基因演算法演化交易策略在股票買賣點的預測

邱寬旭

真理大學工管理與經營資訊學系 副教授 kschiu@email.au.edu.tw

摘要

本研究探討使用基因演算法演化交易策略在股票買賣點的決定。基因演算法演化類神經網路的權重、停損點和停利點。基因演算法 通常使用二進位的染色體,本研究使用浮點數的染色體。浮點數的使用保有最短的界限長度 (每個變數只用一個),如此可防止交換和突變對染色體的破壞。

實驗結果驗證基因演算法演化交易策略 能有效和穩定的做出股票買賣的決定。

關鍵詞:股價預測、類神經網路、基因演算法、 交易策略

Abstract

This paper discusses the use of genetic algorithms (GAs) to evolve the trading strategies for making stock buy and sell decisions. Genetic algorithms evolve the weights for neural networks as well as the parameters for the stop-loss points and the stop-gain points. The GAs usually use binary chromosomes for coding solutions. This using floating-point paper proposes chromosomes. This provides coding schemata of shortest defining length, one for each parameter, which prevents disruption by crossover and mutation.

The experimental results show that the evolutionary trading strategies are both effective and robust for making stock buy and sell decisions.

Keywords: Stock Price Forecasting, Neural Networks, Genetic Algorithms, Trading Strategies.

1. 前言

買賣股票是一種高風險的投資,在幾個月內甚至短短幾週,股價相差往往可以由數成到數倍計算。以最近發生的金融海嘯和不景氣,很多股價由百元以上跌至 30 元左右的不在少數。而在這波空頭走勢之前,很多股票也由數

十元漲至百元以上。在此誘因之下,投資人趨之若鶩。如能在低價時買進,而在高價時賣出,獲利必定可觀。相對的,若是投資時機不對,慘遭套牢也是常有的事。

因此本研究探討使用基因演算法演化類神 經網路、停損點、和停利點,以幫投資者作出 更客觀的投資決策。

本研究分為五部分,第二部分為文獻探討,第三部分為研究方法,第四部分為實驗設計與結果,最後一部分為結論。

2. 文獻探討

效率市場假設(Efficient Market Hypothesis, EMH) 是指市場中所有可能影響股票漲跌的 因素都能即時且完全反應在股票漲跌上面,Fama 依市場效率性質提出弱式(weak form)、半 強式(semi-strong form) 、及強式(strong form) 三種形式[1][2][3]。

1. 弱勢效率市場假設

目前股票價格已充分反應過去股票價格 所提供各項資訊。此市場資訊包括股票的歷史 價格、報酬率和交易量等。所以,投資人無法 運用各種方法對過去股票價格進行分析,再利 用分析結果來預測未來股票價格,投資者無法 再利用過去資訊來獲得超額報酬。所以,弱勢 效率越高,若以過去價量為基礎的技術分析來 進行預測效果將會十分不準確。

2. 半強勢效率市場假設

目前股票價格已充分反應於所有公開資訊上,這些公開資訊包括弱式效率市場所反映的市場價量資訊以及非市場資訊(如股利分配公告、本益比、財務資料、總體新聞事件、政治事件)。所以,投資人無法利用情報分析結果來進行股票價格預測而獲取超額報酬。亦即技術分析和基本分析預測股票價格是徒勞無功。

3. 強勢效率市場假設

目前股票價格充分反應了所有已公開和未公開(內線消息)的所有資訊。雖然資訊未公開,但投資者能利用各種管道來獲得資訊,因此,未公開的消息,實際上是已公開的資訊且已反應於股票價格上。此種情形下,投資者也無法因擁有某些股票內幕消息而獲取超額報酬。

然而也有很多的投資者使用歷史資料作為 買入和賣出股票的依據,這些投資者將歷史料 轉換成圖形或技術指標的形式表示,用以預測 未來股票價格的走勢和變化程度,作為買賣股 票的依據。近年來更有許多的研究,利用人工 智慧的方法根據這些技術指標,用以預測股價 的買賣點。這些常用的人工智慧的方法有類神 經網路和模糊理論,或結合基因演算法進行分 析[4][5][6][7][8]。

