Transformer-based Reinforcement Learning Model for Optimized Quantitative Trading

Aniket Kumar, Rodrigue Rizk, KC Santosh
Applied AI Research Lab, Department of Computer Science, University of South Dakota, Vermillion, SD 57069, USA
aniket.kumar@coyotes.usd.edu, rodrigue.rizk@usd.edu, kc.santosh@usd.edu

Abstract—**Stock market prediction has long been a focal point of financial research due to its immense potential for investors and policymakers. It is a challenging task, as it involves forecasting the future prices of assets based on historical data. Traditional quantitative trading strategies often face challenges in adapting to dynamic market conditions and capturing intricate patterns in financial data. In response, this work explores a novel approach to stock prediction that combines the power of Transformer architecture with the decision-making capabilities of reinforcement learning. We propose a quantitative trading framework that integrates a transformer-based encoder-decoder network to predict future stock prices and a reinforcement learning agent to optimize investment strategies based on these predictions. The Transformer network leverages its self-attention mechanism to capture complex relationships within historical price data, while the reinforcement learning agent learns optimal actions through trial and error in a simulated environment. The experimental results demonstrate the effectiveness of our approach in improving stock prediction accuracy compared to existing methods. This work highlights the potential of combining Transformer and reinforcement learning for efficient and profitable stock market navigation.**

Index Terms—**Stock Prediction, Reinforcement Learning, Transformer, Quant Trading, Algorithmic Trading, Financial Markets**

I. Introduction

Stock is a type of security that represents ownership in a company. When you buy a stock, you are essentially buying a small piece of that company. Stocks can be a risky investment, but they also have the potential to generate high returns. Over the long term, the stock market has historically outperformed other types of investments, such as bonds and savings accounts [1]. Stock prediction is the process of forecasting the future movement of stock prices. It is a challenging task due to the complex and non-linear nature of stock markets. The financial markets operate in an ever-changing and complex environment influenced by a multitude of factors such as economic indicators, geopolitical events, market sentiment, and investor behavior [1]. Investors and financial institutions constantly seek accurate methods to predict stock prices, in order to capitalize on market fluctuations and make informed investment decisions. Stock prediction can be used to make informed investment decisions, such as when to buy or sell a stock. Traditional machine learning models, such as linear regression and random forests [2], have been used for stock prediction. However, these models often suffer from overfitting and are unable to capture long-range dependencies in the data, and struggle to capture the intricate patterns hidden within vast and dynamic financial datasets. Recently, transformer-based models, with their attention mechanism, have achieved state-of-the-art results on a variety of natural language processing and computer vision tasks. Transformers can learn long-range dependencies in sequences of data, making them well-suited for stock prediction. Reinforcement learning (RL) [3] is a machine learning technique that allows agents to learn to behave in an environment by trial and error. RL agents can be trained to make optimal decisions in order to maximize a reward function.

This paper proposes a quantitative trading framework that leverages the attention mechanism of Transformer architectures for efficient sequential data processing and employs RL to make adaptive trading decisions based on market feedback. Combining these two paradigms for stock prediction aims to address the limitations of existing methods.

II. Methodology

In this section, we present a formal mathematical framework for the proposed quantitative trading model. Our proposed approach leverages the power of Transformer models for efficient sequential data processing, coupled with an RL agent for adaptive decision-making in the context of quantitative trading. We modeled our environment in a formal representation using the

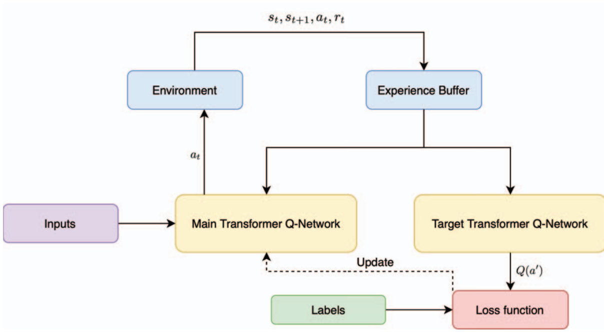

Fig. 1: Proposed quantitative trading framework.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00261

Markov decision process (MDP), specified by $M = (S, A, R)$, where $s \in S$ represents the state space, $a \in A$ denotes the action space, $R(s_t, a_t)$ is the reward function (*ref.* (3)). The fundamental concept is to predict the closing price, given the previous historical sequences of closing prices. Labels representing *buy*, *sell*, or *hold* actions were generated based on a predefined 3-day prediction formula [4] $3DPred$ (1),

$$3DPred = 0.4 \times \frac{closeprice(d-2) - closeprice(n-3)}{closeprice(d-3)} + 0.32 \times \frac{closeprice(d-1) - closeprice(d-2)}{closeprice(d-2)} + 0.28 \times \frac{closeprice(d) - closeprice(d-1)}{closeprice(d-1)}. \quad (1)$$

It uses historical price data to predict future price movements and categorizes them into actionable labels where d is the index of the closing price of the current day.

