

國立臺灣大學電機資訊學院資訊工程學系

碩士論文

Department of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

使用增強式學習法建立臺灣股價指數期貨當沖交易策

略

Using Reinforcement Learning to Establish Taiwan

Stock Index Future Intra-day Trading Strategies

賴怡玲

Yi-Ling Lai

指導教授：呂育道 博士

Advisor: Yuh-Dauh Lyuu, Ph.D.

中華民國 98 年 7 月

July 2009

國立臺灣大學碩士學位論文

口試委員會審定書

使用增強式學習法建立臺灣股價指數期貨當沖交易策略

Using Reinforcement Learning to Establish Taiwan
Stock Index Future Intra-day Trading Strategies

本論文係賴怡玲 君 (R96922117) 在國立臺灣大學資訊工程學所完成之碩士學位論文，於民國 98 年 6 月 25 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：

(簽名)

(指導教授)

系主任、所長

(簽名)

誌謝

經過兩年的學習，論文終於順利完成，首先要感謝的是指導教授呂育道博士的諄諄教誨，帶領我們進入財務工程與金融計算的領域，老師的督促與指導使得本篇論文得以圓滿完成。

接下來要感謝我的家人，你們的支持是我完成論文最大的動力，還有好友君玉，謝謝你在生活上給我的諸多照顧，讓我得以專心完成課業。

此外還要感謝實驗室的各位學長姊、同學及學弟妹，彼此之間的討論和資訊的互相分享，讓我兩年的求學生活順利、愉快。

最後要感謝精誠資訊的主管易建雄先生及元富證券的耿世鈞先生，在我的生涯規劃上提供許多寶貴的建議。



中文摘要

增強式學習法具有與環境互動及延遲報酬兩大特色，適合應用在決策控制系統的問題上，因此本研究採用增強式學習法來建立臺灣股價指數期貨的當沖交易策略。在系統設計上，我們嘗試了三種不同的狀態定義方式、採用 Q-learning 及 SARSA 兩種不同的演算法，另外也針對停損、停利點的設置進行討論。

為檢測其可用性，我們採用 2004 年 1 月 1 日至 2008 年 6 月 30 日之臺灣股價指數期貨歷史資料進行學習訓練及績效檢測。

中文關鍵詞：交易策略、當日沖銷、臺股期貨、機器學習、增強式學習法、SARSA、Q-learning



Abstract

Learning from interacting with environment and delayed reward are the two most important features of reinforcement learning. Because of these two characteristics, reinforcement learning is suitable for control problems. This thesis adopts reinforcement learning to establish several Taiwan stock index future intra-day trading strategies. We design three different definitions of state and use Q-learning and SARSA to implement reinforcement learning. In addition, we discuss the effect of setting maximum acceptable loss and minimum acceptable profit.

To verify the usability of our strategies, we use real historical data for back testing and then examine the performance of the trading strategies.

Keywords: Trading strategy、Intra-day trading、Taiwan stock index future、Machine learning、Reinforcement Learning、SARSA、Q-learning



目錄

口試委員會審定書.....	i
誌謝.....	ii
中文摘要.....	iii
Abstract.....	iv
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 研究目的.....	1
1.3 論文架構.....	2
第二章 背景知識.....	3
2.1 行為財務學.....	3
2.2 交易策略.....	4
2.3 技術分析.....	4
2.3.1 K線.....	5
2.3.2 移動平均線.....	5
2.3.3 平均交易量.....	6
2.4 增強式學習法.....	6
2.4.1 時間差分法.....	7
2.4.2 SARSA.....	8
2.4.3 Q-learning.....	9
第三章 研究方法.....	10
3.1 系統設計.....	10
3.1.1 狀態.....	10
3.1.2 動作.....	10
3.1.3 報酬.....	11
3.1.4 價值函數.....	11
3.2 實驗設計.....	11
3.2.1 實驗資料.....	11
3.2.2 交易環境設定.....	11
3.2.3 實驗模型設計.....	12
第四章 實驗結果與分析.....	13
4.1 各模型績效一覽表.....	13
4.2 各模型績效比較分析.....	18
第五章 總結與展望.....	22
5.1 結論.....	22
5.2 未來展望.....	22



圖目錄

圖 一：K線圖.....	5
圖 二：增強式學習法中學習代理人與環境互動圖.....	7
圖 三：SARSA演算法.....	8
圖 四：Q-learning演算法.....	9
圖 五：總損益比較圖.....	18
圖 六：獲利率比較圖.....	18
圖 七：狀態A之每日損益標準差.....	19
圖 八：狀態K之每日損益標準差.....	19
圖 九：狀態C之每日損益標準差.....	19



表目錄

表 一：各交易模型一覽表.....	12
表 二：績效評估項目說明表.....	13
表 三：交易模型SAQL00P00 績效表.....	14
表 四：交易模型SAQL50P50 績效表.....	14
表 五：交易模型SASL00P00 績效表.....	14
表 六：交易模型SASL50P50 績效表.....	15
表 七：交易模型SKQL00P00 績效表.....	15
表 八：交易模型SKQL50P50 績效表.....	15
表 九：交易模型SKSL00P00 績效表.....	16
表 十：交易模型SKSL50P50 績效表.....	16
表 十一：交易模型SCQL00P00 績效表.....	16
表 十二：交易模型SCQL50P50 績效表.....	17
表 十三：交易模型SCSL00P00 績效表.....	17
表 十四：交易模型SCSL50P50 績效表.....	17
表 十五：SCQL00P00 與SCQL50P50 平均獲利比較表.....	20
表 十六：2008/3/27 交易動作比較表.....	20
表 十七：2008/5/23 交易動作比較表.....	21

第一章 緒論

1.1 研究動機

投資者在進行投資決策時都以為自己的決策過程是嚴謹且符合理性的，但研究結果發現投資者在進行投資決策時，常會受到一些心理偏誤的影響。為了要克服這些心理偏誤並建立投資紀律，投資者依據自己的投資目標發展出各種交易策略，投資者藉由遵守這些策略以避免受到本身心理因素的影響。在電腦科學的領域中，人工智慧與機器學習已被廣泛的應用在交易策略的發展上，其中大部分的研究都是著重於價格變化的預測，交易系統則根據此預測結果進行投資。本研究則從另外一個角度出發，我們不對價格的變動進行預測，而是採用增強式學習法(reinforcement learning, RL)直接學習如何進行交易決策。

