

國立臺灣大學電機資訊學院資訊網路與多媒體研究所

碩士論文

Graduate Institute of Network and Multimedia

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

使用 AdaBoost 之台股指數期貨當沖交易系統

Using AdaBoost for Taiwan Stock Index Future Intra-day Trading System

林典南

Tien-Nan Lin

指導教授：呂育道 博士

Advisor: Yuh-Dauh Lyuu, Ph.D.

中華民國 97 年 7 月

July 2008

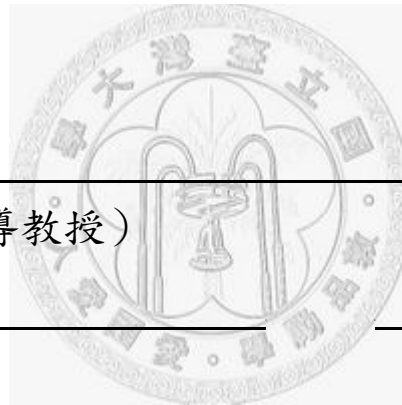
國立臺灣大學碩士學位論文 口試委員會審定書

使用 AdaBoost 之臺股指數期貨當沖交易系統

Using AdaBoost for Taiwan Stock Index Future Intra-day Trading System

本論文係林典南 君 (R95944016) 在國立臺灣大學資訊網路與多媒體所完成之碩士學位論文，於民國 97 年 7 月 28 日承下列考試委員審查通過及口試及格，特此證明

口試委員：



(指導教授)

(簽名)

系主任、所長

(簽名)

誌謝

時光匆匆，兩年的碩士生涯即將畫下句點。能夠順利完成學業，首先要感謝我的指導教授呂育道博士兩年來的教導，引領學生深入了解金融計算的領域，隨時為我們解答研究上的困惑，老師的教誨與督促，使得本篇論文得以圓滿完成。

接下來要感謝愛喝啤酒的承韋老師，無私的傳授業界操作金融商品的技巧，本論文許多靈感來自於老師在金融市場上多年的寶貴經驗。

還要感謝我的家人，父母長輩的關心與體諒，還有我的女友在這段期間不斷的鼓勵，你們的支持是我完成論文最大的動力。

此外還要感謝同學們，學業上、生活上相互幫忙，在論文寫作碰到疑問時熱心解答困惑。

最後要感謝公司主管的體諒，雖然急需人手，仍然同意我以課業為重的請求，讓我得以專心完成學業。



中文摘要

本論文應用AdaBoost於臺灣股價指數期貨當沖交易，我們實作了一個交易系統，可依據盤中即時資料進行實務交易，亦可根據臺灣期貨交易所提供之行情資訊，進行歷史資訊回測模擬交易，檢驗交易策略之可行性。

我們選定臺灣股價指數期貨(以下簡稱臺股期貨)作為研究標的，以2004—2007年每分鐘成交資訊為訓練資料，檢測期間則為2008年1—6月。AdaBoost演算法在物件辨識方面有良好的效果，本交易系統即是用此方法訓練看漲與看跌兩個分類器，藉此找出臺股期貨適當的交易買賣點，提供交易人一個低風險且穩定獲利之投資參考。

中文關鍵詞：臺股期貨、技術指標、程式交易、當沖、機器學習、物件辨識

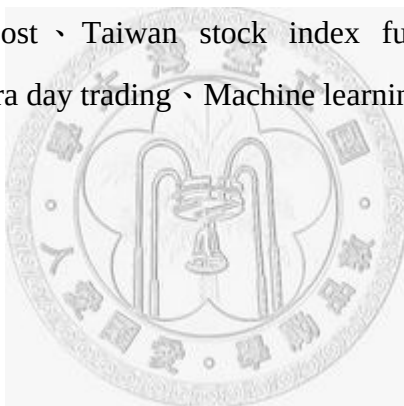


英文摘要

We use AdaBoost for the Taiwan stock index future intra-day trading system. We design a trading system which can trade futures contracts automatically according to real-time streaming quotes. In addition, it allows us to use historical data for back testing and then examines the performance of our trading strategies.

The training data are 1 minute candlesticks from 2004 to 2007. The period of testing starts from January 2008 and ends in June 2008. AdaBoost is an excellent machine learning technique for solving pattern classification problems. We train bull and bear classifiers by AdaBoost. These two kinds of classifiers support our trading system to find the proper time to long or short futures contract. In this thesis we provide low risk and steady profit models for futures traders.

英文關鍵詞：AdaBoost、Taiwan stock index future、Technical index、Programming trading、Intra day trading、Machine learning、Pattern classification



目錄

口試委員會審定書.....	i
誌謝.....	ii
中文摘要.....	iii
英文摘要.....	iv
第一章 緒論.....	1
1.1 研究背景與動機.....	1
1.2 研究目的.....	2
1.3 論文架構.....	4
第二章 文獻探討.....	5
2.1 效率市場假說.....	5
2.2 技術分析.....	6
2.2.1 技術分析理論.....	6
2.2.2 技術指標.....	7
2.2.2.1 K線.....	7
2.2.2.2 平均線.....	8
2.2.2.3 KD隨機指標.....	9
2.2.2.4 CDP逆勢操作系統.....	9
2.2.2.5 相對強弱指數.....	10
2.2.2.6 資金流向指數.....	10
2.2.2.7 能量潮指標.....	11
2.2.2.8 布林帶.....	11
2.2.2.9 停損點轉向指標.....	11
2.2.2.10 指數平滑異同移動平均線.....	12
2.2.2.11 動向指數.....	13
2.2.2.12 乖離率.....	14
2.3 AdaBoost.....	14
2.3.1 AdaBoost概論.....	14
2.3.2 AdaBoost演算法.....	15
2.3.3 Cascade AdaBoost.....	19
2.3.4 相關應用.....	20
第三章 研究方法.....	21
3.1 資料來源.....	21
3.2 辨識系統設計.....	21
3.3 交易系統設計.....	24
3.4 實驗設計.....	28
第四章 實驗結果與分析.....	35
4.1 辨識系統訓練結果.....	35
4.2 辨識系統效能分析.....	36
4.3 交易系統績效分析.....	37
第五章 總結與展望.....	53
5.1 結論.....	53
5.2 未來展望.....	53
參考文獻.....	55

圖目錄

圖一：台股指數期貨年成交量	1
圖二：K線示意圖	8
圖三：錯誤率上限呈指數遞減。	18
圖四：Cascade AdaBoost流程	19
圖五：AdaBoost於人臉偵測應用範例	20
圖六：物件辨識系統設計流程	21
圖七：各層分類器訓練參數設定	23
圖八：交易系統主畫面(日線圖)	24
圖九：交易系統主畫面(5分線圖)	25
圖十：績效模擬畫面(列出每筆交易)	26
圖十一：績效模擬畫面(績效評估表)	26
圖十二：模擬期間損益圖	27
圖十三：報酬率與大盤績效比較圖	27
圖十四：M1、M2、M3 訓練與檢測期	30
圖十五：M4 訓練與檢測期	30
圖十六：M5 訓練與檢測期	31
圖十七：M6 訓練與檢測期	31
圖十八：M7 訓練與檢測期	32
圖十九：M8 訓練與檢測期	32
圖二十：M9 訓練與檢測期	33
圖二十一：交易模型M1 報酬率與大盤績效比較圖	39
圖二十二：交易模型M2 報酬率與大盤績效比較圖	40
圖二十三：交易模型M3 報酬率與大盤績效比較圖	41
圖二十四：交易模型M4 報酬率與大盤績效比較圖	42
圖二十五：交易模型M5 報酬率與大盤績效比較圖	43
圖二十六：交易模型M6 報酬率與大盤績效比較圖	44
圖二十七：交易模型M7 報酬率與大盤績效比較圖	45
圖二十八：交易模型M8 報酬率與大盤績效比較圖	46
圖二十九：交易模型M9 報酬率與大盤績效比較圖	47

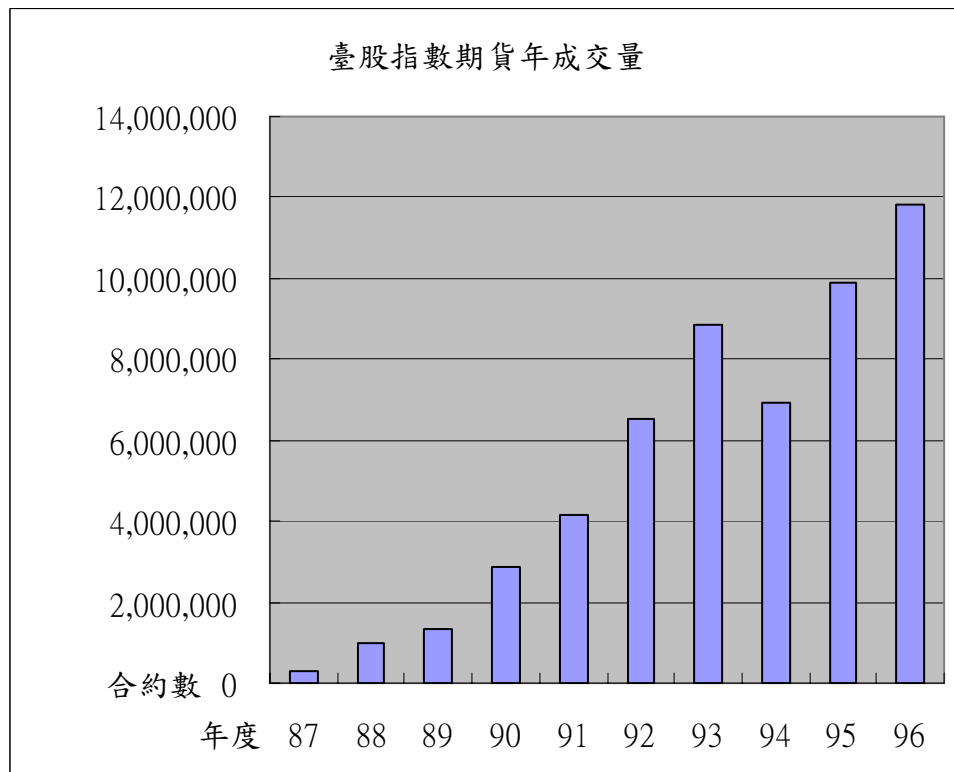
表目錄

表一：當沖與波段系統比較.....	3
表二：AdaBoost演算法.....	16
表三：分類器資料表.....	28
表四：交易系統資料表.....	29
表五：分類器訓練結果.....	35
表六：辨識系統效能.....	36
表七：模擬交易系統使用參數.....	37
表八：績效評估表數據說明.....	38
表九：交易模型M1 績效評估表.....	39
表十：交易模型M2 績效評估表.....	40
表十一：交易模型M3 績效評估表.....	41
表十二：交易模型M4 績效評估表.....	42
表十三：交易模型M5 績效評估表.....	43
表十四：交易模型M6 績效評估表.....	44
表十五：交易模型M7 績效評估表.....	45
表十六：交易模型M8 績效評估表.....	46
表十七：交易模型M9 績效評估表.....	47
表十八：交易模型主要績效數據比較表.....	48
表十九：使用濾網前後提高平均各買賣損益金額.....	49
表二十：交易模型M8-5 口績效評估表.....	50
表二十一：M8-1 口與M8-5 口績效比較.....	50
表二十二：限制 1 口、5 口與辨識系統勝率比較表.....	51
表二十三：交易模型主要績效數據比較表(5 口).....	51

第一章 緒論

1.1 研究背景與動機

臺灣期貨交易所自民國八十七年七月二十一日推出臺灣證券交易所股價指數期貨契約後，由於證券市場避險需求日增，以及期貨交易多空皆宜、高財務槓桿可靈活運用資金等特點，近年來已成為我國交易活絡之金融商品，歷年來成交量亦不斷向上成長(圖 1.1)。



圖一：台股指數期貨年成交量

期貨市場參與人主要為避險者(hedger)、投機者(speculator)與套利者(arbitrager)。避險者在現貨市場持有股票，在期貨市場賣出等值的期貨合約達到規避風險之目的，投機者在市場上直接買賣承擔風險獲取報酬，套利者一買一賣兩種到期時價值收斂的商品，來獲取無風險報酬。避險者有現貨部位卻不願意承擔價格波動之風險，投機者無現貨部位卻願意承擔風險獲取報酬，這種互利關係得以讓金融市場順利發展。投機者的存在承擔了價格風險，活絡市場避免流通性風險，拉近買賣價差降低避險成本，對於整體市場經濟帶來正面效益。

期貨市場上，不管是避險或是投機者，統稱為期貨交易人，因其性質與股票市場不同，故不使用投資人一詞。近年來台股期貨漸成為熱門金融商品，但期貨市場競爭激烈，其為一零合遊戲，有人賺就有人虧損，且根據八二法則，百分之八十的錢會流到百分之二十的人手上，此種高風險、高報酬的商品特性讓許多交易人最後只能認賠出場，難以全身而退。

由於期貨採保證金交易，以台股期貨為例，財務槓桿大約在 10—20 倍之間，若再加上 2007 年 10 月 8 日推出的當沖保證金減半措施，財務槓桿更可達 20—40 倍。期貨市場高槓桿、高波動的特性，如操作不慎，很容易血本無歸。因此期貨交易首重風險控管，嚴守停損與停利之原則。

本研究嘗試以台股期貨為實驗對象，利用程式交易系統化的特性，嚴格控制停損與停利，並採用當沖交易避免留倉風險，如此一來可將每次交易之虧損控制在合理之範圍內，搭配 AdaBoost 演算法，找出適當的買賣時機，使本系統在有限之風險下，仍能獲得不錯的報酬。

1.2 研究目的

大部分期貨交易人使用的程式交易系統屬於波段系統，約佔八成，而剩下的兩成則屬於當沖系統，兩者的差異如下：

	當沖系統	波段系統
交易週期	一天內可進行數次交易，當日沖銷。	持有部位較長，通常為數天。
淨值波動度	小	大
風險	小，沒有隔日跳空風險。	大
交易次數	較多	較少
勝率	較高	較低
盈虧比	較低	較高
優點	無留倉風險	當趨勢來臨時，可賺取大波段行情。
缺點	當日行情波動小，因此交易成本相對較高。	有留倉風險

表一：當沖與波段系統比較

過去的研究多屬於波段系統，其優點是當趨勢出現時順勢交易可獲取不錯報酬。近年來這類的探討已有相當成效，因此我們並不討論波段系統，本研究目的在於採用當沖交易實作交易系統，與波段系統相比，本系統之風險可大幅降低，提供台股期貨交易者一個更安全的交易方式。

本研究另一目的在於檢驗 AdaBoost 演算法在台股期貨市場的有效性。過去使用人工智慧方法預測股價走勢的相關研究方法，大致有類神經網路(neural network)、支持向量機(support vector machines, SVM)、模糊集合(fuzzy set)、基因演算法(genetic algorithm)、灰色系統理論(grey system theorem)、小波轉換(wavelet transform)、群(swarms)、螞蟻演算法(ant algorithm)等。國內文獻中，尚無將 AdaBoost 演算法應用於股價預測之研究，此演算法近年來在物件辨識與電腦視覺領域相當熱門，我們可以把台股期貨股價走勢歸類為三種型態，分別為上漲、下跌、盤整，所謂預測股價走勢即是利用當下現有資訊，判斷未來走勢屬於何種類型，進而產生買賣訊號。

