical_indicator, cluster) is the membership degree between a technical indicator value and a cluster, μ_S (cluster, buy_or_sell) is the membership degree between a cluster and a buying or selling transaction strategy, and μ_T (technical_indicator, buy_or_sell) is the membership degree between a technical indicator value and a buying or selling strategy.

(2) Applying defuzzication method to calculate the membership degree between a technical indicator value and a transaction (buying or selling) strategy

Defuzzication is the process of transferring the conclusion obtained through fuzzy inference to clear messages. We employ Center of area of defuzzication to calculate the buying and selling membership degree.

The equation of coincidence rate is as follows:

(1) If (membership degree of buying inferred by composition of fuzzy relations > membership degree of selling inferred by composition of fuzzy relations), and (the actual share price on the day is rising)
then coincidence count = coincidence count + 1
(2) If (membership degree of selling inferred by composition of fuzzy relations > membership degree of buying inferred by composition of fuzzy relations), and (the actual share price on the day is falling)
then coincidence count = coincidence count + 1
(3) If (membership degree of buying inferred by defuzzication > membership degree of selling inferred by defuzzication), and (the actual share price on the day is rising)
then coincidence count = coincidence count + 1
(4) If (membership degree of selling inferred by defuzzication > membership degree of buying inferred by defuzzication), and (the actual share price on the day is falling)
then coincidence count = coincidence count + 1

Finally, the coincidence rate can be obtained as below:

coincidence rate = total of coincidence count/
(2 * total of trading days)

3.2. Artificial Immune Algorithm Design

Here, we introduce the antibody and fitness design. The antibody includes 14 bits. Each bit represents one technical indicator shown as Table I.

The fitness design is as below. The major factors considered include earning rate, precision, and F-measurement. The weights adopted are 0.8 and 0.2 with trial and error method.

Fitness = (earning rate × transaction precision) × 0.8 + (F1 + F2) × 0.2
Where,

$$\text{earning rate} = \prod_{i=1}^{I_m} (P_{sell_i} - P_{buy_i}) / P_{buy_i}$$

I_m is transaction count of year m, P_{sell_i} is selling price of transaction *i*, P_{buy_i} is buying price of transaction *i*.
transaction precision = count of correct transactions / count of all transactions

$$F1 = \frac{2 \times \text{rising recall rate} \times \text{rising precision rate}}{\text{rising recall rate + rising precision rate}}$$

$$F2 = \frac{2 \times \text{falling recall rate} \times \text{falling precision rate}}{\text{falling recall rate + falling precision rate}}$$

3.3. The Architecture of the Proposed Approach

The architecture of proposed approach is explained as follow.

(1) Data collection: We collect stock data from U.S.A. Yahoo! financial website for 10 years. The extracted period is from 1996/1/1 to 2005/12/31.

(2) Computation and normalization of technical indicator values for specific number of trading days: The range of number of trading days used to calculate technical indicator values is set between 3 and 60 days. Then we normalize the values as below:

normalized value = (original value – average value) / standard deviation.

(3) Calculating coincidence rate of a technical indicator: We employ FCM to cluster normalized technical indicator values for a specific cluster number, calculate the membership degree between a cluster and a transaction, and calculate the membership degree between a technical indicator value and a transaction strategy. Finally, we compute the coincidence rate for the specific number of trading days.

(4) Is maximum cluster number reached? : Here, we examine if the number of clusters employed in the loop reached the maximum limit. If so, go to step (5). Otherwise, the number of cluster is incremented by one. Go to step (3).

TABLE I

14 bits													
DIF	MACD	RS	RSI	RSV	K	D	J	PSY	BIAS	MTM	WMS	AR	BR

(5) Is maximum number of trading days used to calculate technical indicator values reached? : Here, we examine if the number of trading days used to calculate technical indicator values employed in this loop has reached the maximum limit. If so, go to step (6). Otherwise, number of trading days is incremented by one. Go to step (2).

(6) Selection of optimal number of trading days used to calculate technical indicator values: We select the number of trading days which leads to best coincidence rate as the optimal one. Then we form the training and testing data sets by using the selected number of trading days to calculate the corresponding technical indicator values.

(7) Initialization of antibody for AIA process: The first generation of AIA process is initialized at random. A generation includes 20 antibodies and each antibody consists 14 bits. Each bit represents one technical indicator.

(8) Training data selection: Corresponding values of selected technical indicators are extracted as training data to form the input of SVR.

(9) Evaluation of fitness value: The extracted training data is used to train SVR and produce values needed to evaluate fitness value. The kernel function employed is RBF.

(10) Termination criterion of AIA algorithm: the termination criterion is evolution of 50 generations. If the criterion is met, terminate the AIA process and then go to step (12).

