

國立清華大學 電機工程學系

實作專題研究成果摘要

Optimizing Residual Self-Attention  
Models for TXF High-Frequency  
Trading : Triple-Barrier Labeling and  
Loss Function Analysis

優化殘差自注意力模型於台指期高頻  
交易：三重屏障標籤與損失函數分析

專題領域：系統領域

組別：B604

指導教授：翁詠祿 教授

組員姓名：羅弘、何震群

研究期間：114年2月17日 至 114年11月17日止，共 9 個月

## 摘要

高頻交易（HFT）策略高度依賴對限價訂單簿（Limit Order Book, LOB）微觀結構的解讀，以預測極短時間內的價格方向。本研究探討深度學習模型在不同市場微結構中的適用性，並以 ReSaLOB [2]模型應用於台灣指數期貨（TXF）市場為例，檢驗其跨市場遷移的限制。

實驗發現，ReSaLOB 原始針對 FI-2010 [3]基準資料集設計的「固定百分比閾值」標籤法，因未考慮台指期的高價格位階與低短期波動特性，導致短時間內的價格變動難以跨越門檻，使模型陷入僅預測持平的「模式崩塌（Mode Collapse）」困境。為突破此瓶頸，本研究提出一套優化框架：引入基於波動率調整的「三重屏障法（Triple-Barrier Method） [6]」，透過即時波動率動態調整漲跌門檻以過濾噪聲；並結合「加權交叉熵（Weighted Cross-Entropy） [7]」損失函數以改善類別不平衡。

實證結果顯示，此優化策略有效解決了數據訊噪比低的問題，並成功喚醒了模型對上漲與下跌趨勢的辨識能力。新方法於台指期資料上的 Macro F1 與分類表現均顯著優於原始設定，混淆矩陣亦展現了均衡的預測分佈。綜合而言，本研究建立出一套更符合台指期市場特性的預測框架，並證實了針對特定市場微結構進行標籤與損失函數客製化的必要性。

# 報告內容

## 1. 研究動機

高頻交易（HFT）策略的核心在於利用限價訂單簿（Limit Order Book, LOB）的微觀結構預測極短時間內的價格方向。近年來，結合卷積神經網路（CNN）與 Transformer 自注意力機制（Self-Attention）[4]的模型在學術界表現優異。然而，現有研究多以 FI-2010 [3]公開資料集為基準，該市場波動度較高。當我們將相同模型直接套用於台灣指數期貨（TXF）時，發現由於台指期短時間內的價格跳動幅度極小，原始的標籤定義方法會將絕大多數樣本歸類為「持平」，導致模型無法學習到有效的交易訊號。

本計畫的目標是建立一套適用於台指期市場特性的深度學習預測框架。我們旨在解決因市場微結構差異導致的「類別不平衡」與「模式崩塌」問題。具體而言，我們希望透過改進標籤生成機制與損失函數設計，讓 ReSaLOB 模型能夠在充滿噪聲的台指期數據中，精準辨識出具備交易價值的上漲與下跌趨勢。

## 2. 研究方法

### 2-1. 限價訂單簿預測與深度學習

鑑於高頻交易對低延遲的嚴苛要求，本研究選用 ReSaLOB [2]模型作為基礎架構。相較於傳統 DeepLOB [1]模型因 LSTM 序列運算而產生的高推論延遲，ReSaLOB 透過引入自注意力機制（Self-Attention）[4]實現平行運算，成功在維持高預測準確率的同時，大幅降低運算時間以符合實時交易需求。

### 2-2. ReSaLOB 模型架構

本研究採用輕量化的 ReSaLOB（Residual Self-Attention LOB）模型。該架構包含兩個核心部分：首先利用 LOB 感知殘差區塊（LOB-aware Residual Block），結合 SE [5]模組與實例正規化（Instance Normalization），以精確提取限價簿跨檔位與跨買賣方的微觀特徵，並適應金融數據的非平穩性；接著透過自注意力機制（Self-Attention）取代傳統 LSTM，在捕捉長短期時間序列依賴的同時實現平行運算，大幅降低推論延遲，使其更適用於高頻交易場景。

