Advanced Factor Analysis and Ensemble Learning of RL Algorithm

Laxmi Tiwari
Advance AI and RL
Logictronix Technologies
Lalitpur, Nepal
laxmi@logictronix.com

Krishna Gaihre
Advance AI and RL
Logictronix Technologies
Lalitpur, Nepal
krishna@logictronix.com

***Abstract*—**This study explores advanced factor analysis techniques for reinforcement learning (RL)-based trading, focusing on order book dynamics, market impact metrics, volatility, and market regime detection. Traditional OHLCV-based approaches often fail to capture hidden liquidity patterns and execution constraints, leading to suboptimal decision-making. We introduce key factors such as Order Book Imbalance (OBI), Spread-Based Liquidity, Order Flow Imbalance (OFI), Price Impact of Trades, and Market-to-Limit Order Ratio (MLR) to enhance market state representation. Additionally, intraday volatility, momentum scores, and noise-to-signal ratio are incorporated to distinguish between trend-following and mean-reverting environments. As market data is POMDP, for generalization we stack previous information and also find the relative information of the market uncertainty.We employ model-free algorithm ensembles of PPO, A2C, Reinforce and TwinD3QN to dynamically optimize cryptocurrency portfolios. Our findings suggest that integrating these factors significantly improves trading performance by enhancing state representations and optimizing decision-making in high-frequency financial environments.

Keywords—RL, POMDP, Diluted CNN, PCA, XRL, Transformer

I. Introduction

With the advent of AI, financial markets have seen a transformation in data-driven decision-making. From algorithmic trading to risk management, AI-powered models have enhanced efficiency and accuracy. In particular, quantitative trading (QT) has evolved with machine learning and deep learning, enabling automated strategy execution. However, traditional methods often struggle to generalize in volatile and partially observable markets. Reinforcement Learning (RL) offers a promising alternative by dynamically adapting to market changes, optimizing strategies, and maximizing financial returns.

Traditional RL-based financial models primarily rely on OHLCV (Open, High, Low, Close, Volume) data, which fails to capture the complexities of market microstructure. This limitation results in suboptimal decision-making, especially in high-frequency trading (HFT) environments. To address this, we model QT as a Partially Observable Markov Decision Process (POMDP), recognizing that traders lack full visibility of competitor actions and market dynamics. By incorporating advanced factor analysis, we enhance the agent's market understanding, improving decision-making and trading performance.

II. Related work

In RL, we have to design our state to follow Markov Decision Process (MDP) frameworks, which real-world financial markets are inherently noisy and unpredictable, making Partially Observable Markov Decision Processes (POMDPs) a more suitable approach. Prior works have shown promising result in limit order book(LOB) environment for tasks such as market making and trade execution, supporting the design of our purposed ensemble-based model combines LSTM+GRU, Diluted CNN and Transformer architecture. Gašperov and Kostanjčar [1][8] demonstrated that DRL agents could effectively operate within Hawkes process-based LOB models, while Guo et al. [2] reinforced the idea of directly learning from raw LOB data, validating our use of both convolutional and attention-based architectures for pattern extraction. Nagy et al. [3][6] introduced asynchronous deep double dueling Q-learning to improve training stability and performance in LOB trading, a direction aligned with our ensemble policy strategy for robustness. The benefits of multi-agent learning and policy heterogeneity were emphasized by Zhang et al. [4] and Karpe et al. [7], directly supporting our use of multiple DRL models that collaboratively predict through majority voting. Further, Jerome et al. [5] and Wang et al. [9] provided frameworks for model-based LOB trading and scalable FinRL environments, highlighting the importance of combining domain-specific feature representations and parallel policy training. Altogether, these studies substantiate our ensemble learning framework that fuses diverse neural representations into a unified, resilient prediction mechanism for decision-making in LOB-based trading. Despite these advancements, reinforcement learning (RL)-driven trading systems often suffer from a lack of interpretability, which has led to the adoption of explainability methods such as SHapley Additive exPlanations (SHAP). Our work builds upon these studies by formulating QT as a POMDP, integrating advanced factor-based signals, employing an ensemble of RL algorithms (PPO, A2C, REINFORCE and TwinD3QN) with maximum voting, and incorporating explainability techniques to enhance model transparency and decision-making insights.

III. Methodology

A. State Representation

Our dataset consists of high-resolution, second-wise order book data, providing a detailed view of market microstructure dynamics. This rich dataset serves as a foundation for understanding market direction, supplemented by advanced factor-based insights that enhance predictive capabilities.

