Deep Reinforcement Learning for Quantitative Trading: Challenges and Opportunities

Bo An, Shuo Sun, and Rundong Wang, *Nanyang Technological University, Singapore, 639798*

Quantitative trading (QT) has been a popular topic in both academia and the financial industry since the 1970s. In the last decade, deep reinforcement learning (DRL) has garnered significant research interest with stellar performance in solving complex sequential decision-making problems, such as Go and video games. The impact of DRL is pervasive, recently demonstrating its ability to conquer some challenging QT tasks. In this article, we outline several key challenges and opportunities that manifest in DRL-based QT to shed light on future research in this field.

The financial market, which attracts hundreds of millions of investors, is an ecosystem involving over $90 trillion[a] market capitalization globally in 2021. In the last decade, we have witnessed the rise of quantitative trading (QT), which refers to automatically generating trading signals with data-driven techniques, due to its instant and accurate order execution (OE), and increased capability to analyze large amounts of financial market data. QT has become ubiquitous across different financial markets and currently accounts for more than 70% and 40% of the trading volume in developed markets (e.g., USA) and emerging markets (e.g., China),[b] respectively. Traditional QT methods are based on either heuristic rules[1,2] or human-directed predictive algorithms.[3,4] However, due to the high volatility and noisy nature of the financial market, their performance is not stable and highly dependent on market conditions. To tackle these issues, deep reinforcement learning (DRL) has emerged as a promising approach for QT by training end-to-end agents for profitable trading decisions. In this article, we first briefly introduce the background of QT and the most notable DRL-based QT methods. Then, we discuss current challenges and point out future research directions in this field.

[a][Online]. Available: https://data.worldbank.org/indicator/CM.MKT.LCAP.CD/
[b][Online]. Available: https://therobusttrader.com/what-percentage-of-trading-is-algorithmic/

1541-1672 © 2022 IEEE
Digital Object Identifier 10.1109/MIS.2022.3165994
Date of current version 20 May 2022.

QT PROBLEMS

The mainstream QT tasks can generally be categorized into macrolevel tasks and microlevel tasks, as shown in Figure 1. Algorithmic trading (AT) where traders consistently buy and sell one given financial asset to make profits. It is widely applied in the trading of various financial assets, such as stocks, cryptocurrencies, and foreign exchanges. Time is divided into discrete time steps. Traders are allocated some cash at the beginning of a trading period. Later on, at each time step t, traders have the option to buy, hold, or sell some amount of shares for changing positions with a goal is to maximize the final net value at the end of the trading period.

Portfolio management (PM) is a fundamental QT task in which investors hold a number of financial assets and reallocate them periodically to maximize long-term profit. At the beginning of a holding period, the agent holds a portfolio that consists of preselected financial assets with various weights. With price fluctuations in the market, the portfolio manager can build the new portfolio based on market status and personal risk preference.

OE focuses on executing an order of liquidation in a fixed period of time. The objectives of OE is to fulfill the whole order with lower cost. The major challenges of

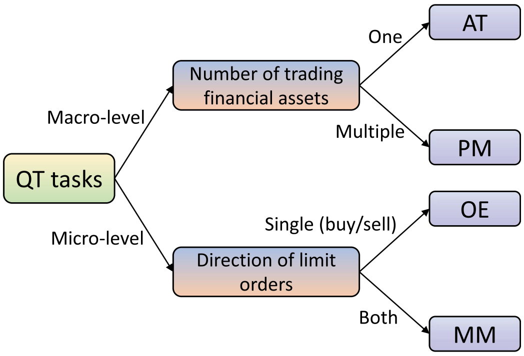

FIGURE 1. Relationships between QT tasks.

designing a good OE strategy are twofold: 1) avoiding harmful market impact caused by large transactions in a short period and 2) restraining price risk, which means missing good trading windows due to slow execution.

Market making (MM) refers to the trading activities that continually quote prices at which market makers are willing to trade on both the buy and sell side for one financial asset. They provide liquidity and make profits from the tiny price spread between buy and sell orders.