增強式學習法是機器學習法中的一種，起源於心理學中的制約操作理論(operant conditioning)，它不同於監督式學習法(supervised learning)必須要有一個監督者來指導學習，增強式學習法的學習代理人沒有明確的指導者，僅憑著與環境的互動進行學習。而在一連串的決策過程中，單一項決策可能會影響後續的狀態與決策，因此增強式學習法學習的目標是極大化所有決策報酬的總合，而非極大化單一決策的報酬。與環境互動及追求未來整體預期收益的特色，讓增強式學習法被認為適合用在決策控制系統上，因此本研究採用此方法學習交易決策以建立交易策略。

1.2 研究目的

在本研究中，我們利用增強式學習法與環境互動及延遲報酬的兩大特色建立一個學習代理人，讓學習代理人反覆地進行投資決策，系統則根據價格歷史資料給予此決策相對應的獲利或損失，學習代理人藉由此回饋進行決策的修正，最終獲得一個當沖交易策略。

為了檢驗其可用性，本研究採用臺灣股價指數期貨歷史資料進行驗證，將此歷史資料分為學習期及績效檢測期，將學習代理人在學習期所學到的交易策略應用到績效檢測期，將此期間的投資績效做為各種系統設計的評比標準。

1.3 論文架構

第一章 緒論

說明研究動機與目的。

第二章 背景知識

描述財務行為學、交易策略、技術分析、增強式學習法。

第三章 研究方法

說明系統的設計方式及實驗方法。

第四章 實驗結果與分析

針對前一章所介紹之研究方法的實驗結果進行分析與討論。

第五章 總結與展望

討論本實驗的結果與未來可改進的方向。



第二章 背景知識

2.1 行為財務學

投資行為是由一連串的投資決策所組成，然而當人們在進行決策時，並不如自己所認為的那樣理性，行為財務學(behavioral finance)正是在探討投資者的心理如何影響其決策的過程。研究結果發現投資者在進行投資決策時，常會受到一些心理偏誤的影響，例如當人們所擁有的知識增加時，人們會傾向於相信自己的預測能力也隨之增加，人們也會自以為對一些不可控制的事件具有影響力，這造成了人們在決策時會發生過度自信(overconfidence)的心理偏誤，而這樣的偏誤會使人高估自己所擁有的知識及低估所面臨的風險。過度自信的投資者會相信自己擁有具高度精確性的資訊，也會過於相信自己對於資訊的判斷能力，因此容易造成過度交易的情況發生，而當投資者對自己深具信心時，也會影響投資者對自己風險承受度的判斷能力，進一步做出不良的投資決策[5]。

另一種常見的心理偏誤是人們在判斷自己的風險承受度時，會受到過去結果的影響，人們在獲利之後會提高自己的風險承受度，此時過度自信也變得較為嚴重；而在損失之後，有時會選擇接受較大的風險，企圖打平損益，有時則會開始不願意接受風險。對於投資者來說，在獲利之後自信心增加，不但會因為過度自信而造成過度交易，也會做出風險較高的投資決策。在投資損失之後，則會選擇將剩餘的資金投入較高風險的投資中，企圖彌補上一次的損失；或是開始趨避風險，甚至因為害怕而不願意回到市場。

除了以上說明的兩種心理偏誤外，還有許多會影響投資決策品質的心理偏誤存在，要克服這些心理偏誤，投資者可透過訂定具體且明確的投資目標及制定符合目標的交易策略，在執行交易時以遵守策略為原則，僅根據市場的即時狀況進行微調。在每一次的投資結束後，也必須針對投資的結果進行回饋檢討與策略調整，如此一來才能不受心理偏誤的影響而做出正確的投資決策。

2.2 交易策略

交易策略是藉由分析歷史交易資料，以技術面、基本面、籌碼面等面向為基礎所建立出來的交易系統。建立的方法可分為兩種[3]：

1. 模式趨導(model-driven)：尋找現有的交易策略模型，了解其原理後對模型現有的規則進行個別或排列組合的分析，藉由反覆地進行回溯測試調整其組合方式及模型參數，進一步超越原本模型的績效。
2. 資料趨導(data-driven)：不預設任何立場，僅憑現有的資料歸納有效的規則以建立操作模型。此種建立方法必須對大量的資料進行分析，可採用資料探勘、人工智慧、機器學習等電腦科學發展出來的技術做為分析的工具與方法。本研究即是以資料趨導方式建立交易策略，並且在分析歷史資料時採用增強式學習法。

無論以何種方式建立交易策略，為確定此交易系統的可行性，均須進行歷史資料的回溯測試，藉由測試的績效來調整策略的參數以達到策略最佳化的結果。

2.3 技術分析

投資者在建立交易策略時，必須先取得歷史交易資料以進行分析，而這樣的資料必須具有「可取得」及「可定量分析」的特性，在此限制之下，交易策略大多以技術分析為基礎。

技術分析認為市場價值是由供給與需求決定的，而市場上每個投資者對未來的預期都已反映在價格和交易量上，於是透過圖表或量化的指標去解析市場過去的價格與交易量，可做為預測未來變化的基礎。技術分析中的技術指標即為將價格與交易量以合理的方式進行數學處理後的結果，若依據使用的資料類別可分成以下幾類：

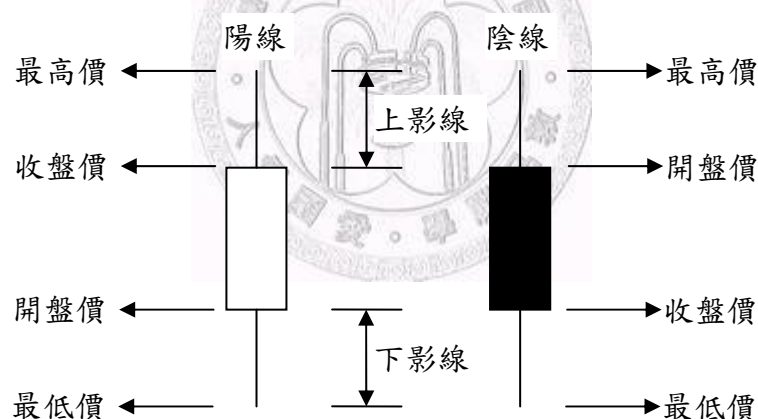
1. 價指標：單純以成交的開高收低價處理的指標，例如移動平均線(MA)、相對強弱指標(RSI)等。
2. 量指標：單純以成交量、值或成交筆數處理的指標，例如交易張數、平均交易量、能量潮(OBV)等。