1.3 論文架構

本文共分為五章，各章節分別如下：

第一章 緒論

描述我國期貨市場的發展與期貨商品的性質，進而闡述本交易系統的特性。

第二章 文獻探討

描述效率市場假說、技術分析理論、AdaBoost 演算法。

第三章 研究方法

本研究使用 AdaBoost 演算法訓練上漲與下跌兩種分類器，藉由分類器的辨識功能，產生買賣訊號供交易系統使用。

首先闡述如何把連續的交易資訊轉化為 AdaBoost 演算法處理的樣本，接著把技術分析指標轉換為樣本特徵值，用來訓練辨識系統，最後介紹建構本交易系統的方法。

第四章 實驗結果與分析

討論前一章所介紹之研究方法的實驗結果，檢測辨識系統正確率，與交易系統績效。

第五章 結論與未來展望

針對實驗過程中觀察到的現象，進行分析與結論，並檢討本實驗之貢獻度，與將來可改進之方向。

第二章 文獻探討

2.1 效率市場假說

效率市場假說(Efficient Market Hypothesis, EMH)，由著名經濟學家 Fama 提出，他認為「在短期，效率市場假說的證據是廣泛的，而反證是零星的。」，該假說認為股票的技術分析無效，因為投資人在買賣股票時會迅速有效地利用目前的訊息，且所有會影響股價的資訊都已經反映在股票的價格上面。

Fama 對效率市場理論提出三個基本假設：

1. 當市場上有新的資訊產生，股價會立即調整為新的價位，因此價格變化是取決於新資訊的產生。
2. 新資訊的出現呈隨機性。
3. 市場上每個投資者是理性且追求最大利潤，且每個人對股票分析是獨立互不受影響。

由以上的假設，Fama 依照市場上資訊的效率性，由弱到強提出三種不同程度的型態：

1. 弱勢效率市場(Weak Form Efficiency Market)

目前股價已經充分反應過去的資訊，因此投資人無法運用對過去股價表現的分析(技術分析)來獲取超額報酬，即技術分析無效。

2. 半強勢效率市場(Semi-Strong Form Efficiency Market)

目前股價已經充分反應過去與現在的所有公開資訊，所以投資人無法藉由研究財務報表、經濟環境等基本面資訊來獲取超額報酬，即基本分析無效。

3. 強勢效率市場(Strong Form Efficiency Market)

目前股價已經充分反應所有公開與未公開之消息，雖然消息未公開，但投資人仍能運用各種管道得到消息，因此實際上所有消息皆已經反應於股價上，所以投資人無法藉由得知某些內線消息來獲取超額報酬，即內線消息無效。

由以上三種效率市場定義可知，這三種型態的效率市場彼此之間是程度上的關係，即強勢效率市場包含半強勢效率市場，半強勢效率市場包含弱勢效率市場。近年來學術界對於資本市場是否為效率市場做了許多的實證性研究，已知一般市場並不屬於強勢效率市場，而由於缺乏技術分析與基本分析獲利的決定性證據，學者們認為資本市場大致上符合弱勢與半強勢效率市場。

2.2 技術分析

由效率市場的定義可知，技術分析與效率市場兩者相互矛盾，因此近年來驗證資本市場是否屬於效率市場的相關研究中，最普遍的檢驗是以技術分析的獲利性為研究方向，如技術分析能獲得超額報酬，即是對效率市場假說提出了反證。Park 與 Irwin (2007)研究了近期 95 篇技術分析相關論文，共有 56 篇獲利結果為正向、20 篇負向，以及 19 篇正負參半。雖然目前沒有決定性的證據證明技術分析確實能夠獲利，但也有相當多的實證論文顯示其使用的技術分析策略獲利超越大盤績效。

2.2.1 技術分析理論

技術分析是藉由研究過去市場交易資訊，來預測將來股價漲跌以作為買賣的依據。其基本精神在於「歷史會不斷重演」，技術分析師認為，投資者並不是理性的，他們的交易行為受到市場氣氛影響，且投資者會重複先前市場上的投資行為。舉例來說，某檔股票總是在 50 元附近反轉向上，當這檔股票又回到 50 元附近時，許多投資人會進場買入股票。這些決策或許並不理性，也可能是錯誤的決定，但確實會影響股價的波動。

Edwards 與 Magee (1966)於股價趨勢技術分析一書中提出，技術分析建立於下列的假設之上：

1. 市場價值由供給與需求決定。
2. 市場存在小型波動，但股票價格基本上是順著長期趨勢發展的。
3. 供給與需求變動使得趨勢反轉。
4. 供給與需求的變動可由圖形(股價走勢圖)觀察與判斷。
5. 圖形的型態經常重複

由以上的描述可知，技術分析理論與效率市場假說最大差異在於價格的決定方式，技術分析理論強調供給與需求決定股票價格，且股票價格隨著長期趨勢發展；效率市場假說則認為市場是隨機漫步，股價呈隨機走勢無法預測。舉例來說，可以把技術分析想成一群人舉手表決，而把效率市場想成一群人擲銅板決定，兩種方式都會讓股價朝人多的方向發展，但決定的方式卻完全不同。

技術分析理論是由經濟學的供給與需求發展而來，由於歷史會不斷的重演，藉由資料探勘(data mining)的技術，從大量的技術指標中擷取有用的資訊來判斷趨勢的方向，就能夠在市場上獲得超額報酬。

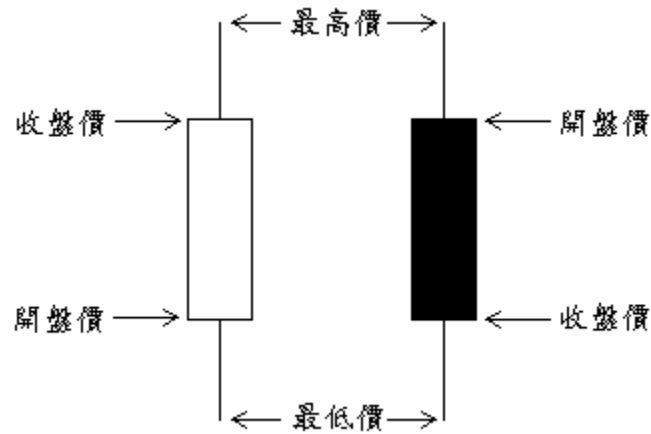
2.2.2 技術指標

技術分析最早起源於十八世紀的日本，當初大阪的米市商人用來紀錄行情所繪製的圖形，經由後代的改良成為今日的 K 線，是現代最常拿來使用的圖形表示工具之一。近代技術分析師則認為道氏理論是現代技術分析的始祖，道氏理論的作者 Charles H. Dow 任《華爾街日報》發行人兼編輯時，每日在專欄選出幾家代表性公司的股價，算出其平均值說明市場趨勢，1902 年道氏去世後，接替他的編輯 William P. Hamilton 整理了其相關見解，並持續發表於《華爾街日報》上，至此道氏理論正式定名。

現今，技術分析主要分為三大類，分別為型態學、指標學與趨勢理論。型態學觀察過去圖形，判斷未來轉折點，例如頭肩頂、頭肩底等；指標學利用量化的數字來判斷轉折，例如 KD、MACD 等；趨勢理論著重在大方向的走勢上，例如道氏理論、波浪理論等。本研究使用的技術分析方法以容易量化的指標學為主，型態與趨勢為輔，由於指標在市場上經過長期的實務驗證與改良，往往會有一種指標數種公式版本的情況，以下分別介紹本研究所使用的技術分析指標，由於本系統不採用一般傳統上由經驗法則定義出的指標使用方法，只取其數值作為系統運算的參數，因此僅針對指標公式部分加以說明，個別指標使用方法可參閱坊間技術分析相關書籍。

2.2.2.1 K 線

K 線是最常被拿來表示股票價格的工具，其圖形如下所示：



圖二：K 線示意圖

如果當天收盤價比開盤價高，以空心矩形表示，稱為陽 K 線；反之當天收盤價比開盤價低則以實心矩形表示，稱為陰 K 線。上引線端點表示當天最高價格所在，下引線端點表示當天最低點位置。

有了 K 線的基本定義之後，即可用 K 線表示每天的開盤價、收盤價、最高價、最低價等四個最關鍵的價位。表示一天內股票關鍵價位的 K 線稱為日線，同理表示一小時內行情的 K 線稱為小時線、15 分鐘內行情稱為 15 分線，本研究採用五種不同時間的 K 線，分別為日線、小時線、15 分線、5 分線與 1 分線來描述行情，不同時間的 K 線描述了不同維度的行情狀況，符合道氏理論主要趨勢、次級趨勢、小型趨勢的精神，或是波浪理論「每一個波都包含了更小規模的波，並且每一個波也都為另一個更大刻度的波所包含。」的觀念。

K 線是所有技術指標的基礎，下面介紹的各種技術指標，大多是把 K 線包含的資訊套入其公式後得到數值，因此我們先介紹 K 線，有了 K 線提供的資訊後才能繼續計算其他的指標。

2.2.2.2 平均線

移動平均線(Moving Average, MA)，是最近一段時間內的股票平均價格，其意義在於表示目前股價的趨勢方向，與這段期間內投資人的平均買入成本。

移動平均線的公式如下：

$$MA(N) = \frac{1}{N} \sum_{i=0}^{N-1} Close_i \quad (2.1)$$

其中 N 為移動平均數的計算天數， $Close_i$ 為前 i 天的收盤價，若 i 為 0 表示當天收盤價。

2.2.2.3 KD 隨機指標

KD 隨機指標(Stochastic Oscillator)是由 George Lane (1957)所提出，這是市場上短期交易常用的一套指標。其意義在於當股票處於上漲波段，收盤價往往接近當日最高價，反之下跌波段時，收盤價往往接近當日最低價，因此可藉由觀察 KD 隨機指標判斷目前股價處於波段的何種位置。

KD 隨機指標的公式如下：

1. 首先計算未成熟隨機值 RSV(raw stochastic value)

$$RSV(N) = \frac{Close - \min(N)}{MAX(N) - \min(N)} \times 100 \quad (2.2)$$

其中 $Close$ 為當天收盤價， $\min(N)$ 為 N 日內最低價， $MAX(N)$ 為 N 日內最高價。

2. 接著計算 K 值與 D 值

$$K = \frac{2}{3} \times K_1 + \frac{1}{3} \times RSV \quad (2.3)$$

$$D = \frac{2}{3} \times D_1 + \frac{1}{3} \times K \quad (2.4)$$

其中 K_1 為前一日 K 值， D_1 為前一日 D 值。

2.2.2.4 CDP 逆勢操作系統

CDP 逆勢操作系統是把前一天的最高價、最低價、收盤價加以計算，分為 5 個價位元作為當日交易的參考。

CDP 的公式如下：

1. 首先計算均價(CDP 值)

$$CDP = \frac{High_1 + Low_1 + 2 \times Close_1}{4} \quad (2.5)$$

2. 再分別計算最高值(AH)、近高值(NH)、近低值(NL)、最低值(AL)

$$AH = CDP + (High_1 - Low_1) \quad (2.6)$$

$$NH = 2 \times CDP - Low_1 \quad (2.7)$$

$$NL = 2 \times CDP - High_1 \quad (2.8)$$

$$AL = CDP - (High_1 - Low_1) \quad (2.9)$$

其中 $High_1$ 、 Low_1 、 $Close_1$ 分別為前一日之最高價、最低價、收盤價。

2.2.2.5 相對強弱指數

相對強弱指標(Relative Strength Indicator, RSI)是市場上普遍使用的技術指標，其概念是把上漲點數視為買方力道，把下跌點數視為賣方力道，利用統計一段時間內買賣雙方的強弱，作為衡量目前是否買賣超之參考。

相對強弱指數的公式如下：

$$RS = \frac{UP}{DN} \quad (2.10)$$

$$RSI = 100 - \frac{100}{1 + RS} \quad (2.11)$$

其中 UP 為 N 日上漲點數平均值，DN 為 N 日下跌點數平均值。

2.2.2.6 資金流向指數

資金流向指標(Money Flow Indicator, MFI)可視為相對強弱指標的成交量加權版本，公式也非常相似。

資金流向指數的公式如下：

1. 首先介紹資金流公式，RSI 指標只考慮收盤價，MFI 指標則是把最高、最低、收盤價平均後再以成交量加權。

$$MoneyFlow = \left(\frac{High + Low + Close}{3} \right) \times Volume \quad (2.12)$$

2. 接著把 N 日內正負資金流分別加總，兩者相除即為類似 RSI 指標的 RS 值，稱之為 MoneyRatio。

$$MoneyRatio = \frac{PositiveMoneyFlow}{NegativeMoneyFlow} \quad (2.13)$$

3. 即可求出 MFI 值。

$$MFI = 100 - \frac{100}{1 + MoneyRatio} \quad (2.14)$$

2.2.2.7 能量潮指標

能量潮指標(On Balance Volume, OBV)是運用成交量累積多寡來衡量目前股票人氣的指標。

能量潮指標的計算公式如下：

若當日收盤價大於前一日收盤價，則

$$OBV = OBV_1 + Volume \quad (2.15)$$

反之

$$OBV = OBV_1 - Volume \quad (2.16)$$

其中 OBV_1 為前一日 OBV 值。

2.2.2.8 布林帶

布林帶(Bollinger Bands)又稱為保歷加通道，指標由三條線組成，中間是平均線，上下兩條線則由平均線(通常為 20 天平均線)加上標準差的倍數(通常為兩倍)得到，其特點在於利用統計學上的標準差求取其信賴區間，其兼具靈活和順勢操作的特點，使它成為資本市場中普遍使用的指標。

布林帶的計算公式如下：

1. 由公式(2.1)可得到平均線

$$BULL_MA = MA(N) \quad (2.17)$$

2. 由標準差公式可得到上下兩條線的值

$$S = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (Close_i - MA(N))^2} \quad (2.18)$$

$$BULL_Up = BULL_MA + 2 \times S \quad (2.19)$$

$$BULL_Down = BULL_MA - 2 \times S \quad (2.20)$$

其中 $MA(N)$ 為 N 天平均線。

2.2.2.9 停損點轉向指標

停損點轉向指標(Parabolic SAR)與濾嘴法則原理非常相似，兩者都清楚定義了買賣的時機，當持有的部位到達停損位置時，會執行平倉並且反向操作的

動作。SAR 的計算起始日為近期的高(低)點，調整係數 AF 的基本值通常為 0.02，每當股價創新高(新低)時 AF 累加 0.02，AF 最高值為 0.2。

停損點轉向指標的計算公式如下：

1. 若開始計算日是上漲波段，則 SAR 值為近期內最低價。

2. 第二天以後的 SAR 為

$$SAR = SAR_1 + AF \times (High_1 - SAR_1) \quad (2.21)$$

3. 若開始計算日是下跌波段，則 SAR 值為近期內最高價。

4. 第二天以後的 SAR 為

$$SAR = SAR_1 - AF \times (High_1 - SAR_1) \quad (2.22)$$

其中 SAR_1 、 $High_1$ 分別為昨日 SAR、昨日最高價。

2.2.2.10 指數平滑異同移動平均線

指數平滑異同移動平均線(moving average convergence/divergence, MACD)是較為穩定的指標，其利用快慢兩條指數移動平均線來研判盤勢，由於 MACD 是經由二次平滑移動平均所得到的值，因此可以消去許多假訊號，當然也相對不敏感，因此適合作為行情已發生後的確認指標。