股票市場很容易受到國際股市、政治、及 經濟等等的影響,突發的事件,會立即反應到 當日的股價,因此停損點及停利點的設定是很 重要的。而前述的研究使用的技術指標,根據 的是歷史資料,即使使用這些指標來訓練類神 經網路,由於技術指標對突發事件反應慢,且 經常發生鈍化,使用類神經網路來決定買賣 點,對詭譎多變的股票市場,獲利並不容易。

因此,本研究假設股票市場有時是理性,有時不是理性,探討以基因演算法演化類神經網路、停損點、和停利點,類神經網路專注於 買點的決定,使用停損點和停利點來減少損失 和確保獲利。

3.研究方法

類神經網路和基因演算法是本研究使用 的兩個主要方法,下面四個小節將簡單介紹這 兩個方法、兩者的結合、以及本研究使用的方 法。

3.1 類神經網路

類神經網路是模擬生物神經系統的一種電腦模式,這種電腦模式是由許多的處理單元 (Processing Units)所組成,處理單元又被稱之為神經細胞(Neurons),每個處理單元之間有不同強度權重(Weights)互相連結。圖1為一類神經元,其架構適足以模擬一生物神經細胞。類神經網路可被建立在特殊的硬體(電腦晶片)或是由軟體模擬。

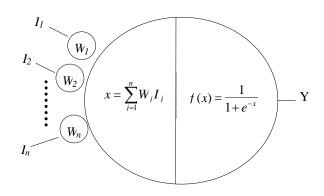


圖 1 類神經元 (Artificial Neuron)

在圖 1 中,輸入訊號乘以權重(Weights) 後傳入神經元,然後所有輸入到神經元的訊號 加總在一起,而加總結果稱為神經元作用程度 (Activation Level),此一作用值(Activation) 再經由一非線性的函數(Sigmoid function)產 生神經元輸出。

生物腦神經系統具有學習的能力,例如幼兒可經由學習而辨識文數字。為了模擬此一特性,類神經網路使用學習策略。學習(Learning)就是調整 Weights 為適當的權重的過程,學習的目的即在尋找適當的 Weights 以便將輸入訊號對映到適當的輸出。三種廣泛被使用的學習策略為:

- 1. 監督式學習 (Supervised Learning)
- 2. 非監督式學習 (Unsupervised Learning)
- 3. 增強式學習(Reinforcement Learning)

監督式學習利用一組輸入和期望的輸出訓練範例,輸入的訊號傳入神經元並產生神經元輸出(如圖 1 所示),而期望的輸出值和神經元輸出的差異則被用來調整 Weights,這種需用到期望輸出值來達到學習的目的,稱之監督式學習。

非監督式學習則不需要期望輸出值,學習 的過程只要輸入值,此種類神經網路根據輸入 值之間的相似性分類,以達到學習的目的。 增強式學習則是介於前兩者之間,這種學習方法和非監督式學習一樣,並無期望輸出值,但卻有一個訊號值可指出網路的成功與否,這種嘗試錯誤法學習(Trial-and-Error)是根據網路的表現回饋值[9][10][11]。本研究使用增強式學習。

3.2 基因演算法

基因演算法是一種模擬生物進化的搜尋方法,該方法將欲搜尋的問題解答編碼成染色體(通常是固定長度的字串),染色體又被稱之為個體(individual),基因演算法隨機產生個體初始值,這些個體的集合稱之為個體群(population)。

使用者需定義一些可量化的評估準則來模擬生物進化的環境,在個體群中的每一個染色體在解碼之後,經由這些可量化的準則評估,而給予一個數字代表該個體在該模擬生物進化的環境中的適應程度 (fitness)。

然後,基因演算法使用三個操作程序:選擇(selection)、交換(crossover)、和突變(mutation)。選擇是一種再製程序,個體會被選取和複製(根據 fitness 值),此一程序模擬物競天擇。

而個體被選取的方法也有很多種,美國University of Michigan 的 Holland 教授所提出的 fitness-proportionate selection 是最簡單的方法之一。該方法根據個體的 fitness 相對於整體 fitness 值的比例選取個體。此方法使個體期望被選取的次數能夠相對於個體在 population中的表現,所以高 fitness 值的個體有更多被選取的機會,反之,低 fitness 值的個體更有可能消失。