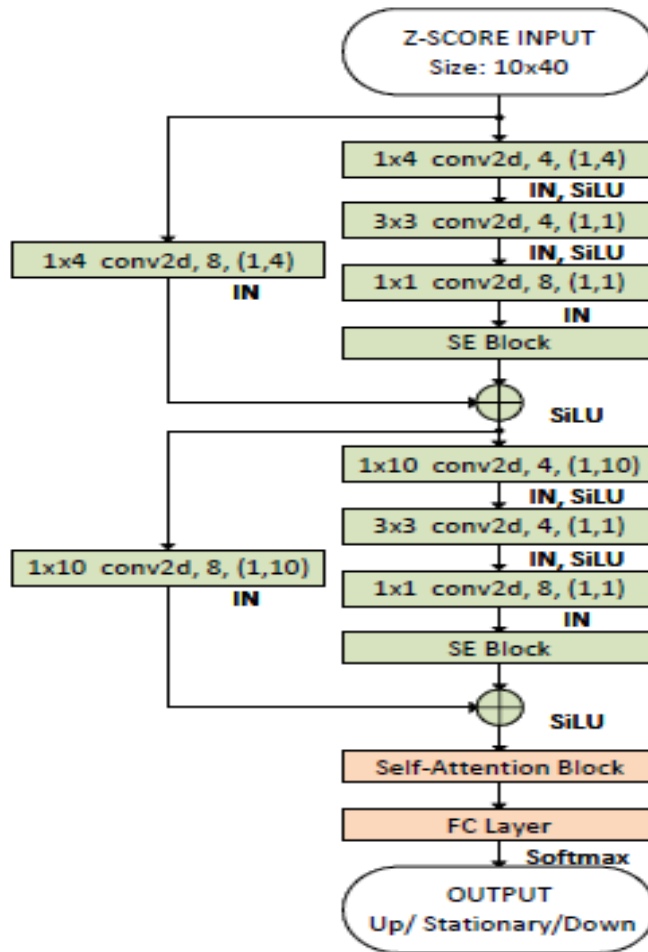


圖 1. ReSaLOB 模型架構圖

### 2-3. ReSaLOB 模型應用於台指期貨資料困境

研究發現，ReSaLOB 原始之固定百分比閾值標籤法不適用於台指期貨，因市場微結構差異導致嚴重類別不平衡。實驗證實，若強行降低閾值 ( $\alpha = 0.000025$ ) 以擬合基準分佈，其對應價格變動僅 0.425 點，遠小於市場最小跳動單位 (1 點)，顯示該標籤僅捕捉無效噪聲而非真實趨勢，無法作為有效預測依據。

### 2-4. 標籤定義優化

針對原始標籤法在台指期貨上的適應性問題，本研究引入「波動調整之三重屏障法 (Triple-Barrier Method) [6]」。此法依據價格路徑是否突破動態波動門檻來定義標籤，而非僅觀察固定時間點的漲跌。

#### 2-4-1. 價格與報酬率定義

令  $p_a^{(1)}(t)$  與  $p_b^{(1)}(t)$  為時間  $t$  之最佳賣價與買價，中間價 (Mid-price)  $p_t$  與對數價格  $l_t$  定義如下：

$$p_t = \frac{p_a^{(1)}(t) + p_b^{(1)}(t)}{2}, \ell_t = \log(p_t),$$

由此定義一階對數報酬 (Log-return)：

$$r_t = \ell_t - \ell_{t-1}.$$

#### 2-4-2. 動態波動度估計

考量高頻資料的異質變異性，本研究採用指數加權移動標準差 (EWM-Std) 估計即時波動度  $\sigma_t$ ，賦予近期數據較高權重：

$$\sigma_t = \sqrt{\frac{\sum_{i=1}^N \alpha (1-\alpha)^{i-1} (r_{t-i})^2}{\sum_{i=1}^N \alpha (1-\alpha)^{i-1}}}$$

為避免市場低迷時估計值過低，設定下限  $\sigma_t \leftarrow \max(\sigma_t, 10^{-6})$ 。

#### 2-4-3. 三重屏障與標籤生成

基於即時波動度，設定動態上下屏障 ( $\kappa = 1.5$ )：

$$U_t = +\kappa\sigma_t, D_t = -\kappa\sigma_t.$$

對於每一預測 horizon  $H$ ，我們比較未來價格相對於目前價格的 log-return 差異：

$$r_{t \rightarrow t+k} = \ell_{t+k} - \ell_t, 1 \leq k \leq H.$$

時間  $t$  的標籤由未來價格路徑首次觸及障礙的情況決定。具體而言，標籤  $l_t^{(H)}$  定義為：

$$l_t^{(H)} = \begin{cases} +1, & \text{若價格在 horizon 內先觸及上界 } U_t \\ 0, & \text{若在 horizon 內未觸及任何上下界} \\ -1, & \text{若在 horizon 內先觸及下界 } D_t \end{cases}$$

不同於傳統僅觀察固定時間點的價格變動，三重屏障法依據價格路徑是否突破波動屏障來定義標籤，能有效過濾市場微觀噪聲。實驗顯示，此機制在短預測視窗下能產出最具可學習性的標籤分布，成功解決了模型無法識別有效交易訊號的困境。