3. 量價指標：同時根據量價關係所發展出來的指標，例如每一加權指數股票成交值(TAPI)等。
4. 綜合指標：綜合不同觀念發展而成的指標，例如隨機指標(KD)、趨向指標(DMI)等。
5. 理論指標：根據特定理論而發展的指標，例如黃金分割率、箱型理論、壓力支撐理論等。

以下分別介紹本研究所採用的三種技術指標。

2.3.1 K線

K線是由一段時間內的開盤價、收盤價、最高價以及最低價等四種價位組成。若收盤價比開盤價高，稱為陽線；若收盤價比開盤價低，稱為陰線，如圖一。而依其代表的時間長短，又可分為日線、小時線、15分鐘線、5分鐘線及1分鐘線等，不同時間長短的K線描述了不同維度的行情狀況。



圖一：K線圖

K線的組成雖然簡單，卻是所有技術指標的基礎，它能反應出市場多空雙方力量的消長變化，而不同型態的K線圖也分別具有不同的市場訊號。

2.3.2 移動平均線

移動平均線(Moving Average, MA)是一段時間的價格平均，其意義在於表示目前股價的趨勢方向及這段時間內投資人的平均買入成本。其公式如下：

$$MA(N) = \frac{1}{N} \sum_{i=0}^{N-1} \text{Close}_i$$

其中 N 為移動平均數的計算天數， Close_i 為前 i 天的收盤價。

2.3.3 平均交易量

平均交易量為某一交易期間分析對象之平均交易量，此為市場流動資金多寡之合計值，可做為股市人氣指標及股價先行指標，通常會搭配股價走勢進行分析。其公式如下：

$$MVM(N) = \frac{1}{N} \sum_{i=0}^{N-1} V_i$$

其中 N 為平均交易量的計算天數， V_i 為前 i 天的成交量。

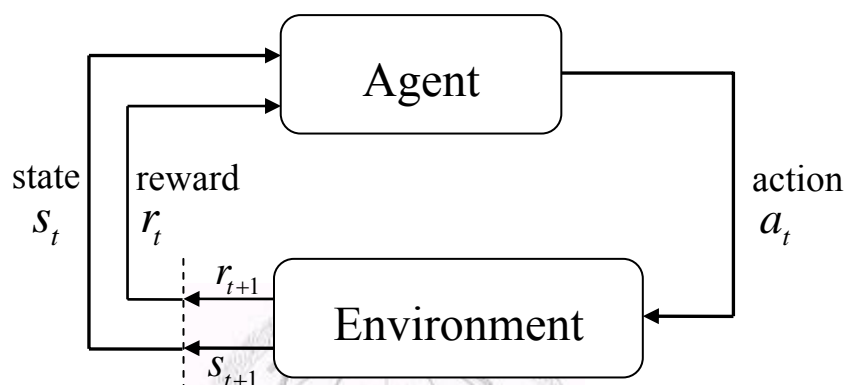
2.4 增強式學習法

增強式學習法會建立一個學習代理人(agent)，此學習代理人會根據現在所處的狀態(state)採取動作(action)，在一開始沒有任何知識基礎下，代理人可任意選擇一項動作。環境(environment)接收到此動作後，會根據此動作回饋給代理人一個報酬(reward)，此報酬可能是正面的獎勵或是負面的懲罰[8]。

學習代理人接收到報酬後，會以極大化未來整體預期收益(expected return)為目標，建立與調整此項狀態與動作的價值函數(value function)，價值函數所代表的意義是在此狀態下採取這個動作後所能獲得的預期收益，意即這個狀態或是這個動作有多好，價值函數的起始值通常設為零，也可根據基礎知識設置其它的數值。在實做增強式學習法的演算法時，要如何正確的估計價值函數是很重要的。若將價值低的狀態或動作錯估為高價值，則將影響其學習的績效。往後的決策若有遇到相同的狀態或動作便可根據這些價值函數來選擇動作，其選擇的方式為貪婪法(greedy method)，但若完全使用貪婪法，將會喪失其他動作被選擇的機會，因此通常會使用改良過後的貪婪法(ϵ -greedy method)，即為在每一次選擇時，給予一個極小的機率 ϵ 去進行探險，也就是放棄最好的選擇，嘗試其它新的

選項，此探險的機率愈高，學習代理人就愈有可能嘗試新的狀態或動作，其學習效果會比較好，但其缺點是收斂的時間將隨探險的比例增加。

環境在提供學習代理人報酬的同時，也會提供下一個狀態給代理人，代理人根據環境所提供的新狀態選擇一個相對應的動作。藉由代理人與環境如此反覆地互動之後，代理人可以習得一個狀態與動作的對應，在增強式學習法中稱此對應為策略(policy)，之後遇到同樣的狀態時，系統便可根據此策略進行決策。其架構如圖二。



圖二：增強式學習法中學習代理人與環境互動圖

2.4.1 時間差分法

增強式學習法的演算法可分為動態規畫(dynamic programming, DP)、蒙地卡羅法(Monte Carlo methods)、時間差分法(temporal-difference, TD)三類[8]。動態規畫的學習效果最好，在給定環境模型之後，動態規畫可求得一最佳解，但其缺點即為必須要知道完整且精確的環境模型，然而大部分的問題所面臨的環境都很複雜，不容易獲得環境完整的資訊。蒙地卡羅法採取抽樣的方式，因此不需要知道環境的模型，但學習的結果會較不正確。而時間差分法介於兩者之間，就像蒙地卡羅法一樣，時間差分法一樣也採取抽樣的方式，因此也不需要知道完整的環境模型；時間差分法也跟動態規畫一樣屬於拔靴法(bootstrapping method)，在更新價值函數時都是基於其它已學習過價值函數值。

時間差分法計算價值函數的方法為：

$$V(s_t) \leftarrow V(s_t) + \alpha [r_{t+1} + \gamma V(s_{t+1}) - V(s_t)]$$

其中， $V(s_t)$ 為狀態 s_t 的價值函數，也就是到達狀態 s_t 後可獲得的預期收益， r_{t+1} 為在狀態 s_t 下選擇該動作，到達一個新的狀態 s_{t+1} 後獲得的報酬， α 稱為 step-size，主要是用來控制學習的速率， γ 則為折扣因子。