指數平滑異同移動平均線的計算公式如下：

1. 計算每日初步平均值 DI。

$$DI = \frac{(High + Low + 2 \times Close)}{4} \quad (2.23)$$

2. 計算首日指數平均線，分別取 m 與 n 天。

$$EMA(m) = \frac{1}{m} \sum_{i=0}^{m-1} DI_i \quad (2.24)$$

$$EMA(n) = \frac{1}{n} \sum_{i=0}^{n-1} DI_i \quad (2.25)$$

3. 指數平均線第二天之後的值。

$$EMA(m) = \frac{1}{m+1} (EMA(m)_1 \times (m-1) + DI \times 2) \quad (2.26)$$

$$EMA(n) = \frac{1}{n+1} (EMA(n)_1 \times (n-1) + DI \times 2) \quad (2.27)$$

4. 計算離差值 DIF。

$$DIF = EMA(m) - EMA(n) \quad (2.28)$$

5. 計算首日 MACD 值，取 N 天移動平均數。

$$MACD = \frac{1}{N} \sum_{i=0}^{N-1} DIF_i \quad (2.29)$$

6. 第二天之後的 MACD 值。

$$MACD = \frac{1}{N+1} (MACD_1 \times (N-1) + DIF \times 2) \quad (2.30)$$

2.2.2.11 動向指數

動向指數(directional movement index, DMI)是用來衡量最高價向上移動、或是最低價向下移動的趨勢，可藉由觀察最高價或最低價的變化，瞭解趨勢的動向。

動向指標的計算公式如下：

1. 計算趨向變動值+DM 與 -DM。

$$+DM = High - High_1 \quad (2.31)$$

$$-DM = Low - Low_1 \quad (2.32)$$

2. 尋找股價的真正波幅(True Range, TR)

$$TR = MAX[(High - Low), (High - Close_1), (Low - Close_1)] \quad (2.33)$$

3. 計算方向指標(Directional Index, DI)

$$+DI = \frac{\sum_{i=0}^{N-1} (+DM_i)}{\sum_{i=0}^{N-1} TR_i} \quad (2.34)$$

$$-DI = \frac{\sum_{i=0}^{N-1} (-DM_i)}{\sum_{i=0}^{N-1} TR_i} \quad (2.35)$$

4. 計算趨向指數 DX

$$DX = \frac{|(+DI) - (-DI)|}{(+DI) + (-DI)} \quad (2.36)$$

5. 計算平均趨向指數 ADX

$$ADX = \sum_{i=0}^{N-1} DX_i \quad (2.37)$$

2.2.2.12 乖離率

乖離率(bias indicator)是目前股價與平均股價的差異程度，其意義在於股價會朝趨勢方向前進，因此當股價偏離平均線過大時，往往會回歸平均線的平衡狀態。

乖離率的計算公式如下：

$$Bias = \frac{Close - MA(N)}{MA(N)} \times 100\% \quad (2.38)$$

2.3 AdaBoost

2.3.1 AdaBoost 概論

Boosting 是一種能增進監督式學習(supervised learning)演算法效能的技術，它的起源是基於 Michael Kearns(1988)提出的問題：「是否可由一群弱學習器(Weak Learner)，創造出一個強效學習器(strong learner)。」對於此問題，Yoav Freund 與 Robert Schapire 兩位學者於 1990 年提出了 Boosting 的概念，其方法是藉由不斷加入效能比平均稍微好一點(大於 50%)的分類器，使整體有良好的決策效能。嚴格來說 Boosting 不能稱為演算法，而是一種分類技術的概念，因此之後發展出許多基於 Boosting 觀念的演算法，例如 LPBoost、TotalBoost、BrownBoost、MadaBoost、LogitBoost 等。其中最熱門的就是 Yoav Freund 與 Robert Schapire 於 1997 年提出的 AdaBoost(Adaptive Boosting)演算法，它可以不斷加入新的弱學習器，直到整體效能滿足設計者訂定的低錯誤率標準為止。

AdaBoost 有以下幾點特性：

1. 監督式學習方法(supervised learning)

監督式學習方法是統計學習方法中最簡單常見的一種方式，所謂監督是指給定學習資料時，同時指定資料是屬於哪種類別。

以二元分類為例，監督式學習法步驟如下：

(1) 每筆訓練資料 $(\bar{X}_i, y_i)$ 由特徵向量 $\bar{X}_i = [F_1^i, F_2^i, \dots, F_n^i]$ ，和類別標籤(class label) $y \in \{+1, -1\}$ 組成。

(2) 假設有一函式 f ，對每筆資料皆滿足 $y_i = f(\bar{X}_i)$ 。

(3) 其目標是找到一個近似函式 h ，使得對每筆新增的資料 $(\bar{X}_{new}, y_{new})$ 計算的結果 $h(\bar{X}_{new})$ 近似於 $f(\bar{X}_{new})$ 。

我們稱此近似函式 h 為分類器(classifier)，由上面可知，可藉由分類器把每筆資料分到它認為正確的類別。

2. 整體學習方法(ensemble learning)

整體學習演算法的分類器由許多近似函數 h_i 組合而成，每個近似函數都有其對應的權重 w_i ，與每個函數逼近的程度有關，我們可以用以下的線性函式表示總體分類器：

$$H(\bar{X}) = w_1 h_1(\bar{X}) + w_2 h_2(\bar{X}) + \dots + w_n h_n(\bar{X}) \quad (2.39)$$

其優點在於使用多個近似函數投票出來的結果，其效果會比單一近似函數來的準確且穩定。其他演算法例如類神經網路(neural network)和決策樹(decision tree)，在尋找整體最小值(global minimum)時，往往採用最陡坡降法(gradient descent)，因此找出來的可能是區域極小值(local minimum)，而非整體最小值。如果改使用整體學習方法，也就是把多個區域極小值加權後輸出結果，可以降低錯誤率。

3. 可搭配所有監督式學習演算法

AdaBoost 的分類方式與其他機器學習演算法有很大的不同，它就像一面濾網，把資料分為正面(positive)與負面(negative)兩類，藉由過濾掉負面的資料，而讓正面資料通過的特性，達到分類目的。因此它也可以搭配所有監督式學習演算法，例如 SVM、類神經網路等。在其他演算法後面串接 AdaBoost 分類器過濾負面被誤判(false negative)的資料，提高整體的效能。由於 AdaBoost 是監督式學習演算法，當然也可以把數個 AdaBoost 分類器串接起來，這個方法稱為“Cascade”，在 2.3.3 有詳細的介紹。

2.3.2 AdaBoost 演算法

AdaBoost 演算法如下：

1 Given: $D = \{(x^1, y_1), \dots, (x^n, y_n)\}$, where $x^i \in \bar{X} = [F_1, \dots, F_m]$, $y_i \in Y = \{+1, -1\}$

2 Initialize weights $W_1(i) = \frac{1}{n}, i = 1, \dots, n$

3 For $k = 1, \dots, K$ do

4 Train weak learner C_k with the lowest training error E_k using D sampled according to $W_k(i)$

5
$$E_k = \sum_{i=1}^n W_k(i) |h_k(x^i) - y_i|$$

6
$$\alpha_k \leftarrow \frac{1}{2} \ln \left[\frac{1 - E_k}{E_k} \right]$$

7
$$W_{k+1}(i) \leftarrow \frac{W_k(i)}{Z_k} \times \begin{cases} e^{-\alpha_k} & \text{if } h_k(x^i) = y_i \\ e^{\alpha_k} & \text{if } h_k(x^i) \neq y_i \end{cases}$$

8 End of for loop

9 Return C_k and α_k , for $k = 1, \dots, K$ (ensemble of classifiers with weights)

10 The final strong classifier is:

11
$$H(x) = \begin{cases} 1 & \sum_{k=1}^K \alpha_k h_k(x) \geq \frac{1}{2} \sum_{k=1}^K \alpha_k \\ 0 & \text{otherwise} \end{cases}$$

表二：AdaBoost 演算法

其中第 4 行是對每個特徵 F_j 所產生的機率分布都找一個最小錯誤值

$\varepsilon_j = \sum_{i=1}^n W_k(i) |h_j(x^i) - y_i|$ ，從 $\varepsilon_1, \dots, \varepsilon_m$ 中選出最小的就是本次訓練的錯誤值 E_k ，

而 C_k 就是產生這個最小錯誤值的弱學習器，也是本次訓練所要找的弱學習器。

第 7 行是降低正確分類的權重且增加分類錯誤的權重，此意義在於每次找弱學習器時，是找錯誤值 E_k 最小的，因此增加分類錯誤的權重有助於降低下次再被分錯的機率。更新權重後還必須標準化，使它滿足機率分配的定義：

$$\sum_{i=1}^n W_{k+1}(i) = 1 \quad (2.40)$$

標準化常數 Z_k 是由所有權重更新後的值加總所得：

$$Z_k = \sum_{i=1}^n \left(W_k(i) \times \begin{cases} e^{-\alpha_k} & \text{if } h_k(x^i) = y_i \\ e^{\alpha_k} & \text{if } h_k(x^i) \neq y_i \end{cases} \right)$$

(2.41)第 6 行是權重調整參數，當 $\alpha_k = \frac{1}{2} \ln \left[\frac{1-E_k}{E_k} \right]$ 時，整體錯誤率 E 可以降到

最低，證明如下：

$$E = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1 & \text{if } H(x^i) \neq y_i \\ 0 & \text{else} \end{cases} \quad (2.42)$$

我們可把演算法第 11 行改寫為

$$H(x) = \begin{cases} 1 & f(x) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{其中 } f(x) = \sum_{k=1}^K \alpha_k h_k(x) - \frac{1}{2} \sum_{k=1}^K \alpha_k \quad (2.43)$$

$$\text{因此原式} = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1 & \text{if } y_i f(x^i) < 0 \\ 0 & \text{else} \end{cases} \quad (2.44)$$

又 $e^1 > 1, e^0 > 0$ ，因此

$$E \leq \frac{1}{n} \sum_{i=1}^n \exp(-y_i f(x^i)) \quad (2.45)$$

由演算法第 7 行可寫成

$$W_{k+1}(i) = \frac{W_k(i) \times \exp(-\alpha_k y_i h_k(x^i))}{Z_k} \quad (2.46)$$

把 k 由 $1, \dots, K$ 依次代入，並把每一項相乘可得

$$W_{k+1}(i) = \frac{1}{n} \times \frac{\exp(-y_i f(x^i))}{\prod_{k=1}^K Z_k}, \text{ 代回公式(2.45)} \quad (2.47)$$

$$E \leq \sum_{i=1}^n W_{k+1}(i) \times \prod_{k=1}^K Z_k \quad (2.48)$$

由公式(2.40)可知 $\sum_{i=1}^n W_{k+1}(i) = 1$

$$\therefore E \leq \prod_{k=1}^K Z_k \quad (2.49)$$

由公式(2.41)

$$Z_k = \sum_{i=1}^n \left(W_k(i) \times \begin{cases} e^{-\alpha_k} & \text{if } h_k(x^i) = y_i \\ e^{\alpha_k} & \text{if } h_k(x^i) \neq y_i \end{cases} \right)$$

$$= e^{-\alpha_k} \sum_{i:h_k(x^i)=y_i}^n W_k(i) + e^{\alpha_k} \sum_{i:h_k(x^i) \neq y_i}^n W_k(i) \quad (2.50)$$

$$= e^{-\alpha_k} (1 - E_k) + e^{\alpha_k} (E_k) \quad (2.51)$$

我們要求 α_k 使得 Z_k 值是最小，取微分為 0

$-e^{-\alpha_k} (1 - E_k) + e^{\alpha_k} (E_k) = 0$ ，整理過後可得所求

$$\alpha_k = \frac{1}{2} \ln \left[\frac{1 - E_k}{E_k} \right]，得證。$$

另外，把 α_k 代回公式(2.51)可得

$$Z_k = 2\sqrt{E_k(1 - E_k)} \quad (2.52)$$

由演算法可知 $0 \leq E_k < 0.5$ ，我們令一正數 G_k ，使得 $E_k = \frac{1}{2} - G_k$ ，代入上式

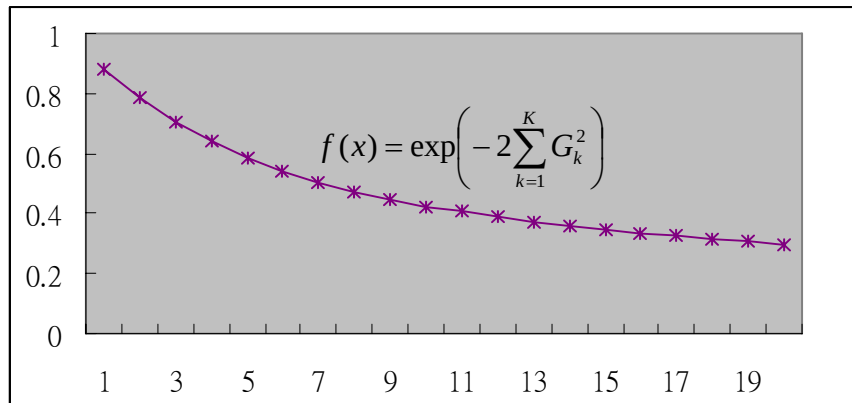
$$Z_k = 2\sqrt{1 - 4G_k^2} \quad (2.53)$$

$$\leq \exp(-2G_k^2) \quad (2.54)$$

代回公式(2.49)可求得錯誤率上限：

$$E \leq \exp\left(-2 \sum_{k=1}^K G_k^2\right) \quad (2.55)$$

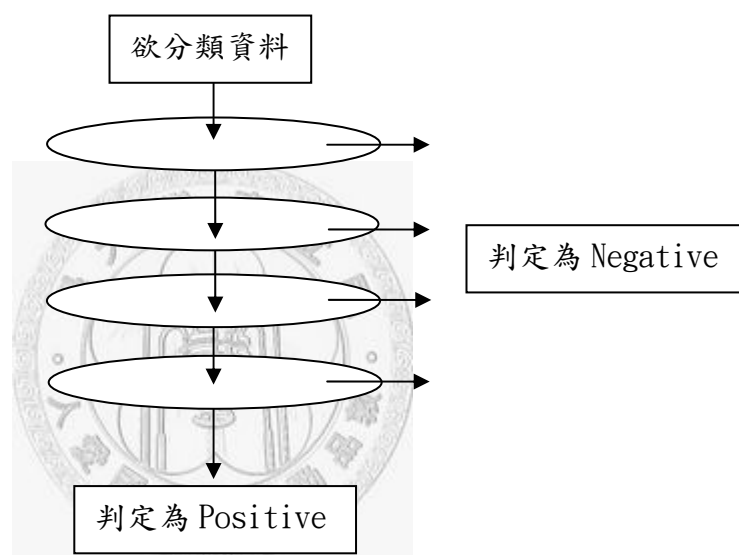
當 K 增加，錯誤率上限會呈指數遞減，因此理論上只要不斷加入新的弱分類器，就可讓錯誤率降低。下圖 X 軸為弱分類器數量，Y 軸為錯誤率：



