國立清華大學 電機工程學系

實作專題研究成果摘要

Reinforcement Learning-Based Strategies for High-Frequency Trading

高頻交易應用強化學習演算法策略

專題領域：系統領域

組 別：B455

指導教授：馬席彬 教授

組員姓名：陳冠菻

研究期間：113 年 1 月至 113 年 11 月止

Abstract

With the rapid advancement of technology and the rise of artificial intelligence, fintech has also flourished. Among these developments, a Q-learning-based high-frequency trading agent system stands out by gradually learning market price fluctuations and trade instructions. The agent system is capable of making optimal trading decisions in a dynamic market environment. The core of this strategy lies in leveraging the Q-learning algorithm to continuously adapt to changes in market conditions, thereby enhancing the expected returns of the trading strategy. Our goal is to maximize cumulative returns and improve the overall performance of the strategy in real-world scenarios that account for trading costs.

摘要

隨著科技與人工智慧的快速發展，金融科技也蓬勃興起。其中，以 Q-learning 為基礎的高頻交易代理系統尤為突出，透過逐步學習市場價格波動與交易指令，能在動態市場中做出最優決策。該策略核心在於利用 Q-learning 算法適應市場變化，提升交易策略的預期收益。目標是在考量交易成本的真實情境下，最大化累積回報並優化整體表現。

1.Introduction

研究動機

隨著金融市場日益複雜化與數據化，高頻交易逐漸成為金融科技領域的重要支柱。高頻交易依靠速度與技術，實現了對市場微小價格差異的迅速反應，極大提高了資本運作的效率。然而，傳統策略往往依賴歷史經驗設計，缺乏對動態市場條件的即時適應能力，難以應對市場中不可預測的波動。基於強化學習的 Q-learning 演算法提供了一種新的解決途徑，透過代理系統與市場環境的互動，持續學習並優化交易策略。本研究希望藉由結合 Q-learning 與高頻交易，探索一種能動態調整並適應市場變化的創新策略，進一步提升交易決策的效率與收益表現。

研究目的

本研究旨在開發一個基於 Q-learning 的高頻交易代理系統，通過動態學習市場價格波動特性，實現對市場趨勢的準確判斷，並制定最優交易策略。研究的核心目標是最大化累積收益，確保在考量交易成本的前提下，達到穩定的資本增長。同時，本研究注重風險控制，通過合理的獎勵設計與適當的交易頻率，減少市場波動對收益的影響。此外，系統的設計還需具備良好的泛化能力，能夠適應不同市場條件，為高頻交易在實際金融應用中提供穩健且高效的解決方案。

2. Research Methodology

研究方法

本研究基於Q-learning演算法設計高頻交易代理系統，核心在於對市場狀態（state）的精確描述、行動（action）的合理設計以及獎勵（reward）機制的嚴謹構建，確保模型能在動態市場中有效學習和適應。

首先，市場狀態的設計以反映市場當前特徵為目標，選擇了一組能描述價格趨勢與交易量變化的技術指標作為狀態變數，包括短期與長期移動平均線（MA）、RSI指標、MACD以及交易量的移動平均值等。例如，短期與長期移動平均線用於捕捉市場的黃金交叉和死亡交叉信號，幫助模型判斷買入或清倉的時機；RSI指標則衡量市場的超買或超賣程度，提供額外的市場動量信息；MACD與信號線差值則輔助模型識別市場上漲或下跌趨勢。此外，交易量移動平均值反映市場活躍度，結合價格變化趨勢，有助於模型更準確地評估當前市場狀況。這些狀態變數的選擇既覆蓋了價格、量能和市場動量，又保持了特徵數量的適度，避免過多變數導致的計算負擔。

行動的設計則聚焦於模擬投資者在高頻交易中的主要操作，包括保持倉位（Hold）、買入（Long）和清倉（Clear）。保持倉位適用於市場信號不明確或等待更佳進場時機的情況；買入行動主要在市場呈現明確上漲趨勢時執行，期望隨價格上漲獲利；清倉則在價格達到目標收益、出現風險信號（如死亡交叉、RSI超買）或價格突破布林帶時觸發，以保護資本或實現收益。本研究採用僅做多（long-only）策略，避免了做空操作帶來的高風險，同時簡化了模型學習的複雜性，確保交易行為更符合大多數市場環境。

獎勵機制的設計是 Q-learning 的核心，直接影響模型對不同行動的學習偏好。本研究的獎勵函數以資本增值為核心目標，並同時考慮了交易成本與風險控制。當執行買入操作時，獎勵根據當前價格與平均買入成本的差額計算，扣除交易成本後即為淨收益；當執行清倉操作時，則根據清倉價格與持倉成本的價差計算收益，若價格上漲則給予正向獎勵，若價格下跌則產生負向獎勵。此外，研究引入了幾項特殊條件來進一步優化獎勵機制。例如，當價格漲幅超過設定的獲利目標（如 5%）時，提供額外的高額獎勵以鼓勵模型識別並抓住盈利機會；若價格跌幅超過止損門檻（如 5%），則減少獎勵以懲罰風險管理不善的行為。同時，當 RSI 超買或價格突破布林帶上限時，設置清倉動作獲得合理獎勵，幫助模型及時退出高風險市場，保護收益。