Figure 1 depicts the proposed quantitative trading framework that combines both RL and Transformer networks. An encoder-based transformer architecture was employed to model the historical price sequence and extract relevant features. The Transformer captured long-range dependencies and complex relationships within the data [5], providing a richer representation for the Q-Network. The state s_t fed into the Q-Network consisted of three key characteristics S: the closing price of the ticker, the total number of shares owned and the available balance, which is the number of liquid assets that are accessible to be used in the process of purchasing or selling the stock at each successive time t step of the process. Target Q-values were calculated based on the predicted future reward r_t and the discounted Q-value of the next state s_{t+1}. The agent then chooses the required action a_t that maximizes long-term rewards, from a discrete number of action sets A: *buy* a stock, *sell* a stock S, or *hold*. The Q network was trained by minimizing the loss between the labels of (1) and the target Q values. Based on the Q-values for *buy*, *sell*, and *hold* actions, the agent employed an epsilon-greedy policy to select an action, favoring exploration in the early stages and exploitation later. Our goal is to encourage profits over the course of time (Earnings);

$$\text{Earnings} = \text{Balance} + \text{Volume} \times P_t - \text{InitialAmount}, \quad (2)$$

where $Balance$ is the remaining amount after trading the stock, $Volume$ is the amount of shares bought, P_t is the price of the stock, and $InitialAmount$ represents the initial amount used, and the respective reward is,

$$R(s_t, a_t) = \begin{cases} \frac{\text{Earnings}}{\text{InitialAmount}} & \text{, if Earnings} > 0 \\ -0.1, & \text{otherwise.} \end{cases} \quad (3)$$

III. Experimental Setup, Results, and Discussion

We conducted experiments using historical stock market data, comparing the performance of our proposed model with traditional methods that include a state-of-the-art (SOTA) machine learning technique, *i.e.* XGBoost, and another SOTA RL-based model, *i.e.* Deep Q-Network (DQN). The training process was conducted on a single node HPC[1]. The data used in this work was obtained from Yahoo Finance using a Python script. It comprises the stock prices of Apple Inc. Table 2 summarizes the experimental results in terms of average reward for RL-based approaches and accuracy. The experimental results indicate that our proposed model outperforms both baseline methods in stock market prediction tasks. The source code for the model implementation is available on GitHub for reproducibility[2].

Table. 2: Performance evaluation comparison

Model	Avg Reward	Accuracy
XGBoost	NA	73
DQN	383.5	43
Proposed Model	**2500**	**76**

IV. Conclusion

This paper proposed a hybrid quantitative trading method integrating a Transformer-based Q-Network with Reinforcement Learning for stock prediction. The proposed method leveraged the strengths of both techniques: Transformers effectively captured long-range dependencies and complex relationships within historical price data, providing a richer representation for the Q-Network, and Reinforcement Learning which enabled the agent to dynamically adjust its investment strategies based on the reward function. The experimental results suggest that the proposed model improves prediction accuracy compared to existing methods. The proposed framework presents a promising solution to the challenges posed by the dynamic nature of financial markets and can be used as an integrated tool for real-world applications in portfolio management.

References

[1] Zhang, W. and Skiena, S. 2010. Trading Strategies to Exploit Blog and News Sentiment. Proceedings of the Int AAAI Conf on Web and Social Media. 4, 1 2010, 375-378.

[2] H. Zhang, "Stock Price Prediction using Linear Regression and LSTM Neural Network," in 2022 Int. Conf. on Machine Learning and Intelligent Systems Engineering, 2022 pp. 182-185.

[3] Zhang, Zihao, et al. "Deep Reinforcement Learning for Trading." J of Financial Data Science, vol. 2, no. 2, Mar. 2020, pp. 25–40.

[4] Fu, Kui, Yu, Yidong, Li, Bing. (2023). Multi-Feature Supervised Reinforcement Learning for Stock Trading. IEEE Access. PP. 1-1.

[5] R. Rizk, D. Rizk, F. Rizk, A. Kumar and M. Bayoumi, "A Resource-Saving Energy-Efficient Reconfigurable Hardware Accelerator for BERT-based Deep Neural Network Language Models using FFT Multiplication," IEEE Int Sympo on Circuits and Systems, 2022, pp. 1675-1679,

[1]Computation was performed on Lawrence Supercomputer at University of South Dakota awarded by NSF.1626516.

[2]https://github.com/2ai-lab/quantitative-trading-rl