圖三：錯誤率上限呈指數遞減。

2.3.3 Cascade AdaBoost

AdaBoost 分類器系統加入越多的弱分類器，可使整體錯誤率不斷降低，但大量的分類器會讓運算時間加長。Paul Viola 與 Michael Jones(2001)提出了“cascade AdaBoost classifier”，原本單一的 AdaBoost 分類器改為數個串接起來的 AdaBoost 分類器。前面幾層分類器內的弱分類器數量較少，目的在於快速過濾明顯可以判定為負面的資料，後面幾層內的分類器數量較多，負責處理較難判定的資料，由於大部分的資料在前面幾層就被過濾掉，不需經過後面大量的分類器，可大幅縮短分類時間。



圖四：Cascade AdaBoost 流程

值得注意的是，每層之間的訓練資料有關聯性。首先我們把訓練資料分為兩類，正面(positive)與負面(negative)並訓練第一層分類器，第一層訓練完畢之後，使用通過分類器的資料繼續第二層的訓練，第二層訓練完畢後，使用通過前兩層的資料進行第三層的訓練，依此類推。所謂「通過分類器的資料」是指被分類器判定為正面的資料，因此除了正面資料外，也包含了原本是負面卻被誤判為正面的資料。資料在上層如被判定為負面就丟棄不用，其原因在於如果正面資料在上層被誤判為負面，也就是在上層已經被刷掉，即使本層訓練結果正確也於事無補，因此不需拿來訓練。同理如果負面資料在上層被刷掉，表示已經被正確分類了，因此也不需要拿來訓練下一層。

2.3.4 相關應用

AdaBoost 最常被提及的應用莫過於人臉偵測，Paul Viola 與 Michael Jones(2001)使用 AdaBoost 由圖片中分辨人臉位置，公認為人臉偵測領域最佳解決方法。除此之外，AdaBoost 近年來也廣泛應用於物件辨識、影像偵測、資訊探勘等議題上。



圖五：AdaBoost 於人臉偵測應用範例

金融交易方面，German Creamer 與 Yoav Freund(2006)使用 boosting 家族的 LogitBoost 預測股票趨勢。作者以微軟前 90 日股價日線資訊為訓練資料，判斷當天收盤價屬於漲或跌，以提高其交易系統準確度。

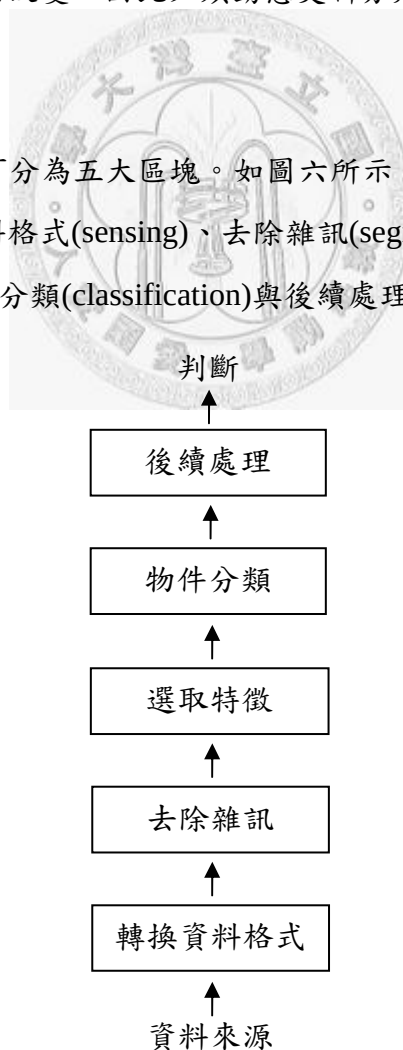
第三章 研究方法

3.1 資料來源

本研究採用臺灣股價指數期貨歷史資料，期間為 2004 年 1 月 1 日至 2008 年 6 月 30 日止，共計 1112 個交易日，333600 筆每分鐘成交資訊，資料來源為臺灣期貨交易所網站(<http://www.taifex.com.tw>)。前四年交易資訊只作為訓練資料用，檢測期間為 2008 年 1—6 月，共計 120 個交易日。由於股價的趨勢會隨著時間而改變，因此本實驗除了檢驗固定分類器績效之外，也檢驗動態分類器績效。固定分類器是用過去一段時間的訓練資料訓練分類器，往後的測試皆使用此分類器而不做更新的動作，適合樣本分佈狀態不隨時間改變。動態分類器則是測試一段時間後，把這段測試過的資料加入訓練樣本重新訓練分類器，由於股價趨勢會隨著時間而改變，因此必須動態更新分類規則。

3.2 辨識系統設計

物件辨識系統設計可分為五大區塊。如圖六所示，資料由下而上(bottom-up)處理步驟分別為轉換資料格式(sensing)、去除雜訊(segmentation)、選取特徵(feature extraction)、物件分類(classification)與後續處理(post-processing)。



圖六：物件辨識系統設計流程

底下依照標準物件辨識系統處理流程說明本實驗的設計方法：

1. 轉換資料格式(sensing)

轉換資料格式是由電腦不可直接處理的形式轉為數位化的資訊，例如紐約證券交易所(New York Stock Exchange)採用人工撮合，再回報數位化的成交資訊。台灣指數期貨開辦之初即採用電子交易，因此省去此一步驟。

2. 去除雜訊(segmentation)

訓練資料來源為期貨交易所提供的每筆成交資料，其中包含十二種商品期貨交易資訊，與不同的到期月份契約。本研究從龐雜的資訊中，抽取出臺指期貨近月合約成交資訊，接著把每筆成交資訊轉為每分鐘成交資訊(1 分 K 線)。1 分 K 線資訊即可分別轉換為 5 分、15 分、小時、日 K 線，因此總共有五種 K 線資訊。

我們定義期貨市場盤中每一分鐘都視為一筆資料，一筆資料包含五種 K 線的最近 10 根 K 棒資訊，因此每一項技術分析指標都有 50 個數值，而我們採用的技術指標有 12 種總計 40 項，藉由這些資訊建立連續性、多維度與多面向的判斷規則。

3. 選取特徵(feature extraction)

由於每種指標可能包含不同參數或細項指標，例如 KD 指標包含了 RSV、K、D、目前交叉關鍵價、下次交叉關鍵價等五項，因此我們把技術指標分為價位指標、擺盪指標、成交量指標與時間四類。

價位指標有 21 項、擺盪指標 16 項、成交量指標 2 項、時間 1 項，依照指標的單位分類。由於除了時間以外，每一項技術分析指標都有 50 個數值，因此每一筆資料當中，價位指標有 1050 個、擺盪指標數值 800 個、成交量指標 100 個、時間有 1 個。

我們認為指標之間的相互關係是一個很重要的特徵，因此除了以上的特徵以外，我們把價位指標兩兩相減，把擺盪指標也兩兩相減、把成交量指標兩兩相除，最後產生了 877226 個特徵。

4. 物件分類(classification)

本研究採用 Cascade AdaBoost 分類物件，訓練的分類器分為上漲與下跌兩種。以訓練上漲分類器為例，選定訓練資料的範圍後，會依據將來的走勢，把

資料分為正面(上漲)與負面(其他情況)兩類，若是下跌分類器，則分為正面(下跌)與負面(其他情況)。

每層的訓練資料是隨機由兩類中各取出 1000 筆，且必須通過整體分類器的篩選。每筆訓練資料包含了 877226 個特徵。在訓練過程中，弱分類器的產生方式是掃過每個特徵，每筆訓練資料對於每個特徵會產生不同的機率分布，而每個機率分布都有一個錯誤值最小的門檻，錯誤值是分錯的資料乘上權重後的加總，所謂弱分類器即是從 877226 個門檻裡面找出最小錯誤值的門檻。每找到一個弱分類器後，要針對這次的分類結果更新每筆資料的權重，把分錯的資料權重加大，對的資料權重縮小，藉由這個動作來降低下次選到錯誤資料的機率。

每層賦予三個參數，分別是最低可接受偵測率(detection rate)、最高可接受誤確認率(false positive rate)與最大弱分類器數量。結束的條件是同時滿足前兩個參數，或達到數量限制。前幾層必須滿足小而快的要求，因此限制數量、犧牲誤確認率，之後則放寬數量限制，使每一層能夠盡量達到前兩項要求。本實驗的分類器訓練參數如下圖七所示，前五層嚴格限制弱分類器數量，同時為了避免因分類器數量不足，造成正面資料在前幾層被丟棄過多的情況，我們把最低可接受偵測率調為 100%，讓正面資料盡可能通過。

Classifier 編號	Positive Detect %	False Positive %	數量限制
0.	100	80	1
1.	100	80	25
2.	100	80	25
3.	100	80	50
4.	100	80	50
5以後	97	50	500
結束條件:	層數	15	OK
Classifier檔案路徑			

圖七：各層分類器訓練參數設定

5. 後續處理(post-processing)

後續處理是考慮除了分類結果之外，影響決策的其他因素。例如本金、風險、時間、持有部位等。本研究的辨識系統只負責依照分類結果產生上漲或下跌的訊號，最後買賣決策留待交易系統解決。

3.3 交易系統設計

本交易系統主畫面如下圖：



圖八：交易系統主畫面(日線圖)

上排按鈕可移動顯示日期、縮放顯示寬度，和選擇不同維度的 K 線交易資訊。交易資訊來源分為資料庫與盤中即時更新，資料庫資料取自於台灣期貨交易所網頁，即時更新資料來源為券商即時報價。

左列為交易文字資訊，上半部為 K 線資訊，包含日期、開盤價、最高價、最低價、收盤價與漲跌。下半部為 KD 技術指標資訊，包含 K 值、D 值、KD 差、本日交叉關鍵價、明日交叉關鍵價。

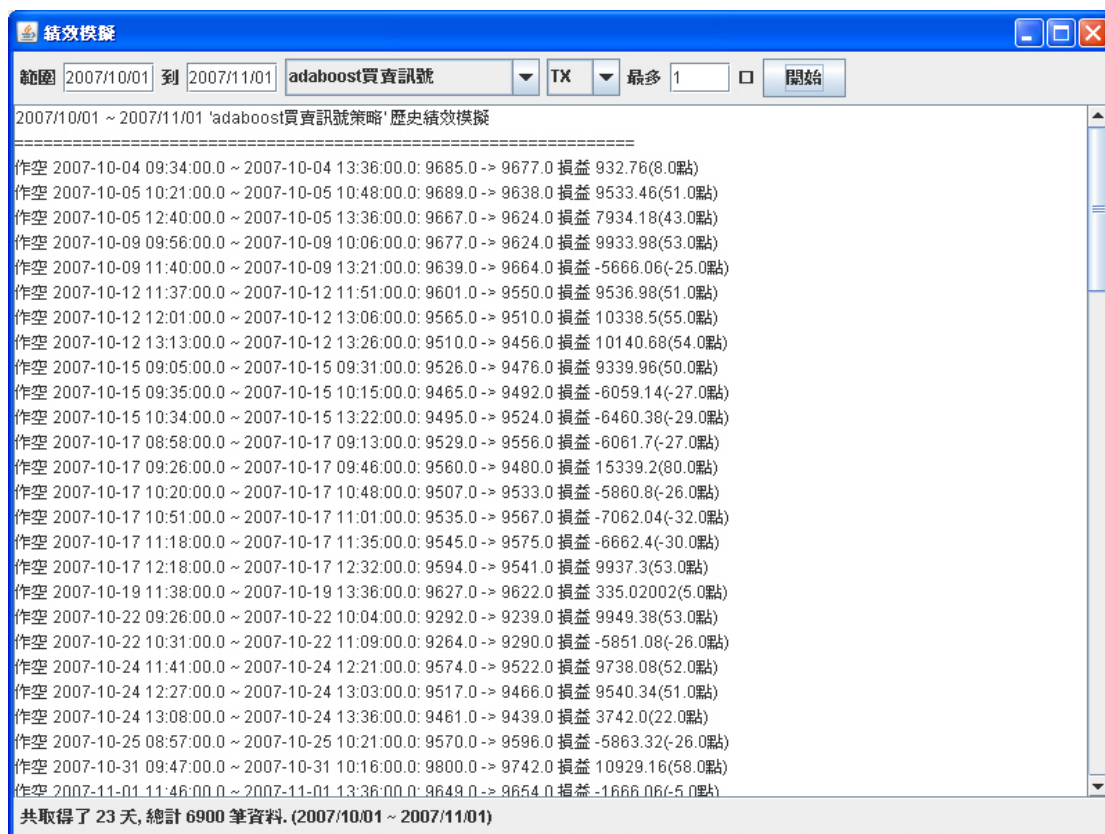
主畫面為臺指期貨 K 線圖，包含 5、10、20、30 日均線，與區域最高價、區域最低價。K 線可選擇 1、5、15 分鐘、小時與日線等五種維度，下圖為 5 分鐘 K 線圖範例。