整體來看，模型通過這一狀態、行動與獎勵的協同設計，實現了對市場趨勢的準確判斷與資本增值的穩健管理。Q-learning 算法通過不斷更新狀態-行動對應的 Q 值，學習如何在不同市場條件下選擇最優行動。隨著訓練的深入，模型逐漸掌握市場規律，並能在測試階段有效適應未見數據，表現出較高的勝率與穩定的收益能力。

3. Experimental Results

勝率(Win Ratio)：

高 Win Ratio 表明模型在策略執行中有較高的精準度，但勝率並非唯一指標，仍需結合獲利幅度與風險進行綜合評估。例如，雖然勝率高，但若每次獲利的金額較低，而虧損時損失較大，則仍可能導致整體收益不足。

$$\text{Win Ratio} = \frac{\text{Number of Winning Trades}}{\text{Total Number of Trades}} \times 100\%$$

總收益率(Profit Percentage)：

Profit Percentage 不僅體現了策略的盈利能力，也反映了 Reward 設計的合理性。在交易次數與資本使用效率之間，模型達成了一個平衡，使得整體收益能持續增長。

$$\text{Profit Percentage} = \frac{\text{Final Capital} - \text{Initial Capital}}{\text{Initial Capital}} \times 100\%$$

測試期間交易次數(Test Trade Times)：

交易次數的設計取決於策略的特性，適量的交易能有效提升收益，但過多交易可能會增加交易成本或導致模型對隨機市場波動過於敏感。此次實驗中的交易頻率控制得當，為策略執行提供了穩定基礎。

$$\text{Test Trade Times} = \sum \text{Trades Executed in Testing Period}$$

Hyperparameters Setting

· **學習率（alpha）**

設置：0.001-0.05

定義：學習率控制每次更新 Q 表時，新信息對已有 Q 值的影響比例。較小的學習率可以穩定更新，但可能導致收斂速度較慢；較大的學習率則 可以快速更新，但可能引入噪聲。

說明：學習率決定了每次 Q 表更新時，新資訊對已有數值的影響程度。 較低的學習率（如 0.001）適合在數據波動較大或模型需要更穩定收斂的 情況下使用，避免更新過快導致結果不穩定。而較高的學習率（如 0.05） 則適合在早期快速適應市場變化，加速學習過程。本實驗中採用可調的學 習率範圍，兼顧穩定性與學習效率。

· **折扣因子（gamma）**

設置：0.995

定義：折扣因子權衡短期收益與長期收益的權重。接近 1 時，模型更關注長期回報；接近 0 時，更注重短期獎勵。

說明：折扣因子控制了模型對長期回報的重視程度。接近 1 的折扣因子（如 0.995）意味著模型會更看重未來的累計收益，適合具有長期趨勢的市場環境。同時，這樣的設置能幫助模型在捕捉短期信號的同時， 保持對整體趨勢的敏感度，適合於金融交易場景。

· **探索率（epsilon）**

設置：0.001 - 0.01

定義：探索率控制隨機行動的比例，用於在學習過程中探索未知環境。較高的探索率增加多樣性，較低的探索率專注於已知最優策略。

說明：探索率控制隨機行動的比例，是平衡探索與利用的重要參數。在較高探索率（如 0.01）時，模型會更多嘗試未見的行動，以發現潛在的收益；而較低探索率（如 0.001）則專注於當前已知的最優策略，避免過多的隨機交易。本實驗通過降低探索率來提高模型的穩定性，同時保留適當的探索機會以避免局部最優。