交換程序根據是否交換的機率(crossover rate)決定是否將選擇出的隨機配對個體(上一代),進行交換配對染色體中的基因(部分字串)。突變程序根據是否突變的機率(mutaion rate),通常此機率的值很小,隨機改變或重新產生基因。

在經過基因演算法的三個操作程序,新的個體群產生,稱之為一代(generation)。這些在隨機產生個體群之後步驟(三個操作程序)會一直重覆,直到符合結束的條件,固定跑幾代是最常使用的結束條件之一 [12]。圖 2 為基因演算法的虛擬碼。

```
\begin{array}{lll} begin\,GA \\ g:=0 & \{ \ generation\ counter\ \} \\ Initialize\ population\ P(g) \\ Evaluate\ P(g) & \{ \ compute\ individuals' \} \\ fitness\ values\ \} \\ while\ not\ satisfy\ the\ stop\ conditions\ do \\ g:=g+1 \\ Select\ P(g)\ from\ P(g-1) \\ Crossover\ P(g) \\ Mutate\ P(g) \\ Evaluate\ P(g) \\ end\ while \\ end\ GA \end{array}
```

圖 2 基因演算法虛擬碼

3.3 基因演算法演化類神經網路

最早使用遺傳演算法演化類神經網路權重的應用是 Montana 和 Davis [13],他們把它用在海底音波的偵測和分析。之後有許多的應用使用此一方法,相關介紹的文章可參考 [14][15]。

圖3為一簡單的染色體解碼成類神經網路的例子,該染色體使用浮點數,共有六個基因。要將染色體解碼成類神經網路的權重,染色體中的基因按順序被讀取而成為權重,該權重順序由輸出層到輸入層(本例不考慮偏移權重一bias)。遺傳演算法以三個操作程序(selection、crossover、和mutation)演化取代倒傳遞的學習法,而類神經網路則根據該網路的表現,並將該表現量化傳回給遺傳演算法,成為原染色體的fitness。

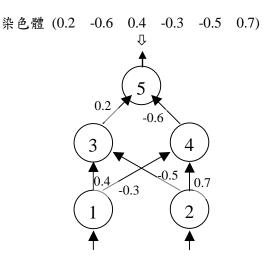


圖 3 染色體解碼為類神經網路的權重

3.4 基因演算法演化交易策略

股票的買點和賣點決定了獲利或虧損。買對了時間點,必須賣對了適當的價位才能確保 穫利;反之,買錯了時間點,也必須設定停損 點,才能減少虧損。股票市場詭譎多變,買點、 停利、和停損是三個確保獲利和減少虧損的重 要關鍵。

賣點的決定則使用停利點和停損點,這兩個點的設定使用下列3個變數:

- 1. 停損點百分比
- 2. 停利點門檻值百分比
- 3. 停利點百分比

停損點百分比為一以買入價為基準的百分比,買入價乘以該百分比,只要開盤價低於此價即以開盤價賣出。停利點門檻值百分比為一個以買入價為基準的百分比,只要股價漲幅(以買入價為基準)超過該百分比,即根據停利點百分比設定停利點。停利點百分比為以最高價為基準的一個百分比,只要開盤價跌幅低於該百分比(以最高價為基準)即賣出。

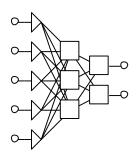


圖 4 類神經網路架構

基因演算法中染色體的編碼使用浮點數,將上述圖4類神經網路架構中的權重及停利點和停損點使用的3個變數編碼為染色體。使用股價的歷史資料模擬,並計算交易成本(手續費和證交稅),統計賺和賠的次數,以賺的次數扣除賠的次數為fitness value。

4.實驗設計與結果

4.1 實驗資料與資料期間

實驗使用的股票為摩台股 (摩根士丹利資本國際公司 MSCI 成分股) 中的 10 支股票 (表 1)。

表 1 研究使用的 10 支股票

股票代號	股票名稱	股票代號	股票名稱
1101	台泥	2301	光寶科
1102	亞泥	2308	台達電
1216	統一	2311	日月光
1301	台塑	2317	鴻海
1303	南亞	2325	矽品