由此算式可看出增強式學習法會將每次選擇動作之後所獲得之報酬用來調整其價值函數，除此之外，也會加上下一個狀態的預期收益來調整，此調整方式是以極大化未來整體預期收益為目標。

2.4.2 SARSA

SARSA 是一種 on-policy 時間差分控制方法，on-policy 是表示它是從現在正在執行的策略中進行學習。而與一般時間差分法不同的是它是以計算每個狀態-動作配對的價值函數取代計算每個狀態的價值函數，主要是應用在控制問題上，其價值函數的計算方式為：

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$$

其中 $Q(s_t, a_t)$ 為狀態-動作配對之價值函數，也就是在狀態 s_t 時選擇的動作 a_t 可獲得的預期收益， r_{t+1} 為選擇該動作後獲得的報酬， α 稱為 step-size，主要是用來控制學習的速率， γ 則為折扣因子。其演算法如圖三。

```

Initialize  $Q(s, a)$  arbitrarily
Repeat (for each episode)
  Initialize  $s$ 
  Choose  $a$  based on  $s$  using policy derived from  $Q$  (e.g.,  $\epsilon$ -greedy method)
  Repeat(for each step of episode):
    Take action  $a$ , obtain reward  $r$  and next state  $s'$  from the environment
    Choose  $a'$  based on  $s'$  using policy derived from  $Q$  (e.g.,  $\epsilon$ -greedy method)
     $Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma Q(s', a') - Q(s, a)]$ 
     $s \leftarrow s'$  and  $a \leftarrow a'$ 
  Until  $s$  is terminal
  
```

圖三：SARSA 演算法

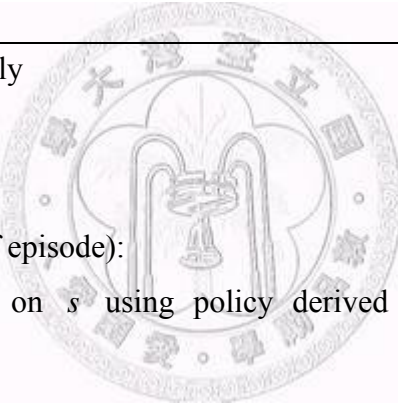
2.4.3 Q-learning

Q-learning 是一種 off-policy 時間差分控制方法，其基本架構與 SARSA 相同，不同的地方在於估算價值函數時，Q-learning 採用的是下一個狀態中價值函數最高的動作之價值函數值，而非學習代理人在下一個狀態真正採用的動作的價值函數值，此即為 off-policy 的含意，其計算方式如下：

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]$$

其中， $\max_a Q(s_{t+1}, a)$ 表示從下一個狀態 s_{t+1} 所有可能的動作中挑選出價值函數最高的動作 a 的價值函數值做為調整之用，而非下一個狀態真正採用的動作的價值函數值，在下一個狀態真正採用的動作的價值函數值應為 $Q(s_{t+1}, a_{t+1})$ 。

其演算法如下：



```
Initialize  $Q(s, a)$  arbitrarily
Repeat (for each episode)
  Initialize  $s$ 
  Repeat (for each step of episode):
    Choose  $a$  based on  $s$  using policy derived from  $Q$  (e.g.,  $\epsilon$ -greedy method)
    Take action  $a$ , obtain reward  $r$  and next state  $s'$  from the environment
     $Q(s, a) \leftarrow Q(s, a) + \alpha \left[ r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right]$ 
     $s \leftarrow s'$ 
  Until  $s$  is terminal
```

圖四：Q-learning 演算法

第三章 研究方法

3.1 系統設計

3.1.1 狀態

狀態是用來描述與代表環境，學習代理人會根據接收到的狀態去選擇相對應的動作，因此狀態的定義方式會是影響學習績效的關鍵。而在本次的研究過程中發現要找到適當的狀態定義方式是很大的挑戰，一個過於簡化的狀態定義方式會無法描述環境的變化，而過於複雜的定義方式又會需要較多的訓練資料與次數，否則在應用到測試環境時，會出現過多未學習過的狀態，導致結果的不穩定或出現較差的學習績效。

在過去的相關研究當中，大部份都是採用各項技術分析指標做為狀態定義的依據，而這些技術指標則是基於市場的價格跟交易量計算出來的，因此本研究根據價與量定義了三種不同的狀態並比較其學習績效：

1. 狀態 A：採用價格平均線與交易量平均線。
2. 狀態 K：採用 K 線的各組成部分的長度，意即最高價-最低價、開盤價-收盤價、最高價-開盤價、收盤價-最低價。
3. 狀態 C：採用狀態 A 與狀態 K 的欄位。

3.1.2 動作

動作定義為：

1. Hold: 維持目前留倉口數，不建立新倉，也不平倉。
2. Long1: 買進一口期貨。
3. Short1: 賣出一口期貨。
4. Clear: 將留倉部位全部平倉。

3.1.3 報酬

學習代理人根據接收到的狀態做出相對應的動作後，本系統會根據其選擇的動作及當時環境的市場價格計算報酬。

3.1.4 價值函數

本研究將價值函數的起始值設定為 0，其後調整的方式則依據該模型採用的演算法進行調整。

3.2 實驗設計

3.2.1 實驗資料

本研究採用 2004 年 1 月 1 日至 2008 年 6 月 30 日之臺灣股價指數期貨歷史資料作為系統的環境設計與績效計算依據。其中 2004 年 1 月 1 日至 2007 年 12 月 31 日的資料為學習代理人的學習訓練期，在此期間學習代理人將反覆地運用此資料學習，最後獲得一個投資策略。而 2008 年 1 月 1 日至 2008 年 6 月 30 日將做為學習代理人的學習績效檢測期，學習代理人將根據其習得之交易策略做出決策，系統將會累計此期間獲得之報酬，做為績效評比的依據。