### 2-5. 損失函數設計

針對台指期數據在標記後呈現的顯著類別不平衡，本研究探討了損失函數的優化策略。由於「持平」類別佔絕大多數，傳統損失函數易導致模型陷入預測多

數類別的局部最優解。為解決此問題，本研究依序引入加權交叉熵（Weighted Cross-Entropy） [7]與 Focal Loss [8]進行實驗。實驗旨在找出能平衡模型對少數波動類別關注度，同時維持訓練穩定性的最佳優化方案。

### 2-5-1. 標準交叉熵（Standard Cross-Entropy）與其限制

多分類任務常用的標準交叉熵損失定義如下：

$$L_{CE} = - \sum_{c=1}^C y_c \log (p_c)$$

其中， $C=3$  為類別總數， $y_c$  為真實標籤的指示變數， $p_c$  為模型預測該樣本屬於類別  $c$  的機率。在嚴重不平衡資料中，多數類別（持平）累積的損失值將主導梯度下降方向，導致模型犧牲對少數類別（上漲/下跌）的判別能力，使其失去捕捉關鍵市場波動的實用價值。

### 2-5-2. 加權交叉熵（Weighted Cross-Entropy）

為了緩解上述樣本數量不均的問題，本研究首先採用 Weighted Cross-Entropy 損失函數。其核心概念是賦予少數類別較高的懲罰權重，迫使模型在訓練過程中對這些類別給予更高的關注。本研究採用「逆類別頻率（Inverse Class Frequency）」方法來計算每個類別  $c$  的權重  $w_c$ ：

$$w_c = \frac{N_{total}}{C \times N_c}$$

其中， $N_{total}$  為訓練集的總樣本數， $N_c$  為類別  $c$  的樣本數。在此機制下，出現頻率越低的類別，其權重  $w_c$  越大。修正後的損失函數公式為：

$$L_{WCE} = - \sum_{c=1}^C w_c y_c \log (p_c)$$

如此透過引入權重  $w_c$ ，即使少數類別的樣本稀缺，其對總損失的貢獻權重將被放大，從而平衡模型對各類別的學習力度，避免模型退化為僅預測多數類別。

### 2-5-3. Focal Loss

除了樣本數量的「靜態不平衡」，訓練過程中還存在樣本難易度的「動態不平衡」。在金融數據中，大量的持平樣本通常特徵明顯且易於分類，而趨勢反轉點（上漲/下跌）通常特徵模糊且較難分類。

為此，本研究進一步引入 Focal Loss。該函數在 Standard Cross-Entropy 的基礎上引入了一個調節因子  $(1 - p_t)^\gamma$ ，旨在降低模型對簡單樣本的關注

度，將優化重心聚焦於困難樣本。其公式定義如下：

$$L_{FL} = -\alpha(1 - p_t)^\gamma \log(p_t)$$

其中  $p_t$  為預測機率， $\gamma$  為聚焦參數。隨著簡單樣本的預測確信度提高（ $p_t \rightarrow 1$ ），調節因子  $(1 - p_t)^\gamma$  將趨近於 0，從而大幅降低其損失權重，使模型聚焦於困難樣本； $\alpha$  則採用前述之逆類別權重  $w_c$  進行平衡。此機制旨在自動抑制大量無效的盤整訊號，迫使模型專注於具交易價值的價格波動特徵。

### 3. 研究結果與討論

本研究將不同模型設定的實驗數據彙整於表 1。首先觀察原始 ReSaLOB 模型，雖然其準確率高達 94.74%，但 Macro F1 僅 32.43%，顯示模型陷入嚴重的模式崩塌，完全喪失對漲跌趨勢的預測能力。再著，針對損失函數的優化嘗試中，引入 Focal Loss 因過度抑制簡單樣本導致訓練不穩定，Macro F1 反降至 29.35%；而改用 Weighted Cross-Entropy 雖將 Macro F1 提升至 36.40%，且少數類別的召回率開始回升，但整體效能仍受限於原始標籤法的噪聲干擾。最後，本研究提出的「三重屏障法（Triple-Barrier Method）」結合「加權交叉熵（Weighted Cross-Entropy）」策略表現最佳。如表 1 所示，在預測視窗  $H = 10$  下，該組合成功解決了類別不平衡問題，將 Macro F1 分數由基準的 32.43% 大幅提升至 46.56%。而圖 2 進一步證實，模型在「下跌（Recall 45.4%）」、「持平（Recall 51.0%）」與「上漲（Recall 44.1%）」三個類別上均展現了均衡且具鑑別力的預測能力，證明此框架能有效從台指期微結構中提取具實戰價值的交易訊號。

表1. 各模型應用在台指期資料之結果比較表

預測視窗 $H = 10$				
Model	Accuracy %	Precision %	Recall %	Macro F1 %
ReSaLOB模型	94.74	31.58	33.33	32.43
ReSaLOB模型 + Focal Loss	78.65	26.22	33.33	29.35
ReSaLOB模型 + Weighted Cross Entropy	42.84	36.41	36.61	36.40
TB標籤法 + Weighted Cross Entropy	46.87	46.61	46.53	<b>46.56</b>