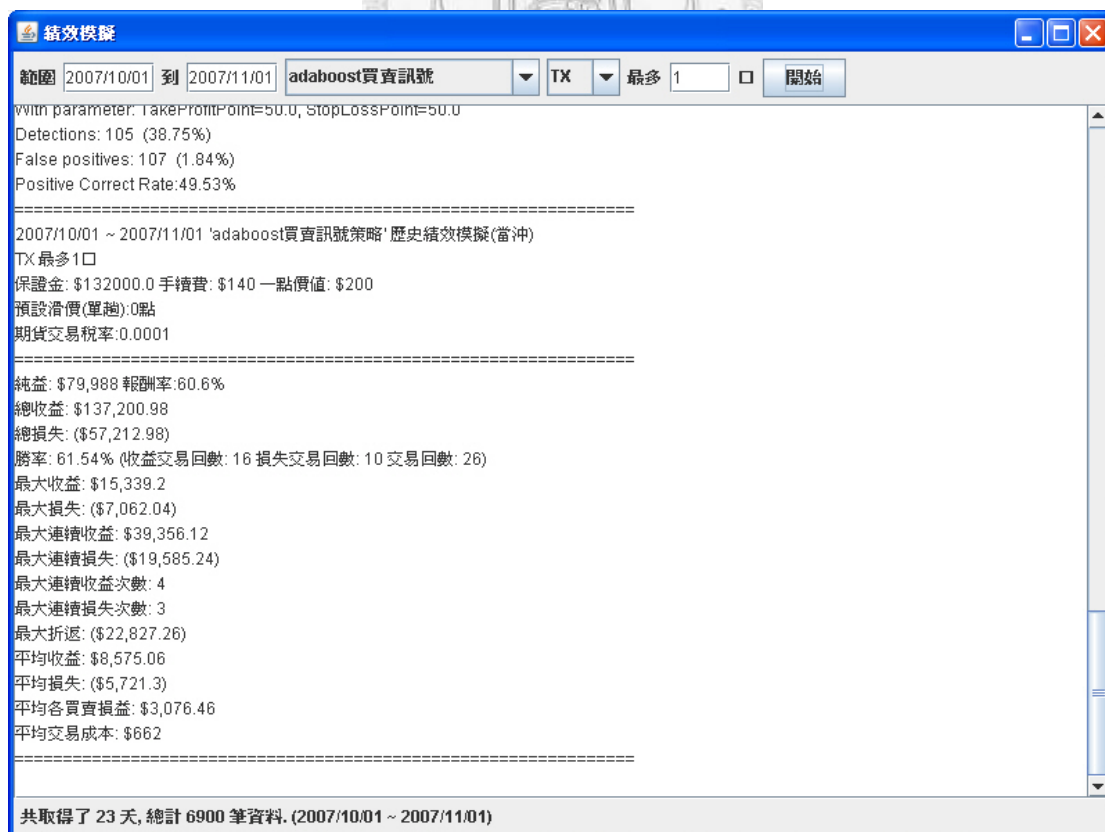
圖九：交易系統主畫面(5分線圖)

除了主畫面之外，本系統提供了績效模擬的功能，讓使用者用歷史交易資訊模擬不同交易策略的交易狀況。

策略績效模擬畫面如下圖十、十一：



圖十：績效模擬畫面(列出每筆交易)



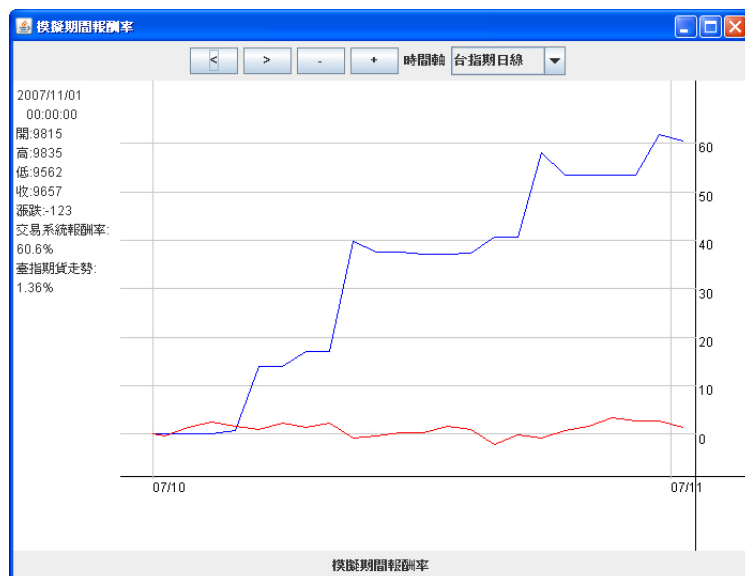
圖十一：績效模擬畫面(績效評估表)

績效模擬畫面的上方分別為模擬期間、交易策略、交易標的(臺指期貨或小型臺指期貨)、部位數量上限，以及開始按鈕。按下開始後，系統會以每次新增一分鐘資訊的方式模擬真實報價狀況，可以模擬當下 K 線每分鐘變動情形，讓回測結果與真實交易更接近。

績效模擬系統可列出每筆成交資訊，包含部位方向、建倉時間、平倉時間、指數變化、損益金額(含交易成本)與損益點數(不含成本)，最後針對整體績效產生績效評估表、損益圖(下圖十二)、報酬率與大盤績效比較圖(下圖十三)。



圖十二：模擬期間損益圖



圖十三：報酬率與大盤績效比較圖

3.4 實驗設計

本實驗採用了數個不同訓練期間的分類器作為績效模擬，各個分類器的資料如下表三：

分類器代號	訓練期間	正面資料條件
B1	2004/01/01 到 2007/12/31 (共 4 年)	漲 50 點
B2	2004/03/01 到 2008/02/29 (共 4 年)	漲 50 點
B3	2004/05/01 到 2007/04/30 (共 4 年)	漲 50 點
B4	2007/10/01 到 2007/12/31 (共 3 月)	漲 50 點
B5	2007/12/01 到 2008/02/29 (共 3 月)	漲 50 點
B6	2008/02/01 到 2008/04/30 (共 3 月)	漲 50 點
S1	2004/01/01 到 2007/12/31 (共 4 年)	跌 50 點
S2	2004/03/01 到 2008/02/29 (共 4 年)	跌 50 點
S3	2004/05/01 到 2007/04/30 (共 4 年)	跌 50 點
S4	2007/10/01 到 2007/12/31 (共 3 月)	跌 50 點
S5	2007/12/01 到 2008/02/29 (共 3 月)	跌 50 點
S6	2008/02/01 到 2008/04/30 (共 3 月)	跌 50 點

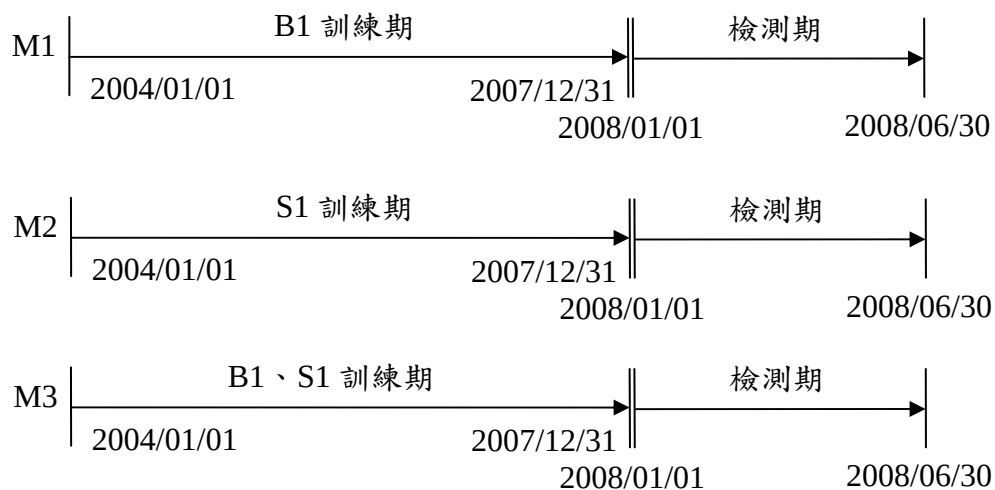
表三：分類器資料表

我們根據四個議題：(一)是否需要定期更新分類器；(二)濾網對效能的影響；(三)只作單邊或多空雙向的績效差異；(四)最大口數限制的影響；設計了九個不同的交易模型(如下表四)，檢測期間皆為 2008/01/01 到 2008/06/30，交易模型使用相同交易策略，差異在於採用不同的分類器。M1、M2、M3 採固定分類器，檢測期間內不更新，其中 M1 只作多、M2 只作空、M3 多空皆宜。M4、M5、M6 採動態分類器，每兩個月更新一次，其中 M4 只作多、M5 只作空、M6 多空皆宜。M7、M8、M9 採動態分類器，每兩個月更新一次，且使用近 3 月分類器作為濾網，其中 M7 只作多、M8 只作空、M9 多空皆宜。

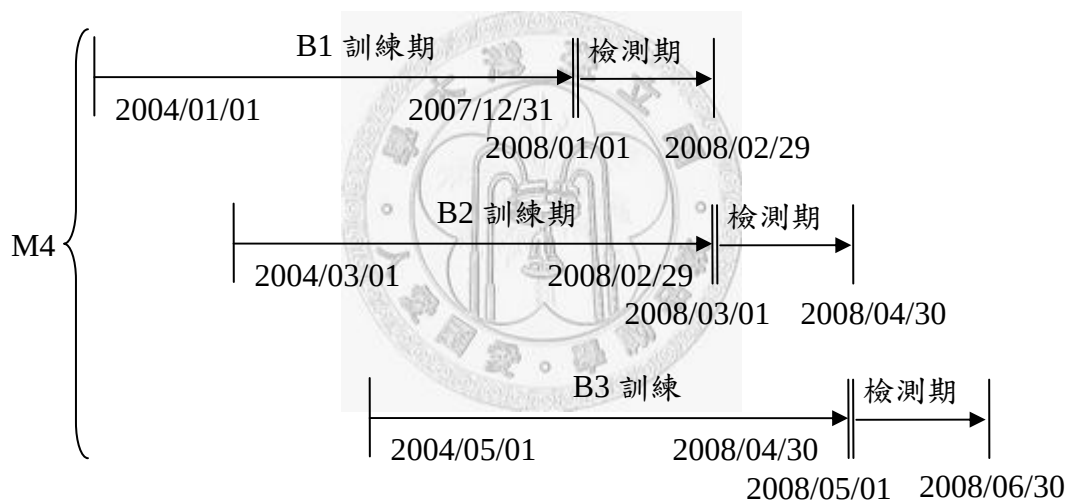
交易模型代號	採用分類器代號	檢測期間
M1	B1	2008/01/01 到 2008/06/30
M2	S1	2008/01/01 到 2008/06/30
M3	B1、S1	2008/01/01 到 2008/06/30
M4	B1	2008/01/01 到 2008/02/29
	B2	2008/03/01 到 2008/04/30
	B3	2008/05/01 到 2008/06/30
M5	S1	2008/01/01 到 2008/02/29
	S2	2008/03/01 到 2008/04/30
	S3	2008/05/01 到 2008/06/30
M6	B1、S1	2008/01/01 到 2008/02/29
	B2、S2	2008/03/01 到 2008/04/30
	B3、S3	2008/05/01 到 2008/06/30
M7	B1+B4	2008/01/01 到 2008/02/29
	B2+B5	2008/03/01 到 2008/04/30
	B3+B6	2008/05/01 到 2008/06/30
M8	S1+S4	2008/01/01 到 2008/02/29
	S2+S5	2008/03/01 到 2008/04/30
	S3+S6	2008/05/01 到 2008/06/30
M9	B1+B4、S1+S4	2008/01/01 到 2008/02/29
	B2+B5、S2+S5	2008/03/01 到 2008/04/30
	B3+B6、S3+S6	2008/05/01 到 2008/06/30