DRL FOR QT

In general, QT tasks focus on maximizing long-term profit under a certain risk tolerance. Traditional QT strategies[1,2] discover trading opportunities based on heuristic rules with the knowledge of financial experts. However, rule-based methods exhibit poor generalization ability and only perform well in certain market conditions. Another direction is to generate trading signals based on financial predictions. In the literature, there are attempts using supervised learn ing methods, such as gradient boosting tree models[3] and deep neural networks[4] for financial prediction. However, there is an unignorable gap between prediction signals and profitable trading actions due to the high volatility and noisy nature of the financial market.

For designing profitable QT strategies, the advantages of DRL methods are fourfold: 1) DRL allows training an end-to-end agent, which takes available market information as input state and outputs trading actions directly. 2) DRL-based methods optimize overall profit directly and bypass financial prediction as the intermediate task. 3) DRL can directly incorporate task-specific constraints (e.g., transaction cost and slippage) into the training process. 4) DRL methods have the potential to be generalized to any market condition.

Recently, there has been an increasing number of research works on DRL-based QT (see Sun *et al.*'s work[5] for a comprehensive survey). For AT, iRDPG[6] is proposed with GRU layers to learn recurrent market embedding and behavior cloning to mimic the trading actions of human experts. For PM, a hierarchical DRL framework[7] is proposed to account for the limited data and high-dimensional sequential decision-making task. DeepTrader[8] focuses on learning risk-adjusted portfolios with a novel reward function and an asset scoring unit. For OE, a model-free DRL framework[9] is proposed with a novel policy distillation mechanism to bridge the gap between the noisy yet imperfect market states and the optimal action sequences for order execution. For MM, a game-theoretic framework based on adversarial DRL learning[10] is proposed as an adaptation of the traditional mathematical MM models.

CHALLENGES AND OPPORTUNITIES

Advanced DRL Techniques on QT

Existing work on DRL-based QT simply applies classic DRL algorithms on different QT scenarios. It is a promising research direction to explore the effectiveness of more advanced DRL techniques on financial data. First, a major challenge on designing profitable DRL-based QT algorithms is data scarcity. Model-based DRL can address this challenge by learning a financial market model[11] to speed up the training process. While maximizing the accumulated reward, the worst case (e.g., financial crisis) can be used as a regularizer. Second, the main goal of different QT tasks is to maintain a good balance between maximizing profit and minimizing risk. Multiobjective DRL techniques provide a weapon to train diversified trading policies with adaptive risk tolerance. Third, graph learning[12] achieves promising results in modeling the interrelationship between stocks in prediction-based methods.[13] Combining graph learning with DRL is also a potential direction. Fourth, current DRL-based QT methods exhibit a poor generalization ability due to the severe distribution shift of the financial market. Meta-RL and transfer learning techniques can help to learn robust trading policies across different financial asset types and markets. Fifth, explainability is of vital importance for high-risk decision-making tasks, such as QT. Hierarchical DRL methods decompose the main goal into subgoals for low-level DRL agents. By learning the optimal subgoals for the low-level agent, the high-level agent forms a representation of the financial market that is interpretable by human experts. Sixth, it is extremely risky to train DRL agents by directly interacting with the real market for QT. Offline DRL, where only historical data are used for training, has the potential to model the distribution shift and risk of the financial market.

New QT Settings

There are still some important QT settings, such as high-frequency trading and pairs trading have not yet been explored. Intraday trading tries to capture the fleeting trading opportunities within the same trading day; high-frequency trading aims at capturing the fleeting microlevel trading opportunities; pairs trading focuses on analyzing the relative trend of two highly correlated financial assets. It is a promising research direction to design the DRL-based methods that incorporate finance knowledge to fit the characteristics of different QT scenarios.

Enhance With Auto-ML

Due to the noisy nature of financial data and the brittleness of DRL methods, the success of DRL-based QT models highly relies on carefully designed DRL components and properly tuned hyperparameters. Auto-ML, which tries to design high-quality machine learning applications automatically, is becoming more and more popular. Auto-ML techniques, such as automatic feature selection, hyperparameter tuning, and neural architecture search can significantly improve the efficiency of designing DRL-based QT methods and make them easy to use for people without in-depth knowledge of DRL.