3.2.2 交易環境設定

本研究採用臺灣股價指數期貨做為投資標的，點數一點為 200 元，為降低投資風險，選擇當日沖銷交易，意即當天建立之新倉部位必須當日平倉，否則期貨商可在收盤前強制平倉。根據臺灣期貨交易所網站公告之「臺灣期貨交易所各期貨及選擇權契約保證金一覽表」，台股期貨之原始保證金為 87,000 元，當日沖銷交易之保證金為一般交易之 50%，因此保證金應為 43,500，但為避免因價格變動劇烈而需時時留意是否要補足保證金，我們將系統中的保證金設定為 100,000，而這樣的設定會使系統的獲利率較實際上為低。計算獲利時，交易本

金設定為保證金×最大口數。交易成本為來回手續費加交易稅，設定為 2 點(400 元)。

3.2.3 實驗模型設計

本研究在實驗模型的設計上，除了考慮影響績效的狀態定義方式與增強式學習法的兩種演算法外，另外加入了停損點與停利點的設置，在沒有設置停損點、停利點的模型當中，其買、賣訊號的決定完全取決於學習代理人所學習到的策略；若有設置停損點、停利點，在持倉的狀態下當獲利已達停利點或損失已達停損點，則不管學習代理人發出的訊號為何，皆執行平倉的動作。

本研究根據三種狀態定義、兩種增強式學習演算法及是否設置停損、停利點，設計了 12 種實驗模型，如下表一。

模型代號	狀態	演算法	停利停損
SAQL00P00	狀態 A	Q-learning	不停利、不停損
SAQL50P50	狀態 A	Q-learning	停利點：50 點；停損點：50 點
SASL00P00	狀態 A	SARSA	不停利、不停損
SASL50P50	狀態 A	SARSA	停利點：50 點；停損點：50 點
SKQL00P00	狀態 K	Q-learning	不停利、不停損
SKQL50P50	狀態 K	Q-learning	停利點：50 點；停損點：50 點
SKSL00P00	狀態 K	SARSA	不停利、不停損
SKSL50P50	狀態 K	SARSA	停利點：50 點；停損點：50 點
SCQL00P00	狀態 C	Q-learning	不停利、不停損
SCQL50P50	狀態 C	Q-learning	停利點：50 點；停損點：50 點
SCSL00P00	狀態 C	SARSA	不停利、不停損
SCSL50P50	狀態 C	SARSA	停利點：50 點；停損點：50 點

表一：各交易模型一覽表

第四章 實驗結果與分析

4.1 各模型績效一覽表

本研究以表二所列之項目做為各操作模型效評估的標準，在損益計算時，以一日為一單位進行之。

績效評估項目		說明
獲利分析	總損益	績效測試期間總獲利或總虧損金額。 計算方式為：總獲利+總虧損
	總虧損	績效測試期間單日交易虧損總合(負值)
	總獲利	績效測試期間單日交易獲利總合
	策略獲利率(%)	績效測試期間之報酬率。 計算方式為：總損益/本金
	策略年化獲利率(%)	以一年 260 個交易日為計算基礎，年化計算策略獲利率。 計算方式為：(策略獲利率(%)/分析期間交易天數) × 260
風險分析	單日最大虧損	績效測試期間單日最大虧損
	單日最大獲利	績效測試期間單日最大獲利
	平均虧損	績效測試期間每日平均虧損。 計算方式為：總虧損/虧損天數
	平均獲利	績效測試期間每日平均獲利 計算方式為：總獲利/獲得天數
	平均每日損益	績效測試期間每日平均獲利 計算方式為：總損益/總投資天數
	每日損益標準差	績效測試期間每日獲利之標準差
週轉率分析	總交易次數	績效測試期間總交易次數
	最大留倉量	績效測試期間最大留倉部位
其它分析	虧損天數	績效測試期間單日投資虧損總天數
	獲得天數	績效測試期間單日投資獲利總天數
	勝率(%)	績效測試期間收益天數佔交易天數之比例，其中交易天數為虧損天數+獲得天數

表二：績效評估項目說明表

模型代號：SAQL00P00			
總損益	(\$319,600)	勝率	48.45%
總虧損	(\$844,600)	總獲利	\$525,000
策略獲利率(%)	-79.9%	策略年化獲利率(%)	-174.59%
平均每日損益	(\$2,685.71)	每日損益標準差	17271
總交易次數	153	最大留倉量	4
單日最大虧損	(\$80,200)	單日最大獲利	\$49,600
虧損天數	50	獲利天數	47
平均虧損	(\$16,892)	平均獲利	\$11,170.21

表 三：交易模型 SAQL00P00 績效表

模型代號：SAQL50P50			
總損益	(\$247,400)	勝率	46.39%
總虧損	(\$794,000)	總獲利	\$546,600
策略獲利率(%)	-82.47%	策略年化獲利率(%)	-180.18%
平均每日損益	(\$2,079)	每日損益標準差	16137
總交易次數	153	最大留倉量	3
單日最大虧損	(\$63,400)	單日最大獲利	\$44,800
虧損天數	52	獲利天數	45
平均虧損	(\$15,269.23)	平均獲利	\$12,146.67

表 四：交易模型 SAQL50P50 績效表

模型代號：SASL00P00			
總損益	(\$236,800)	勝率	45.74%
總虧損	(\$747,000)	總獲利	\$510,200
策略獲利率(%)	-59.20%	策略年化獲利率(%)	-129.34%
平均每日損益	(\$1,990)	每日損益標準差	16399
總交易次數	138	最大留倉量	4
單日最大虧損	(\$80,200)	單日最大獲利	\$49,600
虧損天數	51	獲利天數	43
平均虧損	(\$14,647)	平均獲利	\$11,865.12

表 五：交易模型 SASL00P00 績效表

模型代號：SASL50P50			
總損益	(\$230,800)	勝率	47.87%
總虧損	(\$783,800)	總獲利	\$553,000
策略獲利率(%)	-76.93%	策略年化獲利率(%)	-168.09%
平均每日損益	(\$1,939.5)	每日損益標準差	16123
總交易次數	144	最大留倉量	3
單日最大虧損	(\$63,400)	單日最大獲利	\$44,800
虧損天數	49	獲利天數	45
平均虧損	(\$15,996)	平均獲利	\$12,288.89