表四：交易系統資料表

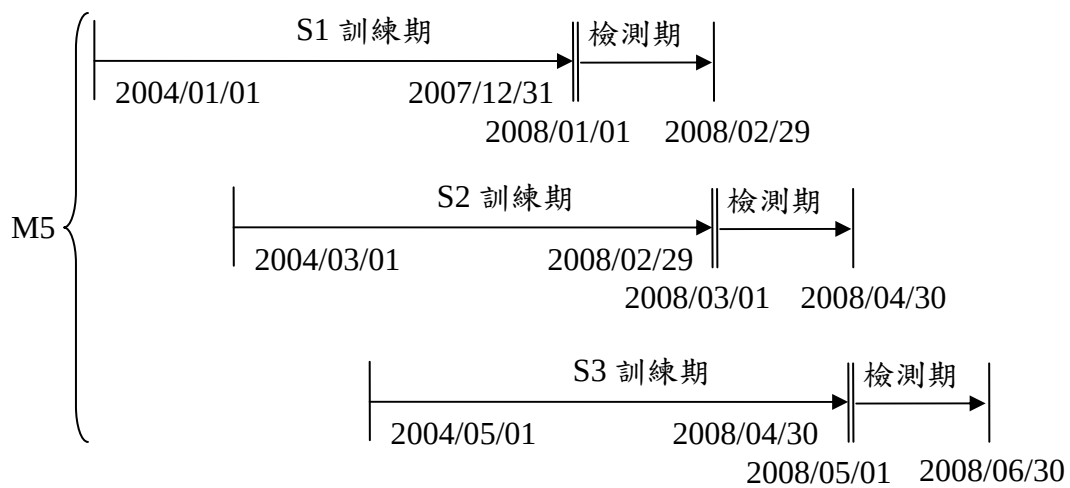
各交易模型使用的分類器訓練期間與檢測期間如下所示：



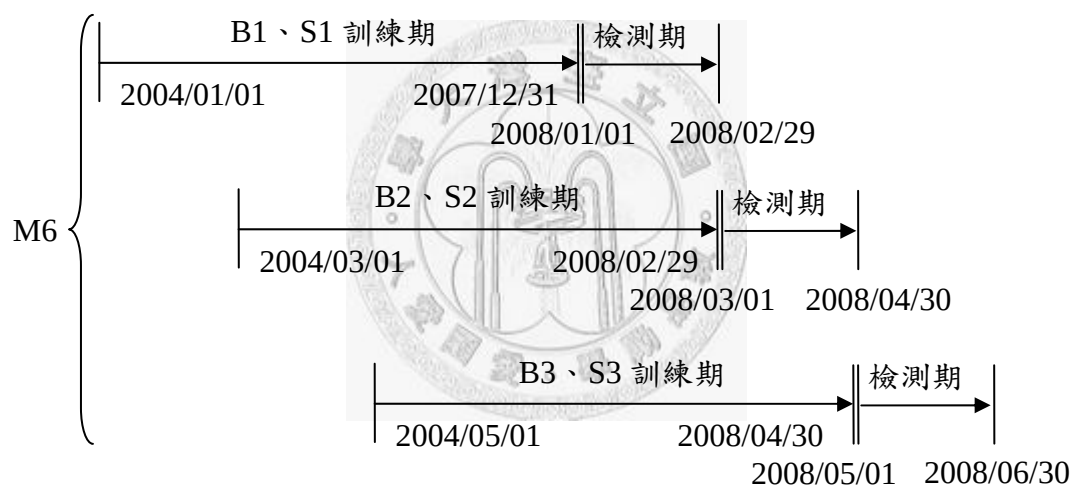
圖十四：M1、M2、M3 訓練與檢測期



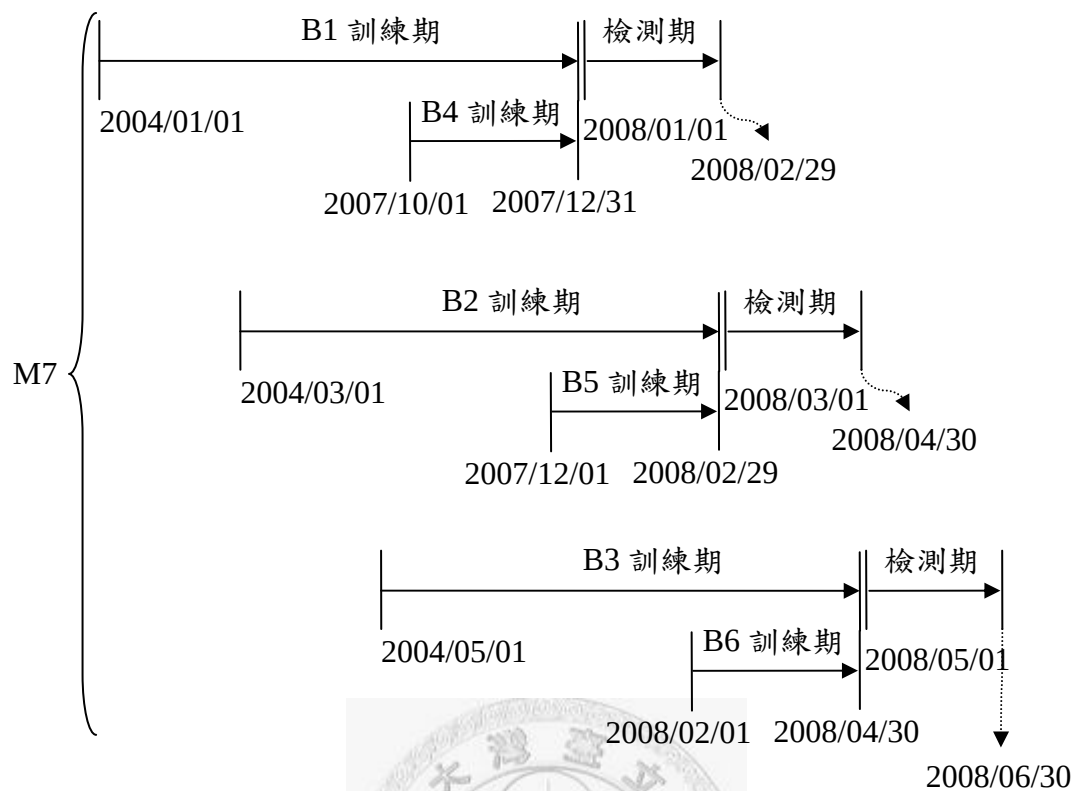
圖十五：M4 訓練與檢測期



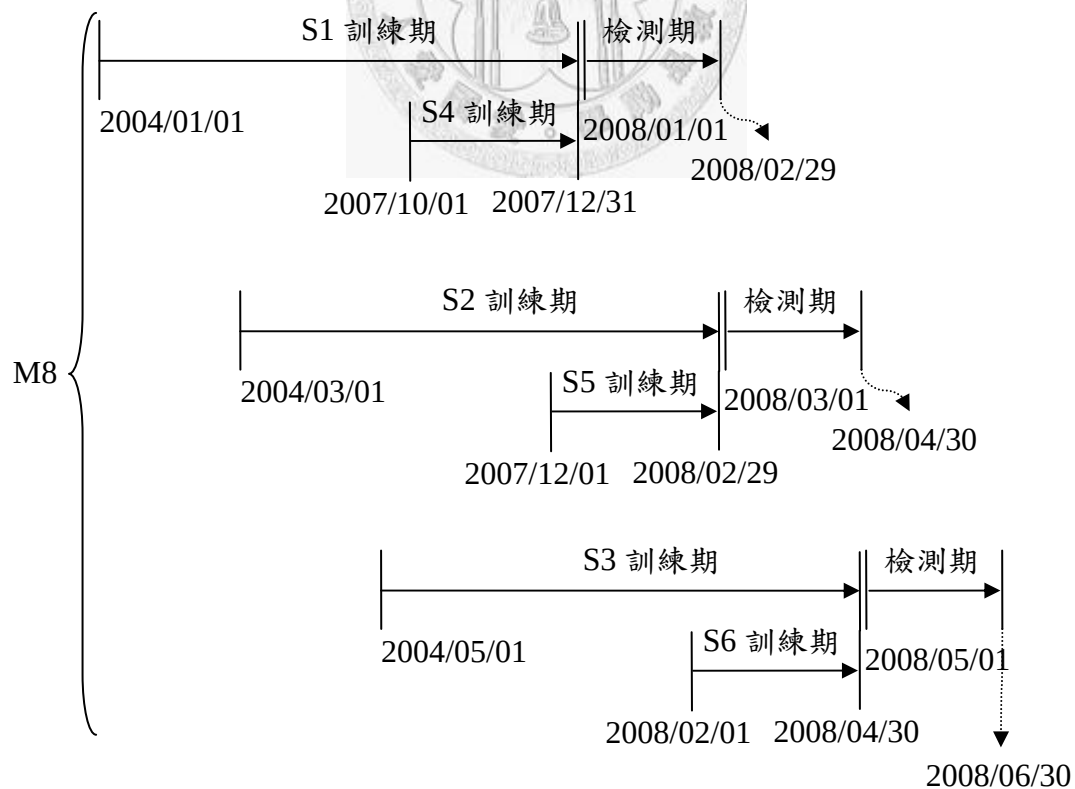
圖十六：M5 訓練與檢測期



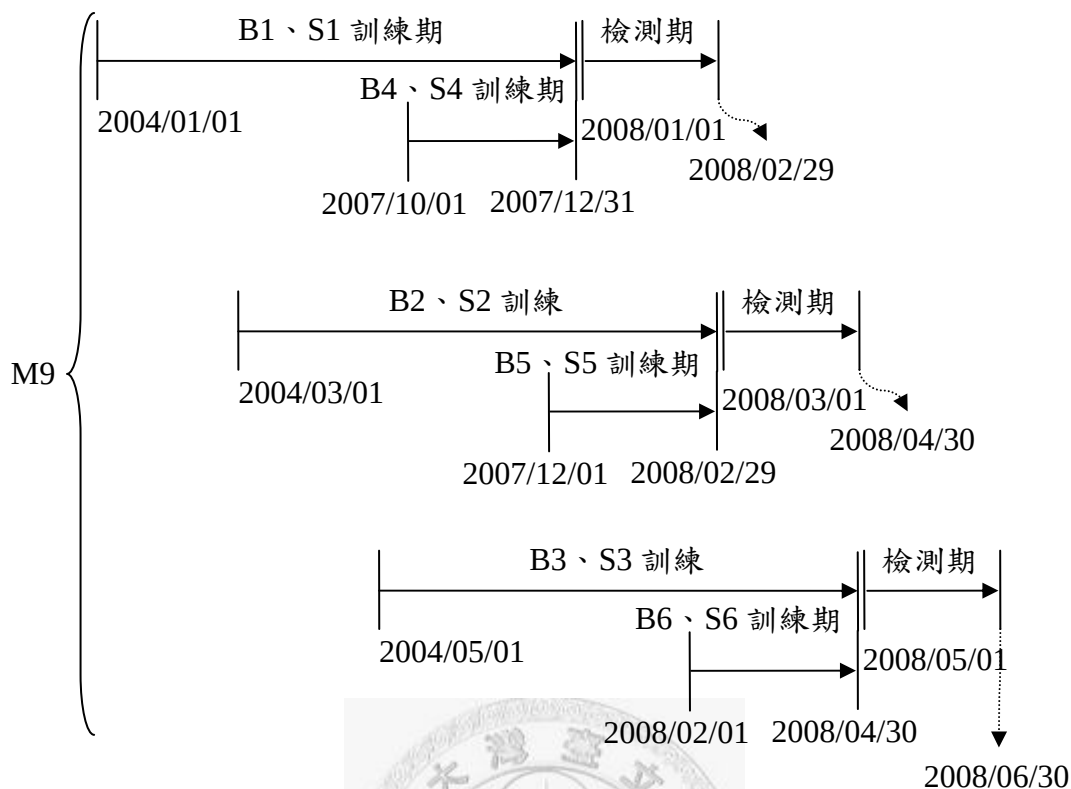
圖十七：M6 訓練與檢測期



圖十八：M7 訓練與檢測期



圖十九：M8 訓練與檢測期



圖二十：M9 訓練與檢測期

分析方式分為兩種：

1. 辨識系統效能分析

檢查分類器的分類結果，在適量的交易次數下，正確率越高表示此方法越佳。

2. 交易系統績效分析

利用分類器產生的買賣訊號模擬交易，搭配一致的買賣策略檢驗各種交易模型的獲利性。

本實驗採用簡易的買賣策略，規則如下：

- (1) 交易開始前，先設定最大部位數量限制，交易期間持有的期貨部位不可超過此限制。
- (2) 買賣訊號由辨識系統產生。買訊產生時，買入一口期貨契約，賣訊產生時，放空一口期貨契約，兩者同時產生則不動作。

- (3) 實務上一個期貨帳戶不允許同時持有相反方向的相同契約，若新的買賣訊號方向與目前持有部位相反時，必須先反向平倉，再依據買賣訊號買入或放空。
- (4) 每筆契約停利 50 點、停損 25 點，到達停損或停利點即平倉。



第四章 實驗結果與分析

4.1 辨識系統訓練結果

我們依照表三定義的訓練期間與漲跌點數，訓練了 12 個分類器，結果如下表五所示：

代號	層數	弱分類器總數	第一層特徵值
B1	10	2099	當下 5 分線 MA10 - 前 1 根 5 分線 LowerBollingerBand
B2	10	2176	當下小時線 MA30 - 當下日線 UpperBollingerBand
B3	10	2266	當下 5 分線 UpperBollingerBand - 前 6 根 5 分線 LowerBollingerBand
B4	7	453	前 6 根小時線 LowerBollingerBand - 當下日線 Open
B5	7	560	前 1 根 5 分線 NL - 前 5 根日線目前 KD 關鍵價
B6	7	437	前 7 根 5 分線 LowerBollingerBand - 前 6 根 15 分線 UpperBollingerBand
S1	10	2245	當下 5 分線 LowerBollingerBand - 前 3 根 15 分線 MA5
S2	10	2292	前 9 根小時線 UpperBollingerBand - 前 2 根日線 LowerBollingerBand
S3	10	2333	前 1 根 5 分線 UpperBollingerBand - 當下 15 分線 MA5
S4	7	463	當下日線 Open - 當下日線 NH
S5	7	709	前 7 根 15 分線 K 值 - 前 7 根日線 MFI(10)
S6	7	441	前 7 根 15 分線 ADX - 前 4 根日線 DX

表五：分類器訓練結果

由於訓練的過程會不斷把被分類器判定為負面的資料丟棄，負面的資料很容易被判定為負面，最後負面資料會全部被丟棄完畢，因此四年的訓練資料只能產生十層的分類器。這是 AdaBoost 處理連續時間資料較為弱勢的一面，我們的解決辦法是在分類器後面再加上近三個月資料訓練成的分類器作為濾網，也就是交易模型 M7、M8、M9 採用的方式。

考量計算速度，每個分類器的第一層只包含一個弱分類器，也是最具有分辨力的分類器。由四年的訓練期間來看，最有影響力的技術指標是保利加通道，時間維度是 5 分線，以當下 K 線的指標最有參考價值。

4.2 辨識系統效能分析

檢驗辨識效能的方式是在不考慮交易成本的前提下，檢查分類器的分類狀況。我們分為兩種情況來考慮，分別為 50 點停利、25 點停損與 50 點停利、50 點停損，當分類器產生買賣訊號後，股價先碰到停利點表示正確分類，反之先來到停損點表示分類錯誤，檢驗結果如下表六。

	50 點停利、25 點停損			50 點停利、50 點停損		
	正確次數	錯誤次數	勝率	正確次數	錯誤次數	勝率
M1	328	267	55.13%	389	206	65.38%
M2	274	367	42.75%	340	301	53.04%
M3	602	634	48.71%	729	507	58.98%
M4	248	200	55.36%	298	150	66.52%
M5	279	313	47.13%	350	242	59.12%
M6	527	513	50.67%	648	392	62.31%
M7	127	96	56.95%	152	71	68.16%
M8	91	85	51.7%	115	61	65.34%
M9	218	181	54.64%	267	132	66.92%

表六：辨識系統效能

可看到動態分類器的效能比固定分類器佳，以 50 點停利 50 點停損為例，勝率增加了 3.33%，顯示判斷股價漲跌的分類器必須動態更新以掌握市場趨

勢。另外，加上濾網後的效果比較好。和不用濾網相比，勝率增加了 4.61%，但正確訊號由 648 次降為 267 次，減少了 58.80%。

4.3 交易系統績效分析

表七是模擬交易使用的參數。根據臺灣期貨交易所最新保證金調整公告(97 年 6 月 19 日)，台股期貨之原始保證金為 87,000 元，我們將模擬交易保證金設為一口 100,000 元，本金為保證金×最大口數。交易成本為來回手續費加交易稅，若臺股指數在 8000 點則交易成本為 600 元，約為本實驗的平均交易成本。

回測參數	設定值
期貨契約	台股期貨
本金	\$100,000
1 點價值	\$200
手續費(單邊)	\$140
期貨交易稅	0.01%
最大口數限制	1 口
可否留倉	不可

表七：模擬交易系統使用參數

交易系統績效評估表數值說明如下表八：

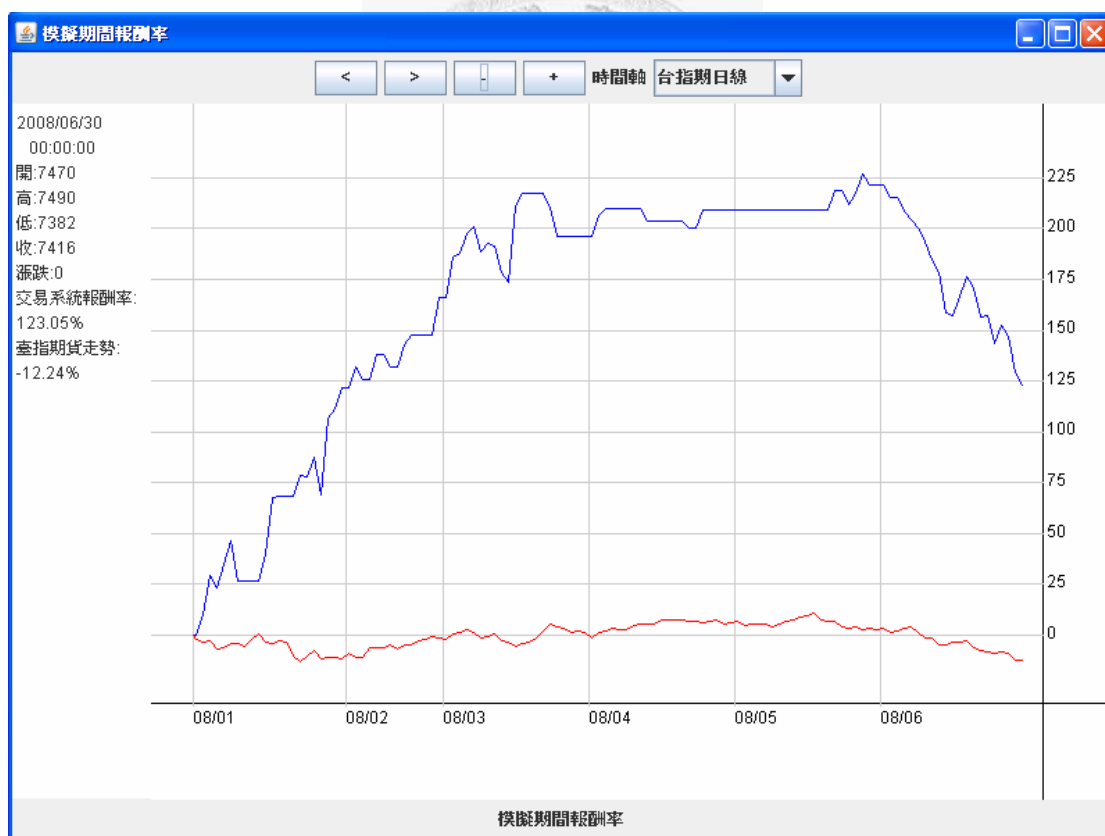
績效評估數值	說明
純益	檢測期間總獲利或虧損金額
報酬率	檢測期間單利報酬率
總收益	所有獲利交易得到的金額
總損失	所有虧損交易失去的金額
勝率	獲利與虧損交易比值
交易回數	總共交易的次數
最大收益	單筆交易最大獲利金額
最大損失	單筆交易最大虧損金額
最大連續收益	連續的交易中最大獲利金額
最大連續損失	連續的交易中最大虧損金額
最大連續收益次數	最多連續獲利幾次
最大連續損失次數	最多連續虧損幾次
最大折返	從績效高點往下滑落到低點的最大距離
單位風險報酬率	報酬率除以最大折返。
平均收益	獲利交易的平均獲利金額
平均損失	虧損交易的平均虧損金額
平均各買賣損益	每筆交易平均獲利或虧損金額
平均交易成本	檢測期間每筆交易平均成本