(11) Process of AIA: In AIA, an evolution includes clonal selection, reproduction, recombination, and mutation. In the clonal selection process, one quarter of antibodies with the highest fitness values are selected and duplicated using the roulette wheel selection method. In mutation, the mutation rate is defined as 1%, and the process is redirected to step (8).

(12) Evaluation of testing data with trained SVR classifier: The trained SVR classifier is used to classify testing data to determine the proper transaction time point.

(13) Performance comparison: The performance of the proposed approach is compared with that of other methods to see how much the proposed method can outperform.

IV. DATA AND EMPIRICAL RESULTS

Here, we introduce experiment description, process and results.

4.1. Experiment Description

We extract stock data from U.S.A. Yahoo! Financial website (http://finance.yahoo.com) for 10 years. The data period is between 1996/1/1 and 2005/12/31. The data count and index levels at the ends of years are as in Table II.

We refer to some researches and adopt 14 technical indicators as input variables [2]. They include Different (DIF), Moving average convergence and divergence (MACD), Relative strength (RS), Relative strength index (RSI), Relative strength volume (RSV), K line (K), D line (D), J line (J), Psychological line (PSY), BIAS, Momentum (MTM), Williams overbought/oversold index (WMS), AR, and BR.

In moving window, data of a past period is treated as training data and data after that period as testing data to form a window. The period of training data moves subsequently to form another window. In seasonal moving window, the training data and testing data are from same seasons of various

years. The data of previous year is treated as training data. The data of the same season of later year is treated as testing data (see Fig. I). The advantage of moving window model is that it

has more training and testing data sets, so that the average of

TABLE II.

INDEX LEVEL AT END OF A YEAR FOR EMPIRICAL DATA

(NUMBER OF TRANSACTION DAYS: 2,519)

Date	Stock Index
Dec 31 1996	740.74
Dec 31 1997	970.43
Dec 31 1998	1,229.23
Dec 31 1999	1,469.25
Dec 31 2000	1,320.28
Dec 31 2001	1,148.08
Dec 31 2002	879.82
Dec 31 2003	1,111.92
Dec 31 2004	1,211.92
Dec 31 2005	1,248.29

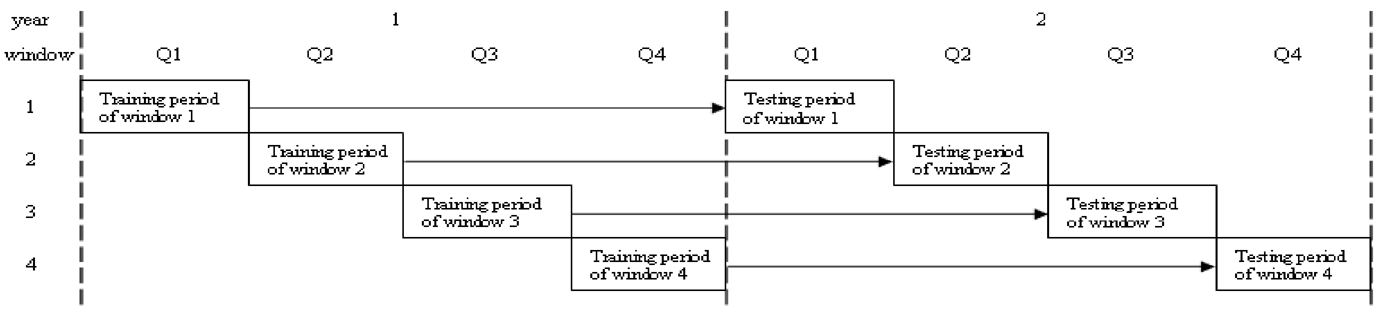

Fig. 1. Seasonal moving window model

all data sets can be more representative. Moreover, the periods of training data and testing data of moving window are close, thus the data relativity is worth referencing. In this study, if the data in a certain season of a certain year is treated as training data, then the data in the same season of the following year is used as testing data of the same window. The data period is 10 years and 4 windows are formed each year from the first to the ninth year. Accordingly, there are 36 seasonal moving windows.

4.2. Experiment Process and Result

In each window, the best cluster number of each technical indicator is derived through Fuzzy c-means, so as to calculate the coincidence rate and determine number of trading days used to calculate technical indicator values. There are 36 seasonal moving windows. As an example, data of the first window is shown in Table III. Through FCM method, the best cluster number for D and J is 4, that for DIF, K and MTM is 5, and that for others is 6. After applying the composition of relations and defuzzication methods, the coincidence rates are in the range from 0.58 to 0.74. The average is 0.70. Number of trading days used to calculate technical indicator values are between 4 and 53 days except K, D, J. Take MACD as an example, its number of trading days used to calculate technical indicator values is 4, that is, when calculating the MACD value of a day, the stock market data of the 4 days before the day should be used. In addition, Stochastic Oscillator (KD) always adapts only one day before to calculate its value. Therefore, the number of trading days used to calculate technical indicator values is always 1.