Toward Realistic Market Simulation

High-fidelity simulation is the key foundation of DRL methods' success. Although existing work takes into account many practical constraints, such as transaction fee, execution cost, and slippage, there is a long way to go to provide realistic financial market simulation due to the ignorance of ubiquitous market impact. Even though there are some efforts to model the impact of the market,[10,14] building high-fidelity market simulators is still a very challenging task.

Unified and Systematic Evaluation

As for the evaluation of DRL-based QT methods, most existing work arbitrarily select a few baselines and datasets, which might lead to inconsistent revenue reporting. There is no broad consensus on the general ranking of DRL-based methods for QT tasks, making it extremely challenging to benchmark new DRL algorithms in this field. The build of a platform with a suite of standardized evaluation datasets and the implementation of SOTA methods is in high demand. As the evaluation criteria, it is necessary to test DRL algorithms on multiple financial assets across different markets for the evaluation of robustness and generalizability. We also note that the split of training, validation, and test set in most QT papers is quite random. However, it is better to split data on a rolling basis due to the significant distribution shift among time in the financial market. In addition, time for tuning hyperparameters on baselines and their own methods should be roughly the same for a more reliable evaluation of DRL-based QT methods.

CONCLUSION

In this article, we first introduce mainstream QT tasks. Then, we discuss the most notable DRL-based QT methods and their advantages over traditional methods. Finally, we point out some of the challenges and promising future opportunities. Both DRL and QT are ongoing hot research topics in the past few decades. There are many emerging techniques and models each year.

REFERENCES

1. N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *J. Finance*, vol. 48, no. 1, pp. 65–91, 1993.
2. T. J. Moskowitz, Y. H. Ooi, and L. H. Pedersen, "Time series momentum," *J. Financial Econ.*, vol. 104, no. 2, pp. 228–250, 2012.
3. G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. 30th Neural Inf. Process. Syst.*, 2017, pp. 3146–3154.
4. Y. Xu and S. B. Cohen, "Stock movement prediction from tweets and historical prices," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 1970–1979.
5. S. Sun, R. Wang, and B. An, "Reinforcement learning for quantitative trading," 2021, *arXiv:2109.13851*.
6. Y. Liu, Q. Liu, H. Zhao, Z. Pan, and C. Liu, "Adaptive quantitative trading: An imitative deep reinforcement learning approach," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2020, pp. 2128–2135.
7. R. Wang, H. Wei, B. An, Z. Feng, and J. Yao, "Commission fee is not enough: A hierarchical reinforced framework for portfolio management," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 626–633.
8. Z. Wang, B. Huang, S. Tu, K. Zhang, and L. Xu, "DeepTrader: A deep reinforcement learning approach to risk-return balanced portfolio management with market conditions embedding," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 643–650.
9. Y. Fang *et al.*, "Universal trading for order execution with oracle policy distillation," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 107–115.

10. T. Spooner and R. Savani, "Robust market making via adversarial reinforcement learning," 2020, *arXiv:2003.01820*.
11. P. Yu, J. S. Lee, I. Kulyatin, Z. Shi, and S. Dasgupta, "Model-based deep reinforcement learning for dynamic portfolio optimization," 2019, *arXiv:1901.08740*.
12. Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.
13. R. Sawhney, S. Agarwal, A. Wadhwa, and R. Shah, "Exploring the scale-free nature of stock markets: Hyperbolic graph learning for algorithmic trading," in *Proc. Web Conf.*, 2021, pp. 11–22.
14. S. Vyetrenko *et al.*, "Get real: Realism metrics for robust limit order book market simulations," 2019, *arXiv:1912.04941*.

BO AN is an associate professor at Nanyang Technological University, Singapore, 639798. Contact him at boan@ntu.edu.sg.

SHUO SUN is a Ph.D. student at Nanyang Technological University, Singapore, 639798. Contact him at shuo003@e.ntu.edu.sg.

RUNDONG WANG is a Ph.D. student at Nanyang Technological University, Singapore, 639798. Contact him at rundong001@e.ntu.edu.sg.