表八：績效評估表數據說明

下面分別列出每個交易模型的績效評估表、報酬率與大盤績效比較圖。報酬率與大盤績效比較圖的 X 軸為交易日期，範圍由 2008 年 1—6 月，Y 軸為報酬率，範圍由每個交易模型的績效決定。

交易模型 M1 績效評估表			
純益	\$123,050.94	報酬率	123.05%
總收益	\$622,944.81	總損失	(\$499,893.88)
最大折返	(\$103,974.95)	單位風險報酬率	118.35%
交易回數	147	勝率	44.22%
最大收益	\$14,798.74	最大損失	(\$9,975.66)
最大連續收益	\$60,473.20	最大連續損失	(\$32,900.68)
最大連續收益次數	7	最大連續損失次數	5
平均收益	\$9,583.77	平均損失	(\$6,096.27)
平均各買賣損益	\$837.08	平均交易成本	\$601

表九：交易模型 M1 績效評估表



圖二十一：交易模型 M1 報酬率與大盤績效比較圖

交易模型 M2 績效評估表			
純益	(\$163,145.06)	報酬率	-163.15%
總收益	\$597,998.81	總損失	(\$761,143.88)
最大折返	(\$211,610.36)	單位風險報酬率	-77.10%
交易回數	181	勝率	35.91%
最大收益	\$25,415.68	最大損失	(\$15,414.52)
最大連續收益	\$43,977.24	最大連續損失	(\$81,056.78)
最大連續收益次數	4	最大連續損失次數	10
平均收益	\$9,199.98	平均損失	(\$6,561.58)
平均各買賣損益	(\$901.35)	平均交易成本	\$608

表十：交易模型 M2 績效評估表



圖二十二：交易模型 M2 報酬率與大盤績效比較圖

交易模型 M3 績效評估表			
純益	(\$22,738)	報酬率	-22.74%
總收益	\$1,091,963.75	總損失	(\$1,114,701.75)
最大折返	(\$162,615.08)	單位風險報酬率	-13.98%
交易回數	356	勝率	43.54%
最大收益	\$25,415.68	最大損失	(\$15,414.52)
最大連續收益	\$54,100.82	最大連續損失	(\$46,602.20)
最大連續收益次數	5	最大連續損失次數	8
平均收益	\$7,044.93	平均損失	(\$5,545.78)
平均各買賣損益	(\$63.87)	平均交易成本	\$604

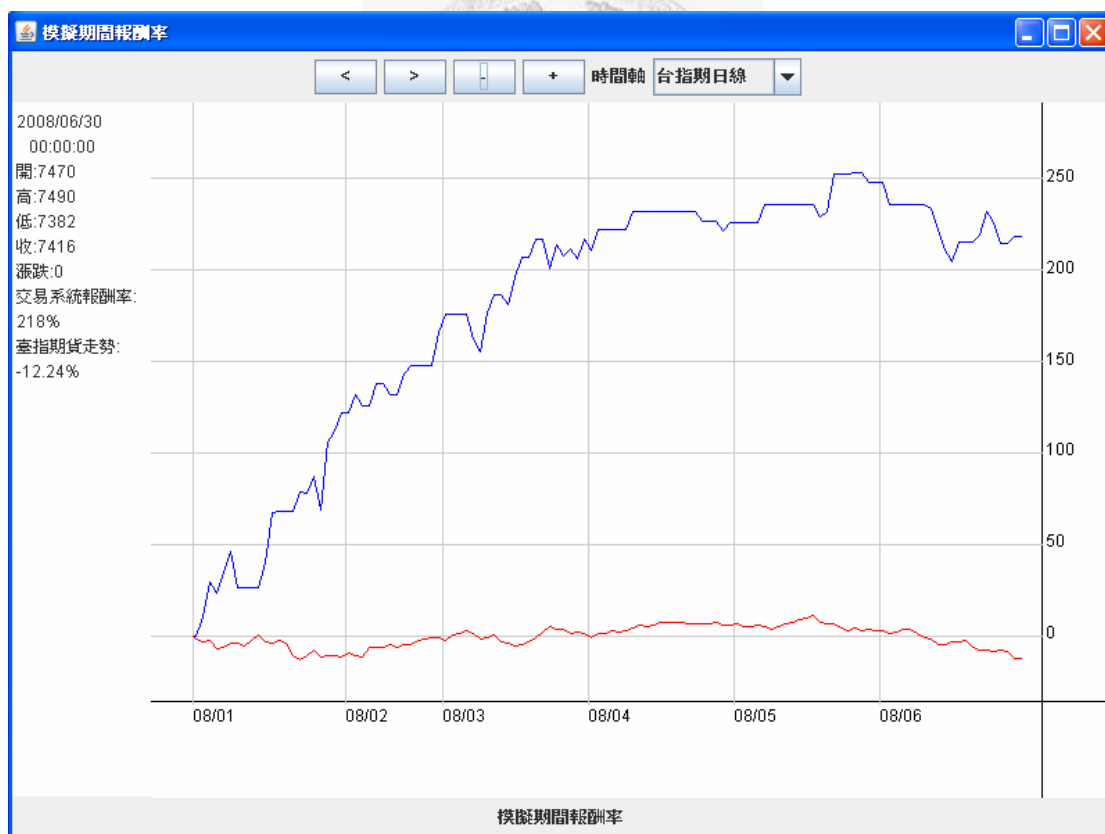
表十一：交易模型 M3 績效評估表



圖二十三：交易模型 M3 報酬率與大盤績效比較圖

交易模型 M4 績效評估表			
純益	\$217,999.47	報酬率	218%
總收益	\$625,675.69	總損失	(\$407,676.22)
最大折返	(\$48,736.32)	單位風險報酬率	447.30%
交易回數	130	勝率	49.23%
最大收益	\$14,798.74	最大損失	(\$10,189.36)
最大連續收益	\$37,280.92	最大連續損失	(\$32,900.68)
最大連續收益次數	4	最大連續損失次數	5
平均收益	\$9,776.18	平均損失	(\$6,176.91)
平均各買賣損益	\$1,676.92	平均交易成本	\$603

表十二：交易模型 M4 績效評估表



圖二十四：交易模型 M4 報酬率與大盤績效比較圖

交易模型 M5 績效評估表			
純益	(\$15,521.94)	報酬率	-15.52%
總收益	\$636,368.25	總損失	(\$651,890.19)
最大折返	(\$122,720.20)	單位風險報酬率	-12.65%
交易回數	161	勝率	40.99%
最大收益	\$25,415.68	最大損失	(\$15,414.52)
最大連續收益	\$55,923.08	最大連續損失	(\$81,056.78)
最大連續收益次數	6	最大連續損失次數	10
平均收益	\$9,641.94	平均損失	(\$6,862.00)
平均各買賣損益	(\$96.41)	平均交易成本	\$606

表十三：交易模型 M5 績效評估表



圖二十五：交易模型 M5 報酬率與大盤績效比較圖

交易模型 M6 績效評估表			
純益	\$271,007.88	報酬率	271.01%
總收益	\$1,186,730.38	總損失	(\$915,722.50)
最大折返	(\$97,918.26)	單位風險報酬率	276.77%
交易回數	315	勝率	51.11%
最大收益	\$25,415.68	最大損失	(\$15,414.52)
最大連續收益	\$55,923.08	最大連續損失	(\$45,662.20)
最大連續收益次數	8	最大連續損失次數	7
平均收益	\$7,371.00	平均損失	(\$5,946.25)
平均各買賣損益	\$860.34	平均交易成本	\$605

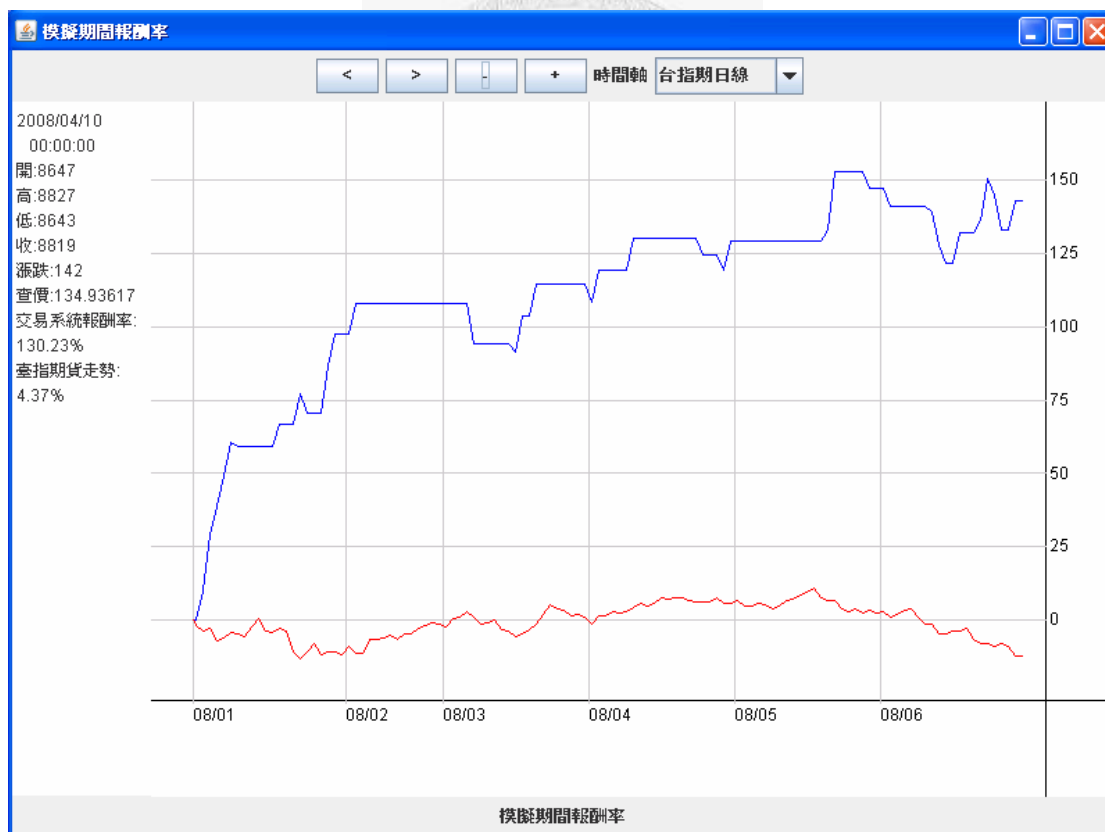
表十四：交易模型 M6 績效評估表



圖二十六：交易模型 M6 報酬率與大盤績效比較圖

交易模型 M7 績效評估表			
純益	\$142,942.41	報酬率	142.94%
總收益	\$338,200.44	總損失	(\$195,258.03)
最大折返	(\$31,302.68)	單位風險報酬率	456.65%
交易回數	66	勝率	53.03%
最大收益	\$14,020.94	最大損失	(\$9,789.32)
最大連續收益	\$60,403.91	最大連續損失	(\$25,320.42)
最大連續收益次數	6	最大連續損失次數	4
平均收益	\$9,662.87	平均損失	(\$6,298.65)
平均各買賣損益	\$2,165.79	平均交易成本	\$600

表十五：交易模型 M7 績效評估表



圖二十七：交易模型 M7 報酬率與大盤績效比較圖

交易模型 M8 績效評估表			
純益	\$13,706.55	報酬率	13.71%
總收益	\$239,713.19	總損失	(\$226,006.64)
最大折返	(\$56,528.36)	單位風險報酬率	24.25%
交易回數	59	勝率	42.37%
最大收益	\$13,021.32	最大損失	(\$12,823.70)
最大連續收益	\$40,539.52	最大連續損失	(\$49,122.60)
最大連續收益次數	5	最大連續損失次數	7
平均收益	\$9,588.53	平均損失	(\$6,647.25)
平均各買賣損益	\$232.31	平均交易成本	\$604