資料蒐集期間為 2000 年 1 月 4 日至 2007 年 12 月 31 日之調整後(還原權息)日資料,將 資料分為兩部分:訓練資料和測試資料。基因 演算法使用訓練資料演化類神經網路的權 重、停損點、和停利點。經過固定代數演化之 後,取出最優的染色體,並解碼為類神經網路 的權重、停損點、和停利點的交易策略。之後, 使用測試資料模擬,並計算賺賠次數及獲利(或 虧損)。

本研究使用移動視窗法 (Moving windows) [16],將資料切成三個期間,每個期間包含訓練資料和測試資料,訓練資料期間為3年,測試資料期間為3年。表2為三個期間的區間:

表2實驗資料的期間

期	訓練期間		測試期間	
間	開始	結束	開始	結束
1	2000/1/4	2002/12/31	2003/1/2	2005/12/30
2	2001/1/2	2003/12/31	2004/1/2	2006/12/29
3	2002/1/2	2004/12/31	2005/1/3	2007/12/31

4.2 實驗設計

由於本研究必需使用基因演算法、類神經網路、和模擬交易的軟體,使用的基因演算法軟體為 SUGAL, The Sudnerland Genetic Algorithms Library [17],類神經網路、股票買賣模擬軟體、以及和基因演算法結合的部分則以 C 語言開發。

基因演算法的 population size 和 generation 都為 100。染色體編碼使用浮點數的資料型態,此編碼的優點是一個浮點數代表一個基因,以避免 crossover 和 mutation 對基因的破壞,並增進收斂的速度及穩定性。為了避免基因演算法過早收斂 (premature convergence)的問題,本研究使用 scale normalization,參數值為 2.0,如此調整 fitness 的值,使得最好的個體和最差的個體的 fitness 為 2 倍。並使用mutation decay,使突變在最初可以大幅改變基因的值,並逐代減少改變幅度。並將每代最佳的個體保留 (elitism)。

基因演算法從產生個體到三大處理程序 (selection、crossover、和 mutation)的發生機 率等等,都是靠亂數隨機產生,為了客觀起 見,每個期間的程式執行各為 5 次,實驗結果 為 5 次的平均值。

4.3 實驗結果

實驗分為三個期間,每個期間資料又分為訓練資料和測試資料,實驗結果的統計資料為使用訓練完成的類神經網路、停損和停利點,根據測試資料決定買和賣,統計了每支股票的交易次數、賺的次數、賠的次數、每股總獲利、手續費和證交稅、及每股的平均持有成本(買入成本),結果分述於下列二小節:獲利率分析和投資報酬率分析。

4.3.1 獲利率分析

期間1實驗結果如表3所示,該交易次數、賺的次數、賠的次數為執行5次程式結果的平均。交易次數1次的計算為買進之後並且賣出。賺賠的定義為賣出股價減去買入股價,並扣除買賣的手續費和證交稅,若為正數,賺的次數加1;若為負,則賠的次數加1。

表 3 中,鴻海的獲利率最高,5 次執行結果平均為 94.5% (37.6/39.8),統一最低為79.9%,10 支股票平均獲利率為88.8%。

表3 期間1股票賺賠次數

17几 西	交易次數	賺的次	賠的次
股票		數	數
1101 台泥	35.2	32.8	2.4
1102 亞泥	34.0	29.0	5.0
1216 統一	30.8	24.6	6.2
1301 台塑	18.8	17.4	1.4
1303 南亞	21.6	19.8	1.8
2301 光寶科	26.0	23.2	2.8
2308 台達電	32.2	29.6	2.6
2311 日月光	40.0	33.4	6.6
2317 鴻海	39.8	37.6	2.2
2325 矽品	41.6	36.0	5.6