表 六：交易模型 SASL50P50 績效表

模型代號：SKQL00P00			
總損益	\$514,200	勝率	54.21%
總虧損	(\$593,400)	總獲利	\$1,107,600
策略獲利率(%)	85.70%	策略年化獲利率(%)	187.24%
平均每日損益	\$4,321	每日損益標準差	22455
總交易次數	215	最大留倉量	6
單日最大虧損	(\$56,600)	單日最大獲利	\$101,400
虧損天數	49	獲利天數	58
平均虧損	(\$11,635.29)	平均獲利	\$19,096.55

表 七：交易模型 SKQL00P00 績效表

模型代號：SKQL50P50			
總損益	\$187,200	勝率	45.87%
總虧損	(\$651,000)	總獲利	\$838,200
策略獲利率(%)	31.20%	策略年化獲利率(%)	68.17%
平均每日損益	\$1,573.11	每日損益標準差	17040
總交易次數	213	最大留倉量	6
單日最大虧損	(\$35,800)	單日最大獲利	\$66,800
虧損天數	59	獲利天數	50
平均虧損	(\$11,033.90)	平均獲利	\$16,764

表 八：交易模型 SKQL50P50 績效表

模型代號：SKSL00P00			
總損益	\$204,000	勝率	50%
總虧損	(\$707,200)	總獲利	\$911,200
策略獲利率(%)	68.00%	策略年化獲利率(%)	148.57%
平均每日損益	\$1,714.29	每日損益標準差	22706
總交易次數	168	最大留倉量	3
單日最大虧損	(\$94,600)	單日最大獲利	\$101,400
虧損天數	51	獲利天數	51
平均虧損	(\$13,866.67)	平均獲利	\$17,866.67

表 九：交易模型 SKSL00P00 績效表

模型代號：SKSL50P50			
總損益	\$39,600	勝率	49.06%
總虧損	(\$727,800)	總獲利	\$767,400
策略獲利率(%)	13.20%	策略年化獲利率(%)	28.84%
平均每日損益	\$332.77	每日損益標準差	17415
總交易次數	176	最大留倉量	3
單日最大虧損	(\$66,600)	單日最大獲利	\$44,400
虧損天數	54	獲利天數	52
平均虧損	(\$13,477.78)	平均獲利	\$14,757.69

表 十：交易模型 SKSL50P50 績效表

模型代號：SCQL00P00			
總損益	\$561,400	勝率	58.33%
總虧損	(\$533,000)	總獲利	\$1,094,400
策略獲利率(%)	93.57%	策略年化獲利率(%)	204.43%
平均每日損益	\$4,717.64	每日損益標準差	26479.9
總交易次數	181	最大留倉量	6
單日最大虧損	(\$75,400)	單日最大獲利	\$191,000
虧損天數	40	獲利天數	56
平均虧損	(\$13,325)	平均獲利	\$19,542.86

表 十一：交易模型 SCQL00P00 績效表

模型代號：SCQL50P50			
總損益	\$374,600	勝率	58.33%
總虧損	(\$453,200)	總獲利	\$827,800
策略獲利率(%)	62.43%	策略年化獲利率(%)	136.41%
平均每日損益	\$3,147.9	每日損益標準差	21374.9
總交易次數	173	最大留倉量	6
單日最大虧損	(\$54,600)	單日最大獲利	\$165,400
虧損天數	40	獲利天數	56
平均虧損	(\$11,330)	平均獲利	\$14,782.14

表 十二：交易模型 SCQL50P50 績效表

模型代號：SCSL00P00			
總損益	\$486,600	勝率	53.33%
總虧損	(\$456,600)	總獲利	\$942,600
策略獲利率(%)	81%	策略年化獲利率(%)	176.97%
平均每日損益	\$4,084.03	每日損益標準差	22741
總交易次數	147	最大留倉量	6
單日最大虧損	(\$75,400)	單日最大獲利	\$155,200
虧損天數	42	獲利天數	48
平均虧損	(\$10,871.43)	平均獲利	\$19,637.5

表 十三：交易模型 SCSL00P00 績效表

模型代號：SCSL50P50			
總損益	\$269,000	勝率	56.04%
總虧損	(\$408,000)	總獲利	\$677,000
策略獲利率(%)	44.83%	策略年化獲利率(%)	97.96%
平均每日損益	\$2,260.5	每日損益標準差	14582
總交易次數	142	最大留倉量	6
單日最大虧損	(\$35,600)	單日最大獲利	\$56,600
虧損天數	40	獲利天數	51
平均虧損	(\$10,200)	平均獲利	\$13,274.51

表 十四：交易模型 SCSL50P50 績效表

4.2 各模型績效比較分析

我們從上一節各模型績效一覽表中的總損益、獲利率及每日損益標準差繪製成圖表以便進行比較。選擇總損益及獲利率是為了觀察其獲利能力，每日損益標準差則用來當作風險觀察指標。