The optimal antibody for each moving window is shown as Table IV. We can see that the bits of each selected antibody, its corresponding technical indicator name, and its fitness value. The number after the technical indicator name is the optimal number of trading days to calculate the technique indicator value, such as RSV_4 of the second window.

The final experimental results are shown as in Table V. We can see that the yearly earning rate of the proposed fuzzy AIA-SVR approach is 15.07% for USA S&P 500 which outperforms buy_and_hold method by 7.31%. We can find that the number of trading days used to calculate technical indicator values and optimal combination of input technical indicators should be various.

TABLE III.

Technical Indicator	DIF	MACD	RS	RSI	RSV	K	D	J	PSY	BIAS	MTM	WMS	AR	B
Optimal number of trading days used to calculate	22	4	50	14	22	1	1	1	53	17	44	22	42	4
Optimal cluster number	5	6	6	6	6	5	4	4	6	6	5	6	6	
Coincidence rate	0.70	0.70	0.71	0.72	0.71	0.64	0.58	0.69	0.70	0.74	0.74	0.71	0.70	0.7

V. CONCLUSIONS

We propose an approach of artificial immune algorithm, fuzzy theorem, support vector regression method, and seasonal of buy-and-hold method. Moreover, as seen we find that the

TABLE IV.

OPTIMAL ANTIBODY FOR EACH MOVING WINDOW

Window	Antibody	Corresponding Technical Indicator	Fitness Value
1	00000001000000	J	1.0598
2	00001000000100	RSV_4, WMS_4	1.2428
3	00001000100001	RSV_52, PSY_11, BR_49	0.8512
4	00001000000000	RSV_3	1.3380
5	01000000000100	MACD_49, WMS_4	1.0239
6	10000010000000	DIF_22, D	1.2023
7	01000000000011	MACD_10, AR_17, BR_17	1.1660
8	00101000000000	RS_38, RSV_40	0.8895

9	11010010000000	DIF_51, MACD_13, RSI_23, D	1.0475
10	00000101000000	K, J	1.2034
11	10000000000000	DIF_55	1.1115
12	10000010000000	DIF_31, D	1.2929
13	01000001000000	MACD_14, J	1.1589
14	00000110000000	K, D	1.2020
15	01000000000000	MACD_17	0.7756
16	11000100000000	DIF_15, MACD_16, K	1.1829
17	10000100000000	DIF_30, K	0.7952
18	00000011000000	D, J	1.0205
19	11000010000000	DIF_21, MACD_16, D	0.9501
20	01000000000000	MACD_6	1.0935
21	00000011000000	D, J	0.6522
22	10000000001000	DIF_33, MTM_23	1.1774
23	10000010100000	DIF_10, D, PSY_36	0.9358
24	11000001000000	DIF_20, MACD_49, J	1.1594
25	00000010000000	D	1.0331
26	00000100000000	K	0.8925
27	10000110000000	DIF_19, K, D	1.2178
28	01000100000000	MACD_22, K	1.1943
29	10000000000000	DIF_12	0.8092
30	01000000000000	MACD_21	1.1498
31	00000100000000	K	1.0968
32	01000100000000	MACD_7, K	1.1704
33	10000000000000	DIF_29	1.0942
34	01000010000000	MACD_4, D	1.1051
35	11000010000000	DIF_9, MACD_12, D	0.9111
36	10100000000000	DIF_22, RS_27	0.8745

TABLE V.
THE EMPIRICAL RESULTS

Method	Earning Rate
Fuzzy AIA-SVR	15.02%
Buy-and-hold	7.71%

number of trading days used to calculate technical indicator values and optimal combination of input technical indicators should be an important issues to discuss.

The future directions of this research can be as follows:

(1) In order to reveal more trancsaction messages in stock market, future studies can employ fuzzy rules to combine different fuzzied indicators to provide more messages to decision-making model of the stock market.
(2) The proposed approach can be revised to explore other derivatives, such as individual share or warrant.
(3) Other artificial intelligent techniques can be used to study the seasonality of stock market, such as artificial neural network, grey theorem, and genetic algorithm.