下跌	0.454	0.329	0.217
持平	0.247	0.510	0.243
上漲	0.249	0.319	0.441
預測/真實	下跌	持平	上漲

圖2. TB標籤法 + Weighted Cross Entropy應用在台指期貨資料之混淆矩陣圖

#### 4. 結論

本研究針對台指期 (TXF) 高頻交易方向預測進行了系統性的優化研究。我們以 ReSaLOB 模型為基礎架構 [2]，深入探討了標籤定義與損失函數在不同市場微結構下的適用性，主要結論如下：

首先，標籤定義對模型效能具有關鍵影響。實驗證實，直接沿用適用於成熟市場的固定百分比閾值標籤法，在台指期貨資料上會導致嚴重的類別不平衡與模式崩塌 (Mode Collapse)，使模型喪失預測能力，Macro F1 僅有 32.43%。

其次，損失函數的選擇需考量資料的訊噪比特性。在低訊噪比的環境下，高強度的 Focal Loss [8] 反而會因過度抑制簡單樣本而導致訓練不穩定。相比之下，加權交叉熵 (Weighted Cross-Entropy) [7] 能提供較穩健的梯度回饋，更適合處理此類高噪聲數據。最終，本研究提出的最佳解決方案為結合「波動調整之三道屏障法 (Triple-Barrier Labeling)」[6]與「加權交叉熵」。此策略成功解決了類別不平衡問題，在預測視窗  $H = 10$  下，將 Macro F1 分數由基準的 32.43% 大幅提升至 46.56%。混淆矩陣分析顯示，模型在三類方向上均具備均衡的辨識能力，證實了此框架在台指期貨高頻預測上的實用價值。

#### 5. 參考資料

- [1] Z. Zhang, S. Zohren, and S. Roberts, "DeepLOB: Deep Convolutional Neural Networks for Limit Order Books," *IEEE Transactions on Signal Processing*, vol. 67, no. 11, pp. 3001-3012, Jun. 2019.
- [2] T. Z. Lin, H. W. Wu, H. F. Chou, T. M. Huang, and Y. L. Ueng, "High-efficiency Residual Neural Networks with a Self-attention Mechanism for High-frequency Trading Systems" (實驗室文獻)
- [3] A. Ntakaris, M. Magris, J. Kannianen, M. Gabbouj, and A. Iosifidis, "Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods," *Journal of Forecasting*, vol. 37, no. 8, pp. 852-866, Dec. 2018.
- [4] A. Vaswani et al., "Attention Is All You Need," in *Advances in Neural Information Processing Systems (NIPS)*, 2017, pp. 5998-6008.

- [5] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132-7141.
- [6] M. Lopez de Prado, Advances in Financial Machine Learning. Hoboken, NJ, USA: Wiley, 2018.
- [7] H. He and E. A. Garcia, "Learning from Imbalanced Data," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263-1284, Sept. 2009
- [8] M. A. M. Al-Zeyadi, A. A. Al-Sultani, and A. H. Al-Sultani, "False awareness stock market prediction by LightGBM with focal loss," in Proceedings of the International Conference on Data Science and Intelligent Analysis of Information, 2022, pp. 145-154.

## 6. 心得感想

本專題讓我們針對高頻交易預測題目，從理論基礎到系統實作進行了完整的研究。

前期以研讀原文書與期刊論文為主，每週與博士班學長進行會議，透過「指定閱讀 → 摘要整理 → 口頭報告 → 問題延伸」的紮實訓練循環，我們大幅提升了文獻閱讀、知識整合與簡報表達的能力，並學會如何從艱澀的數學理論中萃取核心概念。

中期進入實作階段後，我們遭遇了意料之外的挑戰。由於台指期與文獻中資料集的市場特性差異，模型初期發生了嚴重的「模式崩塌」現象，失去有效的預測功能。這段經歷雖令人挫折，卻也讓我們體會到理論與實務的落差，並促使我們針對問題進行探討。

後期我們針對此問題進行了多方方法的比較與嘗試。我們主動引入「三重屏障法」重新定義標籤，並深入分析 Focal Loss 與 Weighted Cross-Entropy 在不平衡資料上的效益，最終收斂出最佳的優化策略。在驗證階段，我們透過混淆矩陣確認模型確實習得了捕捉漲跌訊號的能力，而不僅僅是追求表面的準確率。

總結來說，專題實作讓我們在查閱文獻、解決問題、時間分配和團隊分工方面等能力都有顯著的提升，相信這些經驗將為我們未來進入碩士班從事學術研究奠定堅實的基礎。