Key market microstructure features include:

- **Order Book Imbalance (OBI):** Quantifies the difference between bid and ask orders, indicating directional market pressure.
- **Spread-Based Liquidity:** Captures bid-ask spread fluctuations to assess liquidity conditions.
- **Order Flow Imbalance (OFI):** Analyzes order placements, cancellations, and executions to reveal supply-demand dynamics.
- **Price Impact of Trades:** Measures how executed trades influence future price movements.
- **Market-to-Limit Order Ratio (MLR):** Evaluates the ratio of market orders to limit orders, reflecting execution aggressiveness.

Beyond market microstructure, we incorporate additional features to refine the agent's understanding:

- **Intraday Volatility:** Tracks price fluctuations across different time intervals.
- **Momentum Scores:** Identifies potential trends based on recent price movements.
- **Noise-to-Signal Ratio:** Differentiates meaningful price signals from random fluctuations.

To further optimize feature selection, we design advance Factors above 500, which includes the market liquidity, volatility, LOB 15 level data with leveraging Principal Component Analysis (PCA) to extract the top 200 with 96 % information. Additionally, we incorporate essential financial metrics such as balance, profit, and shares, ensuring a holistic view of market conditions. By combining historical data with these enriched features, we provide the RL agent with a deeper, more structured market representation, enhancing its ability to make informed and confident trading decisions.

IV. Algorithm

A. *Supervised Learning*

To enhance the performance and robustness of our reinforcement learning (RL) models, we integrate a supervised learning approach, particularly useful for testing and creating sequential data. This approach, inspired by the FinRL Starter Kit, leverages historical market data for training RL agents in a supervised learning setup before applying them in reinforcement learning environments. Specifically, sequential data, which is critical for understanding temporal dependencies in financial markets, is constructed by aligning historical price actions and order book data into consistent time series. This structure enables the RL agent to learn from past data in a supervised fashion, mimicking real trading scenarios where decisions are often made based on historical trends and patterns.

In the FinRL Task -2 Competition, supervised learning is employed to pre-train agents on sequential data, where the model predicts future price movements or market events based on previous observations. By initially training on this labeled data, the agent learns key patterns that define market behaviors. This pre-training step helps the model better generalize when transitioning to reinforcement learning, reducing the time required for the agent to adapt to more complex and noisy real-time trading conditions. The generated sequential data also provides a basis for understanding how past price and order book dynamics affect future trading actions. This hybrid approach, combining supervised learning with RL, facilitates a more robust model that is capable of making informed decisions based on both historical data and real-time market conditions.

B. *Reinforcement Learning Framework*

1. Policy

   Financial markets, particularly Limit Order Book (LOB) data, exhibit complex time-series patterns, uncertainty, and rapid fluctuations. To effectively model market depth and price movements, we integrate multiple architectures into our reinforcement learning (RL) policy. As a baseline, we use LSTM and GRU, which capture sequential dependencies and evolving market trends. Beyond these recurrent models, we explore advanced architectures such as Transformers for their self-attention mechanisms and Dilated CNNs, which enhance feature extraction by capturing multi-scale dependencies. Inspired by AlphaZero's structured decision-making with ResNet, we design an optimized stacking mechanism tailored for trading. The Dilated CNN framework, widely used in time-series applications, provides additional market insights by expanding the receptive field without excessive computational overhead, improving the agent's ability to detect latent trading opportunities.

2. Algorithm

   We adopt a **model-free reinforcement learning (RL) approach**, leveraging an ensemble of deep RL algorithms to ensure adaptability, stability, and robustness in volatile crypto markets:

   1. **Twin Delayed D3QN (TwinD3QN):** Reduces overestimation bias through clipped double-Q learning, improving trade execution in noisy LOB environments.
   2. **Proximal Policy Optimization (PPO):** Maintains policy stability by preventing overly aggressive updates, ensuring resilience to sudden market regime shifts and flash crashes.
   3. **Advantage Actor-Critic (A2C):** Utilizes parallelized learning to efficiently process high-frequency LOB updates across multiple crypto pairs.
   4. **REINFORCE:** Employs Monte Carlo policy gradient updates to capture long-term liquidity dynamics, benefiting market-making strategies.

C. *Explainable RL*

To achieve interpretability, we employ **S**HAP (Shapley Additive Explanations**)**, a popular method for feature importance in machine learning. SHAP values allow us to decompose the agent's decisions by evaluating the contribution of each feature (e.g., Order Book Imbalance, Spread-Based Liquidity, Order Flow Imbalance) to the final action, such as buying, selling, or holding. This transparency is essential for understanding how specific market conditions, like liquidity pressure or momentum shifts, influence trading decisions. Additionally, we leverage saliency maps in CNN models, which visualize which portions of the input data (such as bid-ask spreads or order flow dynamics) are most influential in driving the agent's decisions.