References

[1] G. Armano, A. Murru, F. Roli, "Stock market prediction by a mixture of Genetic-Neural experts," *Int. J. Pattern Recognit. Artif. Intell.*, Vol. 16, 2002, pp.501-526.
[2] L. Cao, F.E.H. Tay, "Financial forecasting using support vector machines," *Neural Comput. Appl.*, vol. 10, 2001, pp.184-192.
[3] P.C. Chang, C.H. Liu, "A TSK type fuzzy rule based system for stock price prediction," *Expert. Syst. Appl.*, vol.34, 2008, pp.135-144.
[4] D. Dasgupta, "Advances in Artifical Immune Systems," *IEEE Computational Intelligence Magazine*, vol.1, 2006, pp.40-49.
[5] L.N. De Castro, J. Timmis, "Artificial Immune Systems: A Novel Paradigm to Pattern Recognition," *J. M. Corchado, L. Alonso, and C. Fyfe (eds) Artificial Neural Networks in Pattern Recognition, SOCO-2002*, University of Paisley, UK, 2002 pp.67-84.
[6] L.N. De Castro, J.V. Zuben, "Learning and Optimization using Clonal Selection Principle," *IEEE Trans on Evolutionary Computation, Special Issue on Artificial Immune Systems*, vol. 6, 2002, pp.239-251.
[7] H. Dourra, P. Siy, "Investment using technical analysis and fuzzy logic," *Fuzzy Sets and Systems, vol.*127, 2002, pp.221-240.
[8] M. Glickman, J. Balthrop, S. Forrest, "A Machine Learning Evaluation of an Artificial Immune System," *Evolutionary Computation Journal*, vol. 13, 2005, pp.179-212.
[9] C.L. Huang, C.Y. Tsai, "A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting," *Expert Syst. Appl.*, vol.36, 2009, pp.1529-1539.
[10] W. Huang, Y. Nakamori, S.Y. Wang, "Forecasting stock market movement direction with support vector machine," *Comput. Oper. Res.*, vol.32, 2004, pp.2513-2522.
[11] R. Khemchandani, Jayadeva, S. Chandra, "Regularized least squares fuzzy support vector regression for financial time series forecasting," *Expert Syst. Appl.*, vol.36, 2009, pp.132-138.
[12] K.J. Kim, I. Han, "Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index," *Expert Syst. Appl.*, vol.19, 2000, pp.125-132.
[13] K.J. Oh, K.J.Kim, "Analyzing stock market tick data using piecewise nonlinear model," *Expert Syst. Appl.*, vol.22, 2002, pp.249-255.
[14] L.A. Zadeh, "Fuzzy sets," *Information and Control*,vol.8, 1965, pp.338–353.

國科會補助計畫衍生研發成果推廣資料表

日期:2011/10/24

<table>
<tr><td rowspan="3">國科會補助計畫</td><td colspan="2">計畫名稱: 使用人工智慧的方法來探討金融衍生性市場之認購權證的動態變化</td></tr>
<tr><td colspan="2">計畫主持人: 邱登裕</td></tr>
<tr><td>計畫編號: 99-2221-E-216-052-</td><td>學門領域: 人工智慧</td></tr>
<tr><td colspan="3">無研發成果推廣資料</td></tr>
</table>

99 年度專題研究計畫研究成果彙整表

計畫主持人：邱登裕	計畫編號：99-2221-E-216-052-
計畫名稱：使用人工智慧的方法來探討金融衍生性市場之認購權證的動態變化	

成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	1	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	2	2	100%	人次	
		博士生	3	3	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	無

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：□已發表 □未發表之文稿 ■撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：（以 100 字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

對於學術研究之貢獻

(1)藉由 GM 模型與 GM-C3LSP 模型的加入，探討整體預測能力的改良效果。

(2)藉由基因演算法動態挑選輸入變數，探討對於認購權證具有影響力之變數，並評良 GM-C3LSP 模型的引進是否有效果。

對於參與之工作人員可獲之訓練

(1)參與人員可從中學習到變數的收集、整理以及其它前置處理的過程，以瞭解如何處理收集來的資料。

(2)在模型的開發過程中，培養參與人員對整體系統設計流程的熟稔度。

(3)在模型進行的同時，培養參與人員瞭解欲探討之變數的重要程度，對於實驗設計也會更加熟悉。

(4)強化相關專業技術培養，讓相關人員於未來可以發展出更多關聯性的研究與學術作品，有助於政府所提倡的資訊技術研究與發展。

(5)強化相關研究之文獻探討，使其參與人員可以瞭解近年來相關研究的進展情形，進而提出更不錯的想法於本模型上。

(6)培養參與訓練人員之分析能力，使其人員可以思考更多相關領域之分析與設計，有助於學術界相關作品之研究。

(7) 培養參與人員書面文件的撰寫，如何寫出一篇讓人清楚而又容易瞭解的文件，以及訓

練參與人員對外發表的能力。