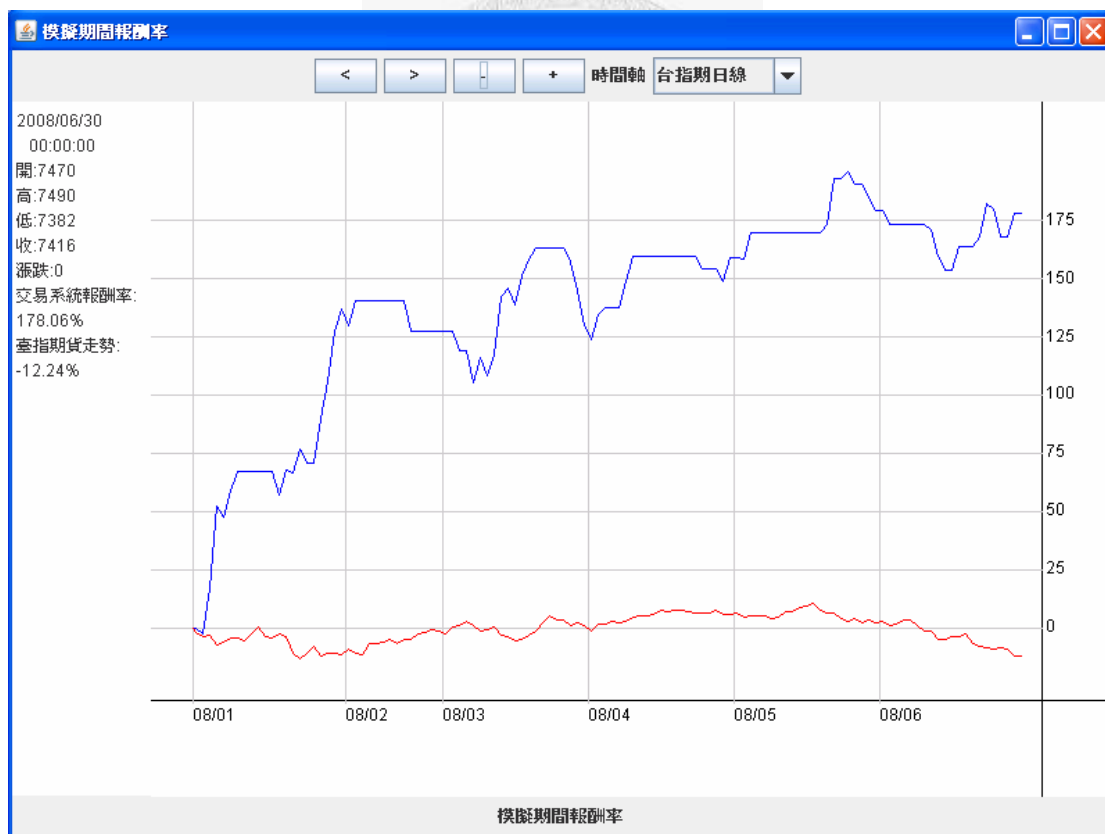
表十六：交易模型 M8 績效評估表



圖二十八：交易模型 M8 報酬率與大盤績效比較圖

交易模型 M9 績效評估表			
純益	\$178,063.00	報酬率	178.06%
總收益	\$582,732.62	總損失	(\$404,669.62)
最大折返	(\$56,528.36)	單位風險報酬率	374.08%
交易回數	125	勝率	50.40%
最大收益	\$14,020.94	最大損失	(\$12,823.70)
最大連續收益	\$69,010.94	最大連續損失	(\$45,510.62)
最大連續收益次數	8	最大連續損失次數	6
平均收益	\$9,249.72	平均損失	(\$6,526.93)
平均各買賣損益	\$1,424.50	平均交易成本	\$602

表十七：交易模型 M9 績效評估表



圖二十九：交易模型 M9 報酬率與大盤績效比較圖

我們把以上九種交易模型的績效取出最重要的四項數據整理如下表十八：

交易模型	報酬率	單位風險報酬率	勝率	平均各買賣損益
M1	123.05%	118.35%	44.22%	\$837.08
M2	-163.15%	-77.10%	35.91%	(\$901.35)
M3	-22.74%	-13.98%	43.54%	(\$63.87)
M4	218%	447.30%	49.23%	\$1,676.92
M5	-15.52%	-12.65%	40.99%	(\$96.41)
M6	271.01%	276.77%	51.11%	\$860.34
M7	142.94%	456.65%	53.03%	\$2,165.79
M8	13.71%	24.25%	42.37%	\$232.31
M9	178.06%	374.08%	50.40%	\$1,424.50

表十八：交易模型主要績效數據比較表

報酬率是衡量交易模型獲利能力的指標，至少須為正值，否則處於長期虧損。單位風險報酬率是純益除以最大折返，用來衡量風險與報酬的比值，至少須為正值，否則沒有意義。注意的是我們不使用夏普指數(Sharpe index)作為衡量單位風險所獲得的報酬，我們認為最大折返比標準差更適合描述本交易模型的風險值。勝率的高低對於判斷交易系統的好壞並非決定性的因素，只要大賺小賠，即使勝率低也一樣可以獲利。但本實驗的所有模型採用相同的停損停利點數，因此勝率的高低直接決定了交易模型的績效。平均買賣損益是每次交易的期望值，可以看出平均每筆交易扣除手續費後還有多少獲利。

由以上九個交易模型的績效可以觀察到幾個現象：

1. 採用動態分類器且多空都作可以獲得最佳報酬率

報酬率前三名分別是 M6、M4、M9，由於 M6 交易次數大約是 M4、M9 的 2.5 倍，因此雖然 M6 的平均買賣損益不高，但乘上交易次數得到的純益卻是最高的。

2. 採用動態分類器加濾網只作多交易模型每筆交易平均獲利最高且風險最低

單位風險報酬率與平均各買賣損益前三名皆為 M7、M4、M9。我們認為多空都作可以降低風險，因此預期單位風險報酬率最高的應該是 M9，但由於本實驗結果作空的交易系統(M8)績效不佳，只好屈居第三。下面第 5 點會探討為何在交易系統績效實驗中，作空的績效相對於辨識系統效能實驗差

很多，以及如何改善此問題。

3. 採用固定分類器的交易模型在檢測一段時間後績效明顯下降

由 M1、M2、M3 的報酬率走勢(圖二十一、圖二十二、圖二十三)，可以明顯看出四、五、六月的報酬率向下，與第三章的假設相符，由於股價趨勢會隨著時間而改變，因此必須動態更新分類器。

4. 採用濾網可以提高平均各買賣損益

M4、M5、M6 加上近三個月訓練資料的分類器作為濾網，分別成為 M7、M8、M9，平均各買賣損益提高的金額如下表十九：

交易模型	平均各買賣損益	交易模型	平均各買賣損益	增加
M4	\$1,676.92	M7	\$2,165.79	\$488.87
M5	(\$96.41)	M8	\$232.31	\$328.72
M6	\$860.34	M9	\$1,424.50	\$564.16

表十九：使用濾網前後提高平均各買賣損益金額

可以看出採用濾網後提高了每筆交易的預期報酬，但同時也大量降低交易次數，因此這之間必須做取捨，在保持一定的交易次數下，盡可能提高平均各買賣損益。

5. 作多的模型績效比作空的模型佳

由上一節表六顯示作多分類器效能比作空分類器佳，因此已經可以預估出作多的交易系統績效比作空的佳，實際的結果也是如此，且作空交易系統績效比預期還差。主要是因為本實驗的最大口數為 1 口，當已經持有期貨部位，又出現相同方向的買賣訊號時，依照第三章訂定的交易策略，這個訊號必須忽略。假設辨識系統產生的訊號有 300 個，限制 1 口的情況下，交易次數可能只有 100 次，而本實驗的作空交易系統挑出的這些交易中，較大的比例是屬於虧損交易。

為了證明我們的假設，我們以 M8 為例，把最大口數限制改為 5 口，重新檢測其績效，結果如下表二十：

交易模型 M8-5 口績效評估表			
純益	\$326,304.75	報酬率	65.26%
總收益	\$811,946.50	總損失	(\$485,641.75)
最大折返	(\$94,599.44)	單位風險報酬率	344.93%
交易回數	151	勝率	53.64%
最大收益	\$15,805.12	最大損失	(\$13,006.80)
最大連續收益	\$170,016.23	最大連續損失	(\$94,599.44)
最大連續收益次數	16	最大連續損失次數	15
平均收益	\$10,024.03	平均損失	(\$6,937.74)
平均各買賣損益	\$2,160.96	平均交易成本	\$603

表二十：交易模型 M8-5 口績效評估表

把 M8-1 口與 M8-5 口的績效比較如下：

交易模型	報酬率	單位風險報酬率	勝率	平均各買賣損益
M8-1 口	13.71%	24.25%	42.37%	\$232.31
M8-5 口	65.26%	344.93%	53.64%	\$2,160.96

表二十一：M8-1 口與 M8-5 口績效比較

由於 M8-5 口的交易回數為 151 次接近訊號產生次數 176 次，因此勝率會接近辨識系統的勝率 51.7%(50 點停利、25 點停損)，與 M8-1 口比較，整體績效大幅提高。

6. 交易系統部位限制 1 口會降低勝率

由上面第 5 點可知，只用 1 口的交易次數比訊號產生次數少很多，因此造成勝率的偏差，且一段走勢中，第一個訊號較容易是錯誤的判斷，因此只用 1 口的策略其勝率往往比辨識系統差。我們把最大口數限制改為 5 口，重新檢測每個交易模型的勝率，並與限制 1 口和辨識系統的勝率比較，結果如下表二十二：

	限制 1 口勝率	限制 5 口勝率	辨識系統勝率
M1	44.22%	47.88%	55.13%
M2	35.91%	39.92%	42.75%
M3	43.54%	45.5%	48.71%
M4	49.23%	51.78%	55.36%
M5	40.99%	48.46%	47.13%
M6	51.11%	55.14%	50.67%
M7	53.03%	52.32%	56.95%
M8	42.37%	53.64%	51.7%
M9	50.40%	53.67%	54.64%

表二十二：限制 1 口、5 口與辨識系統勝率比較表

辨識系統不含交易成本，理論上勝率會比交易系統稍高一些。可看到除了 M7 之外，其餘的交易模型把口數限制提升為 5 口後，勝率都有提高，尤其以 M5、M8 這兩個作空交易模型提升效果最為顯著。

7. 部位限制改為 5 口情況下，採用動態分類器加濾網多空都作的單位風險報酬率最高。

我們把最大口數限制改為 5 口，重新檢測每個交易模型的績效，主要數據如下表二十三所示：

交易模型	報酬率	單位風險報酬率	勝率	平均各買賣損益
M1	85.79%	196.95%	47.88%	\$1,134.82
M2	-50.31%	-57.06%	39.92%	(\$534.11)
M3	39.47%	55.98%	45.50%	\$227.87
M4	120.87%	348.45%	51.78%	\$1,955.87
M5	84.43%	111.95%	48.46%	\$1,002.72
M6	240.84%	407.54%	55.14%	\$1,627.32
M7	68.44%	334.35%	52.32%	\$2,266.38
M8	65.26%	344.93%	53.64%	\$2,160.96
M9	127.02%	611.83%	53.67%	\$2,117.07

表二十三：交易模型主要績效數據比較表(5 口)

可以看到 M9 的單位風險報酬率最高，平均各買賣損益前三名則是 M7、M8、M9，三者差距不大。因此若要兼顧低風險與高報酬，可以採用 M9 交易模型搭配最大部位 5 口的交易策略。



第五章 總結與展望

5.1 結論

本論文提出以 AdaBoost 為辨識方法，搭配停損停利操作的台股期貨當沖交易系統。我們根據四個議題：是否需要定期更新分類器、濾網對效能的影響、只作單邊或多空雙向的績效差異、最大口數限制的影響，設計了九個交易模型，分別以辨識系統分類效能與交易系統回測績效兩種方式檢驗，歸納出以下五點結論：

1. 定期更新分類器可以讓辨識系統分類規則隨著市場趨勢修正，有助於維持獲利能力。
2. 使用濾網可以增加交易勝率，但會減少交易次數。
3. 多空都作是兼具低風險與高獲利性的交易方式。
4. 提高最大口數限制可以縮短交易系統與辨識系統勝率的差距。
5. 本研究提出一個低風險且穩定報酬的交易系統，提供期貨交易人參考。

5.2 未來展望

本研究著重於 AdaBoost 對於台股期貨趨勢的辨識性，實際上一個完整的程式交易系統還包括許多元件，例如定期訓練最新的分類器、風險控管、系統狀態監控、交易績效回報等，這些工作留待將來完成。

以下提出七點建議，提供給未來在這方面的研究者參考：

1. 市場上的技術指標種類很多，除了本研究採用的 11 種指標之外，可以嘗試加入不同的指標。另外，可以用不同方式取出特徵值，或是先把指標標準化後再取得特徵，或許有更好的效果。
2. 本研究採用 50 點停利、25 點停損的當沖交易策略，遇到大行情時獲利比波段系統低。我們建議是可以增加容忍度較大的分類器，當判斷為波段行情時動態調整停損停利點數，並適度考慮留倉以獲取波段報酬。
3. 本實驗的系統使用 4 年資料訓練的分類器，加上 3 個月資料訓練的分類器作為濾網。要以多大的資料訓練並無一定的規範，理論上訓練資料越多，越能得到較佳的分類效果，但另一方面來說，久遠的資料意味著與檢測資訊有較大落差。另外，辨識系統不見得一定要主分類器加濾網這

個架構，也可以訓練三個相同的分類器，投票表決多數同意。

4. 修改 AdaBoost 演算法，在訓練分類器時，依照資料的漲跌幅不同給予不同的權重，讓趨勢明顯的資料先被正確分類。或是採用 Boosting 家族其他的演算法，或許可以得到更佳的檢測資料分類效果。
5. 修改分類器的輸出結果，由原本的正面或反面，改為正面的機率，讓交易系統依據漲跌的機率分配資金的運用。
6. 如果最大口數限制不只 1 口，可以搭配順勢加碼、逆勢減碼的資金規劃，有系統的利用資金。
7. 本論文的研究標的為臺股期貨，未來的研究者可檢驗本交易系統於其他期貨商品是否具有獲利性。

立法院於 2008 年 7 月 15 日通過修正期貨交易稅條例，將股價類期貨契約法定稅率下限由現行的千分之 0.1 降為百萬分之 0.125，相信未來核定稅率可望由目前的萬分之 1 調降為 10 萬分之 1。若能降稅，將有助於臺灣期貨市場的成長，且對當沖交易者來說是一項利多，可望大幅改善被交易成本侵蝕獲利的情況。



參考文獻

- [1] Creamer, German G. and Freund, Yoav, “A Boosting Approach for Automated Trading” In *Proceedings of the Data Mining for Business Applications Workshop on International Conference on Knowledge Discovery and Data Mining (KDD 2006)*, Philadelphia, August 20–23, 2006.
- [2] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. (Second Edition), Wiley-Interscience, New York, 2001.
- [3] Robert D. Edwards and John Magee. *Technical Analysis of Stock Trends*, AMACOM, New York, 1966.
- [4] Yoav Freund and Robert E. Schapire. “A decision-theoretic generalization of on-line learning and an application to boosting.” *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.
- [5] Irwin Scott H. and Park Cheol-Ho. “What Do We Know About the Profitability of Technical Analysis?” *Journal of Economic Surveys*, 21(4):786–826, 2007.
- [6] Michael Kearns. Thoughts on hypothesis boosting. Unpublished manuscript. 1988.
- [7] Paul Viola and Michael J. Jones. “Robust Real-Time Face Detection,” *International Journal of Computer Vision*, 57(2):137–154, May 2004.
- [8] 吳典林, “基於模糊聚類之台股盤整區間期貨交易系統”, 國立臺灣科技大學資訊工程研究所碩士論文, 2005.
- [9] 周俊志, “自動交易系統與策略評價之研究”, 國立臺灣大學資訊工程研究所碩士論文, 2007.
- [10] 林俊吉, “應用灰聚類進行台股指數期貨程式交易設計”, 國立臺灣科技大學資訊工程研究所碩士論文, 2005.
- [11] 洪育安, “台灣股價指數期貨波動度預測模型之建立與績效比較：兼論納入成交量、到期期間與 VIX 資訊內涵”, 國立台北大學合作經濟研究所碩士論文, 2007.
- [12] 曾永政, “期貨操作不靠內線：程式交易打造富足生活”, 2007.
- [13] 蔡明欽, “台股指數期貨之預測與交易策略之研究”, 輔仁大學管理學研究所碩士論文, 2005.