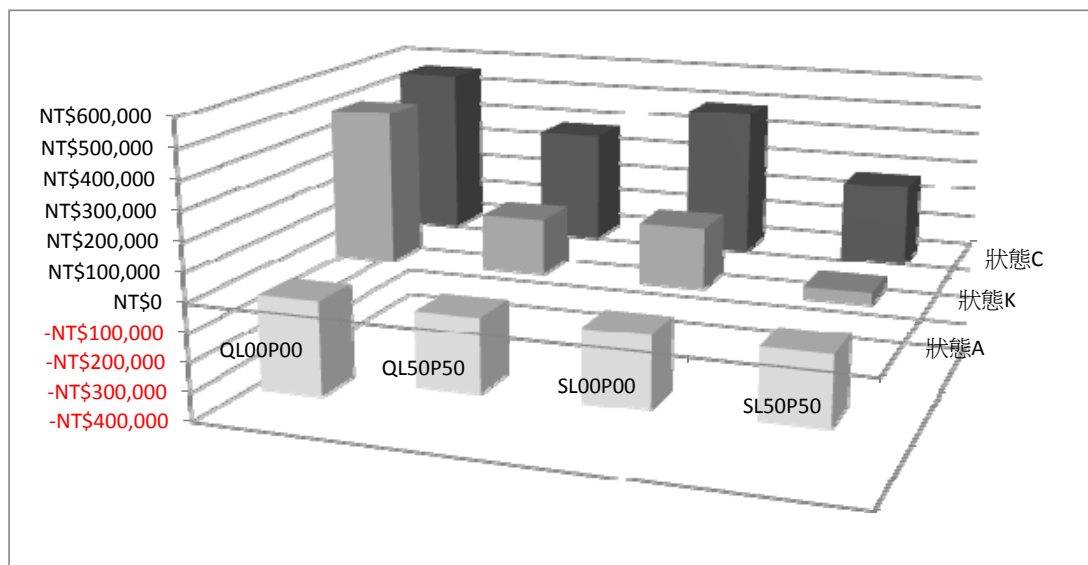


圖 五：總損益比較圖

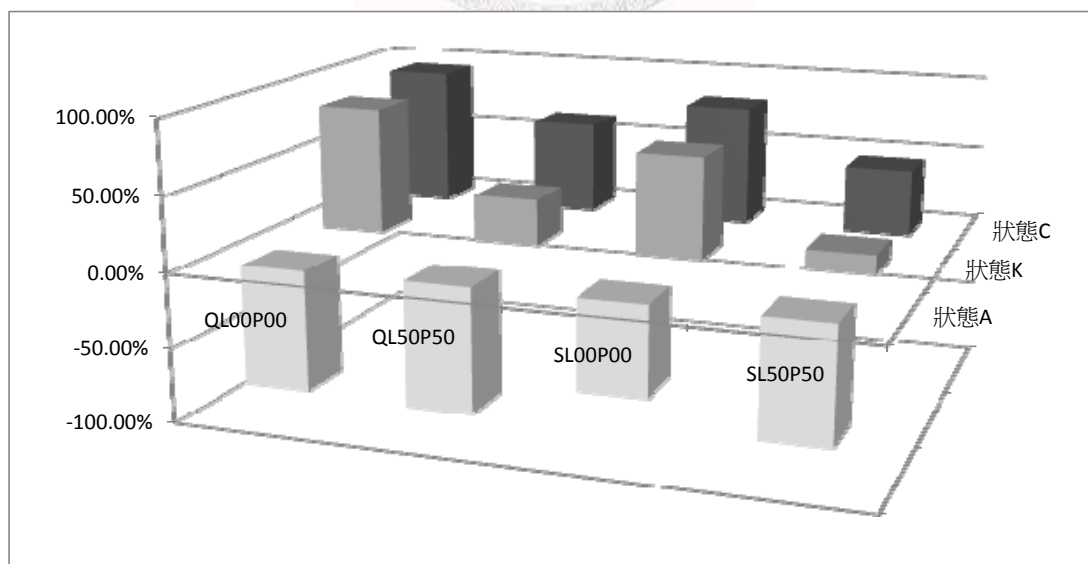


圖 六：獲利率比較圖

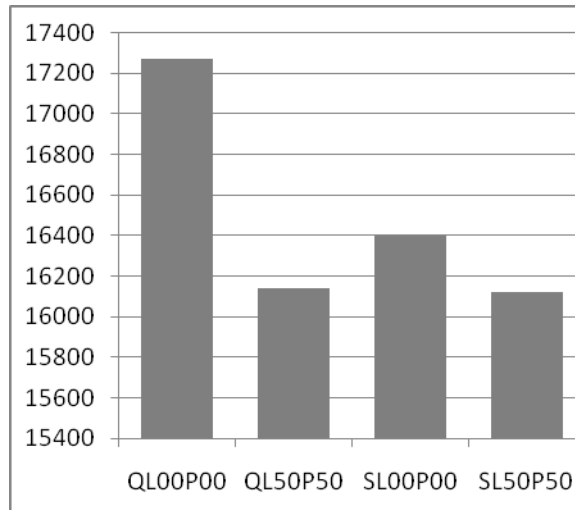


圖 七：狀態 A 之每日損益標準差

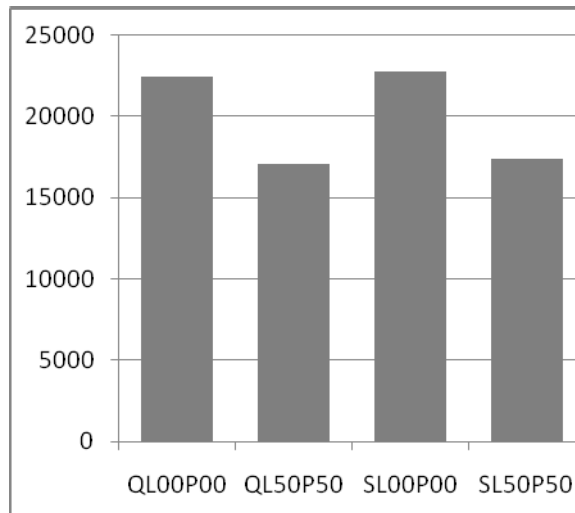


圖 八：狀態 K 之每日損益標準差

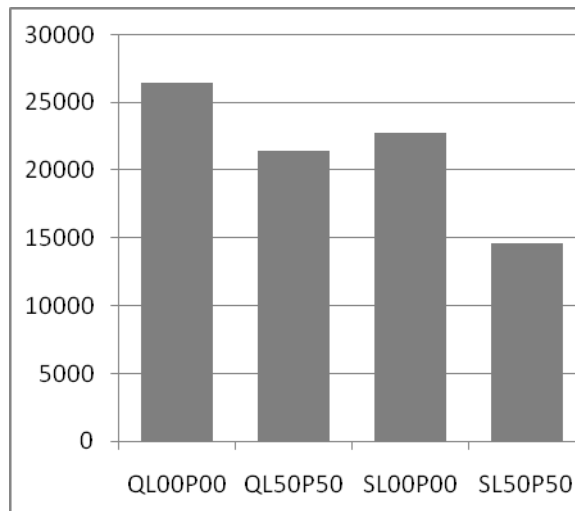


圖 九：狀態 C 之每日損益標準差

由以上的圖形可觀察到幾個現象：

1. 不同的狀態定義方式其獲利能力不同。

由圖五及圖六可知不同狀態定義方式的投資績效相差很多，在本研究設計的三種狀態定義方式中，狀態 A 無論以何種演算法或是設置停損、停利點與否，其投資獲利率皆為負值，而狀態 C 的表現最好，可知狀態 C 的定義方式最能描述環境。