期間2實驗結果如表4所示,南亞獲利率最高,5次執行結果平均為95.4%(16.6/17.4),日月光最低為82.3%,10支股票平均獲利率為88.4%。

表 4 期間 2 股票賺賠次數

நா. க	交易次數	賺的次	賠的次
股票		數	數
1101 台泥	40.4	35.2	5.2
1102 亞泥	38.4	33.6	4.8
1216 統一	38.6	33.8	4.8
1301 台塑	18.2	16.8	1.4
1303 南亞	17.4	16.6	0.8
2301 光寶科	29.6	25.4	4.2
2308 台達電	35.0	31.8	3.2
2311 日月光	43.0	35.4	7.6
2317 鴻海	44.8	41.2	3.6
2325 矽品	48.4	40.0	8.4

期間 3 實驗結果如表 5 所示,南亞獲利率 最高,5 次執行結果平均為 91.9% (27.4/29.8), 日月光最低為 81.8%,10 支股票平均獲利率為 85.5%。

表 5 期間 3 股票賺賠次數

股票	交易次數	賺的次 數	賠的次 數
1101 台泥	45.8	39.4	6.4
1102 亞泥	40.4	34.8	5.6
1216 統一	50.6	41.6	9.0
1301 台塑	26.2	23.8	2.4
1303 南亞	29.8	27.4	2.4
2301 光寶科	32.8	27.4	5.4
2308 台達電	41.6	35.0	6.6

2311 日月光	51.6	42.2	9.4
2317 鴻海	49.6	42.0	7.6
2325 矽品	60.2	50.4	9.8

4.3.2 投資報酬率分析

投資報酬率實驗結果如表 6 所示,其中總 穫利、手續費及證交稅、和每股持有成本是以 一張股票計算,因一張股票為一千股,所以它 們的單位是千元。總獲利為該股票在該期間 (測試期間)所有的交易獲利(或虧損),即 每次賣出該股票的價錢減去買入時的價錢 ,即 每費及證交稅的總和。手續費及證交稅的總行分之 1.425,和賣出股票時的 手續費千分之 1.425 和證交稅千分之 3。每股 持有成本為在測試期間,平均買入該股票的價 格。

表 6 中所示,矽品的投資報酬率最高,5 次執行結果平均為 116.3% ($\frac{25.47-4.44}{18.09}$),台塑最低為 31.4%,10 支股票平均報酬率為 61.5%。

表 6 期間 1 股票獲利

股票	總獲利	手續費	每股持
及示		及稅	有成本
1101 台泥	10.96	2.46	12.04
1102 亞泥	7.29	1.96	9.89
1216 統一	6.71	2.02	11.18
1301 台塑	13.44	3.49	31.68
1303 南亞	15.47	3.11	24.47
2301 光寶科	18.87	4.83	31.71
2308 台達電	28.39	5.89	31.60
2311 日月光	11.06	3.90	16.65
2317 鴻海	68.26	13.67	59.18
2325 矽品	25.47	4.44	18.09

期間 2 實驗結果如表 7 中所示,統一的投資報酬率最高,5 次執行結果平均為 98.2% $\left(\frac{20.51-3.83}{16.98}\right)$,日月光最低為 14.1%,10 支股票平均報酬率為 60.4%。

表 7 期間 2 股票獲利

நா. க	總獲利	手續費	每股持
股票		及稅	有成本
1101 台泥	15.46	3.98	16.94
1102 亞泥	14.58	3.28	14.74
1216 統一	20.51	3.83	16.98
1301 台塑	12.26	3.82	36.21
1303 南亞	15.65	3.50	35.15
2301 光寶科	15.40	6.47	37.77
2308 台達電	51.57	9.54	46.24
2311 日月光	7.91	5.07	20.18
2317 鴻海	111.70	24.09	92.16
2325 矽品	26.92	7.09	24.84

期間 3 實驗結果如表 8 中所示,亞泥的投資報酬率最高,5 次執行結果平均為 85.2% $\left(\frac{28.28-6.18}{25.93}\right)$,光寶科最低為 9.9%,10 支股票平均報酬率為 57.3%。