- **交易成本（transaction cost）**

設置：500 單位

定義：每次交易的固定成本，用於模擬市場手續費或其他費用。

說明：交易成本的引入可以限制不必要的頻繁交易，讓模型在成本和收益之間找到最佳平衡。

- **初始資本（initial capital）**

設置：1 億

定義：初始資本是模擬交易的起始資金，用於計算收益率和資本管理策略。

說明：設置為較大金額，可以更好地模擬實際市場中大規模交易的場景。

- **獲利目標與止損條件**

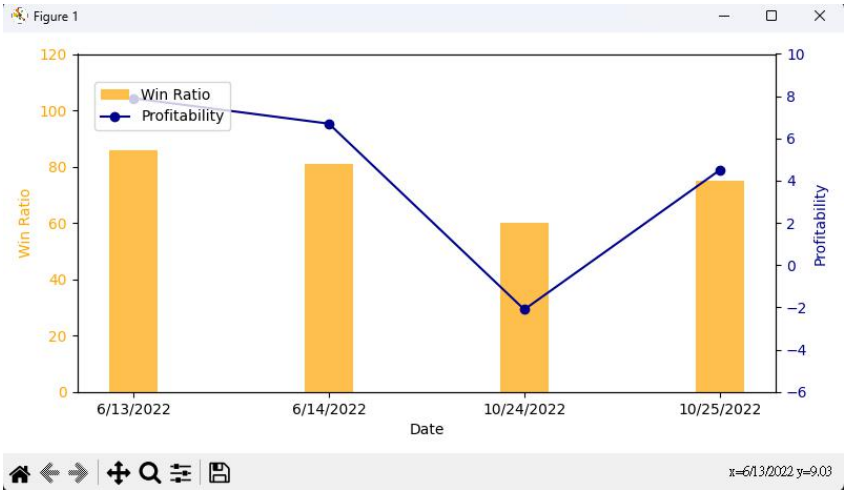
設置：獲利目標 5%，止損條件 5%。

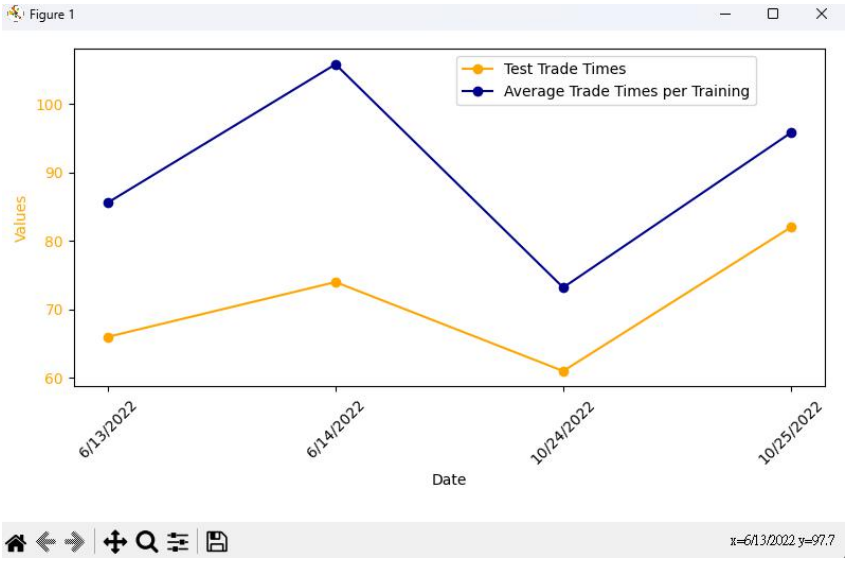
定義：這些參數控制交易的風險管理邏輯，當收益或損失超過指定閾值時執行清倉。

說明：設定對稱的獲利和止損條件，可以有效控制風險，同時確保在合適的時機實現收益。

Testing Result

simulation	value
initial cash	10000000
trading days	2022/6/13
	2022/6/14
	2022/10/24
	2022/10/25
trading time	09:00~12:30
product id	6/13~14:MXFF2
	10/24~25:MXFK2





在測試階段，我們將訓練後的 Q-learning 模型應用於歷史市場數據，檢驗其在未見數據上的表現。結果顯示，模型在其中 3 天的交易勝率接近 80%，顯示其具備良好的趨勢判斷能力與穩定性。初始資本 1 億元中，3 天均達正收益，最高達 8%，而表現最差的一天僅損失 2%。測試期間模型平均執行約 90 次交易，既敏感捕捉機會又避免過度交易，合理控制成本。交易行為與市場信號一致，進場基於黃金交叉，清倉則根據死亡交叉、RSI 超買等指標觸發，進一步提升策略的收益穩定性。結果證明模型在動態市場中具備穩定的資本增長能力與良好的應用價值。

Conclusion

本實驗結論顯示，基於 Q-learning 的交易策略能有效應用於未見數據，展現高勝率與穩定收益率，驗證了其對市場趨勢的準確判斷能力。多天測試設計幫助模型適應不同市場波動，全面評估其泛化能力。交易行為分析表明，模型進場基於黃金交叉，清倉則依據 RSI 超買或布林帶突破等條件，符合設定邏輯。合理的交易頻率與風險管理使策略在波動市場中平衡收益與風險，展現穩健的中短期持倉表現。未來可通過優化獎勵設計、加入做空操作及更高頻數據，提升模型靈敏度與收益能力，為高效可靠的交易系統奠定基礎。

Review and Reflection

一開始選擇這個專題的原因是對軟體方面及演算法有興趣，原本以為可以像程式設計課一樣只要有想法就能將程式打出來跑出結果。沒想到在一開始找資料就遇到很多問題，如找的論文沒有用或太深奧看不懂。且我在專題一開始是對機器學習和 python 不熟悉的，因此前期大部分時間都花在學習上。後期有了基本概念後開始自己一步步照著自己的想法設計演算法，有問題就上網找資料。在每

周一次的 meeting 中，教授也會針對我的報告提出想法，其中我覺得最有幫助的建議是要先找出問題的根本才從那下手，不要底都還沒搭好就想一步登天把後面的都完成。的確，我一開始的想法就希望能將整個模型做出來，但我在毫無基礎的情況下做只是事半功倍。於是之後我從最基本的 Q learning model 做起，在熟悉演算法後才開始加入專題的高頻交易部分。成效也在後面都看的出來。

這次的專題訓練了我很多能力，從最一開始的找資料和後來的打程式都是靠自己獨力完成，感謝教授及同專題不同組別的同學在旁的協助才能讓我成功完成。希望能將此份學習的能力帶給以後的我很多幫助。