2. 設置停損、停利點之獲利能力較差。

由圖五及圖六可知有設置停損、停利點的模型其獲利能力低於沒有設置停損、停利點的模型。為探討此一現象，我們針對獲利表現最佳的 SCQL00P00 及 SCQL50P50 進行觀察，發現在 119 天的測試期間中，有 76 日這兩種模型的交易動作與獲利表現是一致的。我們進一步針對表現不同的 43 個交易日進行分析，發現其中有 26 日 SCQL00P00 的表現較佳，有 17 日 SCQL50P50 的表現較佳，而 SCQL00P00 表現較佳的 26 日中，SCQL00P00 平均獲利較 SCQL50P50 多了 17,446 元，而 SCQL50P50 表現較佳的 17 日中，SCQL50P50 平均獲利較 SCQL00P00 只多了 15,694 元，其相關比較如表十五。

	SCQL00P00 的平均獲利	SCQL50P50 的平均獲利	獲利差
SCQL00P00 表現較佳的 26 日	24,931	7,485	17,446
SCQL50P50 表現較佳的 17 日	-9,718	5,976	15,694

表 十五：SCQL00P00 與 SCQL50P50 平均獲利比較表

觀察 SCQL50P50 表現較佳的交易日中的交易動作，由表十六可知，有設置停利點的 SCQL50P50 模型能在獲利時及時獲利了結。

價格	SCQL00P00	SCQL50P50
8548	做多一口	做多一口
8598	無動作	達停利標準，作空一口。 獲利 50 點。
8555	當沖結倉，獲利 7 點。	無動作

表 十六：2008/3/27 交易動作比較表

觀察 SCQL00P00 表現較佳的交易日中的交易動作，由表十七可知，有設置停利點的 SCQL50P50 模型會較早獲利了結而喪失後續較多的獲利機會。

價格	SCQL00P00	SCQL50P50
8999	做空一口	做空一口
8940	無動作	達停利標準，做多一口。 獲利 59 點。
8920	做空一口	做空一口
8867	無動作	達停利標準，做多一口。 獲利 53 點。
8788	當沖結倉，獲利 343 點。	無動作

表 十七：2008/5/23 交易動作比較表

以一段期間累積的結果來看，沒有設置停損、停利點的模型表現較佳，但藉由表十六與表十七的分析，我們發現設置停損、停利點有時能幫助投資人即時獲利了結，但有時卻會使投資人錯過後續較大幅度的獲利。

3. 設置停損、停利點之投資風險較低。

由圖七、圖八及圖九可知有設置停損、停利點的模型其風險較低。雖然由點 2 可知有設置停損、停利點的模型其獲利能力較差，但是投資績效的好壞不僅要考慮獲利能力，也必須將投資風險包含在內，因此若風險承受度較低的投資者可選擇手動設立停損、停利點以降低風險。

4. Q-learning 的獲利率較 SARSA 高。

由圖五及圖六可知在獲利能力為正的兩種狀態定義中，Q-learning 的獲利能力高於 SARSA 的獲利能力，這符合 Q-learning 與 SARSA 的特性，也就是 Q-learning 較 SARSA 易獲得最佳解。

5. Q-learning 的投資風險較 SARSA 高。

由圖七、圖八及圖九可知無論何種狀態定義方式及是否有設置停損、停利點，Q-learning 的投資風險較 SARSA 來得高，這符合 Q-learning 與 SARSA 的特性，即為 SARSA 所學到的策略其風險較低。

第五章 總結與展望

5.1 結論

本研究以增強式學習法建立台股期貨當沖交易策略。我們根據三種不同的狀態定義、兩種不同的演算法及是否設置停損、停利點設計出 12 種不同的交易模型，針對各模型分別進行訓練及測試。由各模型的測試結果可知，只要狀態定義設計足以描述環境，就能獲得不錯的投資績效，證明了增強式學習法在交易策略發展上的可用性。

採用不同的增強式學習演算法，其結果也不同，SARSA 與 Q-learning 最大的差別在於其價值函數的估算方式，因其估算方式的不同造成兩種演算法具有不同的特色，而這些特色也反映到模型的投資績效上，投資者可依其投資目標及風險承受度選擇適合的投資模型使用。

本研究也發現設置停損、停利點會干擾增強式學習法的學習效果，設置停損、停利點會造成獲利能力的下降，但是也發現設置停損、停利點可降低風險，投資者亦可考量自身情況選擇是否要設置停損、停利點。

5.2 未來展望

本研究已證明增強式學習法在建立交易策略上的可用性，將來可擴充系統的相關設定與參數，以期獲得更穩定的投資績效，以下提出 3 點建議：

1. 擴充動作的選擇：目前的動作設計為每次建倉或是平倉時皆以一口為單位，未來在設計上可進一步擴充，增加資金運用的彈性，應可獲得較佳的投資績效。
2. 找出更適切的狀態定義方式：本研究已找出狀態 K 與狀態 C 的設計方式可獲得不錯的投資績效，相信應有更進一步的設計方式，可讓交易策略獲得更好的投資績效與更低的投資風險。

3. 嘗試不同的停損、停利點的組合：本研究僅採用停損點 50 點與停利點 50 點的組合進行實驗，將來可嘗試不同的停損、停利點設置組合，以找到最符合該模型的風險控管。
4. 結合監督式學習與增強式學習法：監督式學習法與增強式學習法各有所長，若能找出結合雙方優點的方式，應可比採取單一學習法獲得更佳的表现。



參考文獻

- [1] 林典南，“使用 AdaBoost 之臺股指數期貨當沖交易系統”，國立臺灣大學資訊工程研究所碩士論文，2008。
- [2] 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。
- [3] 姜林杰祐，“程式交易系統 設計與建構”，新陸，2007。
- [4] Jae Wan Lee, “Stock Price Prediction Using Reinforcement Learning”, *IEEE International Joint Conference on Neural Networks*, 690–695, Washington D.C., 2001.
- [5] John R. Nofsinger, *Psychology of Investing*. Prentice Hall, NJ, 2002.
- [6] R. J. Kuo, “A Decision Support System For The Stock Market Through Integration of Fuzzy Neural Networks and Fussy Delphi”, *Applied Artificial Intelligence*, 6:501–520, 1998.
- [7] Peitsang Wu, Kung-Jiuan Yang, Zhao-Jung Lian, and Zi-Po Lin, “The Intelligent Trading Strategy System on Taiwan Stock Index Options”, *Proceedings of the 1st International Conference on Information Management and Business*, Taipei, Taiwan, 2005.
- [8] Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 1998.