表 8 期間 3 股票獲利

நா. 西	總獲利	手續費	每股持
股票	總復刊	及稅	有成本
1101 台泥	28.14	7.43	27.64
1102 亞泥	28.28	6.18	25.93
1216 統一	20.08	7.00	23.82
1301 台塑	48.14	8.81	57.30
1303 南亞	42.48	9.43	53.71
2301 光寶科	12.64	8.41	42.64
2308 台達電	69.57	18.43	74.88
2311 日月光	14.23	7.10	23.42
2317 鴻海	85.77	37.74	128.28
2325 矽品	42.52	12.89	36.41

5. 結論

買賣股票是一種高風險的投資,因為股票市場很容易受到國際股市、政治、及經濟等等的影響,突發的事件,會立即反應到當日的股價,因此停損點及停利點的設定是很重要的。

本研究假設股票市場有時是理性,有時不是理性,探討以基因演算法演化類神經網路、停損點、和停利點,類神經網路專注於買點的決定,使用停損點和停利點來減少損失和確保獲利。實驗結果證明此方法可幫助投資者作出更客觀的投資決策,其獲利率平均可達85%到89%,即每次的買賣扣除手續費和證交稅,獲利的機會有85%以上。而三年的平均投資報酬率也介於51%到61%之間。

参考文獻

- [1]徐中琦,投資學,滄海書局,2007。
- [2]謝劍平,投資學—基本原理與實務,智勝文 化事業有限公司,2008。
- [3] Fama, E. F. "Efficient Capital Market: A Review of Theory and Empirical Work," *Journal of Finance*, Vol. 25, pp. 607-636, 1970.
- [4] Chang, P., Liu, C. "A TSK type fuzzy rule based system for stock price prediction," *Expert Systems With Applications*, Vol. 34, Issue 1, pp. 135-144, 2008.
- [5]Liu, J., Kwong, R. "Automatic extraction and identification of chart patterns towards financial forecast," *Applied Soft Computing Journal*, Vol. 7, Issue 4, pp. 1197-1208, 2007.
- [6]Tsang, P., Kwok, P., Choy, S., ...etc., "Design and implementation of NN5 for Hong Kong stock price forecasting," *Engineering Applications of Artificial Intelligence*, Vol. 20, Issue 4, pp. 453-461, 2007.
- [7]Kimoto, T., Asakawa, K., Yoda, M., Takeoka, M., "Stock Market PredictionSystem with Modular Neural Networks," *International Joint Conference* on NeuralNetwork, Vol. 1, pp. 1-6, 1990.
- [8]Kim, K. "Artificial neural networks with evolutionary instance selection for financial forecasting," *Expert Systems With Applications*, Vol. 30, Issue 3, pp. 519-526, 2006.
- [9]Beale, R. and Jackson, T. *Neural Computing, an Introduction*, Adam Hilger, IOP Publishing Ltd:Bristol, 1990.
- [10] Chiu, K. Adaptive Optimization of Intelligent Flow Control, Ph.D. Thesis, School of Computing, Engineering and Technology, University of Sunderland, 1999.
- [11] Haykin, S. *Neural Networks: a Comprehensive Foundation*, Prentice Hall, London, 1994.
- [12] Goldberg, D. Genetic Algorithms in

- Search, Optimization, and Machine Learning, Addison-Wesley, 1989.
- [13] Montana, D. and Davis, L. "Training Feedforward Neural Networks using Genetic Algorithms," *Proceedings of Eleventh International Joint Conference on Artificial Intelligence*, San Mateo, CA, Morgan Kaufmann, pp. 762-767, 1989.
- [14] Schaffer, J. and Whitley, (ed.) Proceeding of the Int'l Workshop on Combinations of Genetic Algorithms and Neural Networks (COGANN-92), IEEE Computer Society Press, Los Alamitos, CA, 1992.
- [15] Yao, X. and Liu, Y. "Towards Designing Artificial Neural Networks by Evolution," *Applied Mathematics and Computation*, pp. 83-90, 1998.
- [16] Skabar, A., Cloete, I. "Neural networks, financial trading and the efficient markets hypothesis," *Proceedings of the Australian Computer Science Conference*, Vol. 4, pp. 241–249, 2002.
- [17] Hunter, A. "Crossing Over Genetic Algorithms: The Sugal Generalised GA," *Journal of Heuristics*, Vol. 4, No. 2, pp. 179-192, July 1998.