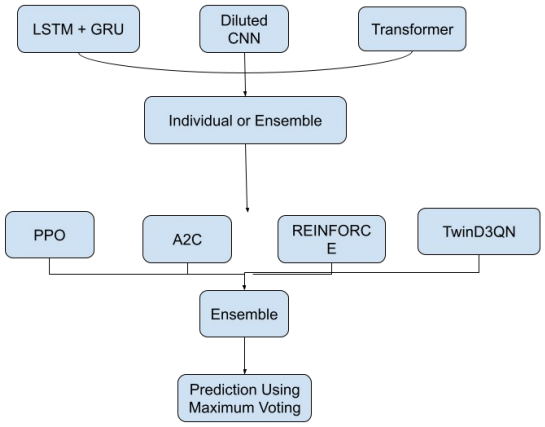

Fig: Architecture of RL Algorithm

This ensemble-based voting mechanism uniquely strengthens RL-driven trading by combining the precision of TwinD3QN, the stability of PPO, the scalability of A2C, and the long-term adaptability of REINFORCE. Also ensemble of Policy LSTM, Transformer, Diluted CNN added a super generalize agent. These algorithms complement each other, overcoming the partial observability, adversarial conditions, and noise prevalent in crypto markets. By transforming raw LOB data into structured, adaptive trading signals, our framework enables a more intelligent and resilient trading agent, capable of navigating complex market dynamics with confidence. Given in Tradesimulator uses a profit and stop loss information as reward, we have added additional reward like looking forward to analysis how the agent is actually taking action. IN addition we implement explainable RL to make

more clear why this action is taking, it has several benefit while training and evaluation. By making the agent's reasoning interpretable, we not only improve decision confidence but also enable easier detection of potential flaws or biases in the model, further enhancing the robustness of the trading system.

CONCLUTSION

In conclusion, our approach significantly enhances trading performance by integrating advanced market factors and employing a robust ensemble of policy and model-free RL algorithms. This methodology allows RL agents to gain a deeper understanding of market dynamics, resulting in more adaptive and profitable trading strategies in high-frequency environments. However, there are limitations, including the absence of sentiment analysis, which could be addressed by incorporating Large Language Models (LLMs) to factor in market news and trader sentiment. Additionally, the potential for further improvements lies in incorporating Reinforcement Learning from Human Feedback (RLHF) and Reinforcement Learning from AI Feedback (RLAIF), which could refine decision-making. Future work also includes exploring multi-agent trading strategies to enhance agent cooperation and competition, thereby advancing the adaptability and robustness of trading systems in volatile markets.

REFERENCES

[1] G. Gašperov and T. Kostanjčar, "Deep reinforcement learning for market making under a Hawkes process-based limit order book model," arXiv preprint arXiv:2207.09951, 2022. [Online]. Available: https://arxiv.org/abs/2207.09951

[2] L. Guo, Y. Lin, and Y. Huang, "Market making with deep reinforcement learning from limit order books," *arXiv preprint arXiv:2305.15821*, 2023. [Online]. Available: https://arxiv.org/abs/2305.15821

[3] M. Nagy, N. Calliess, and R. Zohren, "Asynchronous deep double dueling Q-learning for trading-signal execution in limit order book markets," *arXiv preprint arXiv:2301.08688*, 2023. [Online]. Available: https://arxiv.org/abs/2301.08688

[4] C. Zhang, Y. Zhang, and J. Ding, "Multi-agent reinforcement learning for market-making with limit order book," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 7, pp. 2303-2312, July 2023. DOI: 10.1109/TNNLS.2023.3156825

[5] Joseph Jerome, Leandro Sánchez-Betancourt, Rahul Savani, and Martin Herdegen, "mbt-gym: Reinforcement learning for model-based limit order book trading," *in Proc. of ICAIF*, 2023. [Online]. Available: https://dl.acm.org/doi/fullHtml/10.1145/3604237.3626873

[6] Peer Nagy, Jan-Peter Calliess, and Stefan Zohren, "Asynchronous Deep Double Duelling Q-Learning for Trading-Signal Execution in Limit Order Book Markets," *arXiv preprint arXiv:2301.08688*, 2023. [Online]. Available: https://arxiv.org/abs/2301.08688

[7] Michaël Karpe, Jin Fang, Zhongyao Ma, and Chen Wang, "Multi-Agent Reinforcement Learning in a Realistic Limit Order Book Market Simulation," *in Proc. of ICAIF*, 2020. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/3383455.3422570

[8] Bruno Gašperov and Zvonko Kostanjčar, "Deep Reinforcement Learning for Market Making Under a Hawkes Process-Based Limit Order Book Model," *arXiv preprint arXiv:2207.09951*, 2022. [Online]. Available: https://arxiv.org/abs/2207.09951

[9] Keyi Wang, Kairong Xiao, and Xiao-Yang Liu Yanglet, "Parallel Market Environments for FinRL Contests," *arXiv preprint* arXiv:2504.02281, 2025. [Online]. Available: https://arxiv.org/abs/2504.02281