Trends in Reinforcement Learning Methods for Stock Prediction

Jaehyun Chung†, Hyunseok Choi†, Seokhyeon Min†, Soohyun Park§, and Joongheon Kim†
†Department of Electrical and Computer Engineering, Korea University, Seoul, Republic of Korea
§Division of Computer Science, Sookmyung Women's University, Seoul, Republic of Korea
E-mail: rupang1234@korea.ac.kr, sos153661@gmail.com, msh981117@gmail.com,
soohyun.park@sookmyung.ac.kr, joongheon@korea.ac.kr

***Abstract*—A popular and lucrative area of research has always been stock prediction. Stock prediction using traditional deep learning has been proven to provide better accuracy and returns. However, as artificial intelligence developed, the idea of reinforcement learning (RL) emerged. The rise of RL in the financial markets is fueled by a number of benefits that are specific to this area of artificial intelligence (AI). RL, in particular, enables the combination of the "prediction" and "portfolio construction" activities into a single integrated step, allowing machine learning challenges to be precisely customized to investors' objectives. Conveniently, significant limitations like transaction costs, market liquidity, and investor risk aversion can be considered simultaneously. Despite the fact that supervised learning techniques continue to receive the majority of attention, the RL research community has achieved great strides in the financial field during the previous several years. This paper introduces the overall concepts and applications of RL and stock prediction. Additionally, current technology trends are presented based on several application domains. In summary, this paper explores RL-based research trends in the field of stock prediction and makes suggestions for future research avenues.**

I. Introduction

Stock market prediction has been an important research topic among researchers and investors for a long time. Initially, traditional deep learning techniques have proven to be highly accurate and profitable in stock prediction, but with the recent development of artificial intelligence, RL is opening up new possibilities in the financial field [1]. RL has the ability to effectively respond to market uncertainty and complexity, showing its strengths in establishing more sophisticated and flexible investment strategies in a changing environment [2]. In particular, RL can comprehensively consider various factors such as transaction costs, market liquidity, and investors' risk preferences, making it possible to construct customized strategies that reflect investment goals and constraints [3]. The use of RL in financial markets has grown rapidly in recent years, making RL-based stock prediction a powerful tool that can complement or replace existing traditional prediction methods [4]. This paper presents a general overview of how these RL-based algorithms are applied to stock prediction and analyzes their advantages and limitations through examples of RL use in the stock market. Lastly, an overview of RL-based algorithms and their use in the field of stock prediction is suggested in this paper.

The remainder of the paper is structured as follows. Sec. II presents the fundamental features of portfolio theory and RL. Sec. III examines how RL-based stock forecasting research is progressing. Lastly, Sec. IV wraps up the work and offers suggestions for further investigation.

II. Preliminaries

A. Portfolio Theory

Portfolio theory is a financial strategy that combines assets with various risk characteristics into a diversified portfolio to minimize risk or maximize profit [5]. This theory is primarily based on the work of Harry Markowitz, and its core concepts include diversification, efficient frontier, risk-return tradeoff, efficient portfolio, market portfolio, and the capital asset pricing model (CAPM) [6]. Diversification is an approach to reducing the risk of individual assets by investing in multiple assets when constructing a portfolio. It can reduce the volatility of your overall portfolio as your assets react to different economic factors [7]. The effective frontier visualizes the return and risk of a portfolio based on a combination of assets and includes portfolios that seek the maximum return at a given level of risk or the minimum risk at a given level of return. The risk-return tradeoff is the concept that you must give up a certain level of return to reduce the risk in your portfolio. The efficient portfolio lies along the effective frontier, providing either a higher return for the same level of risk or a lower risk for the same level of return than other portfolios. The market portfolio represents the entire asset market and includes all assets in proportion to the market [8]. CAPM is based on portfolio theory and describes the relationship between the expected return of an asset and its risk, assuming that the return of an asset is determined by the risk-free rate and the excess return of the market portfolio. Typically, the CAPM is stated as,

$$ER_i = R_f + \beta_i(ER_m - R_f), \tag{1}$$

where ER_i is the expected return on investment, R_f is the risk-free rate, β_i is a measure of the sensitivity of an asset to the movements of the overall market, and ER_m is the market return, respectively [9]. These principles provide the foundation of modern financial theory and provide practical help to investors in constructing the optimal portfolio to suit their investment goals [10].

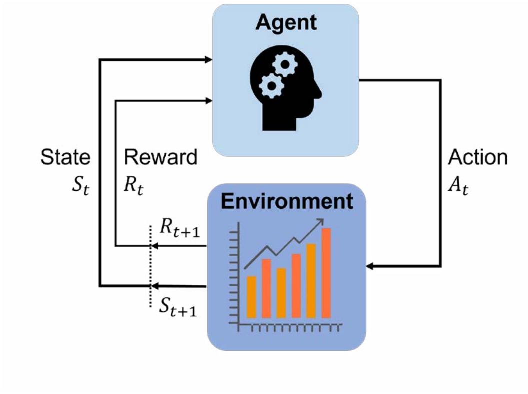

Fig. 1. The overall structure of RL.

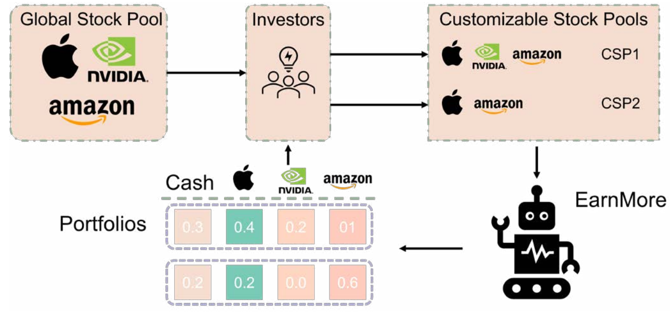

Fig. 2. EarnMore's portfolio management overview for CSPs.

B. Reinforcement Learning

RL, a branch of machine learning, focuses on how an agent interacts with its environment to maximize its performance for a given task [11]. The RL method is expressed in terms of a Markov decision process (MDP), which is a mathematical framework introduced by $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$ for solving problems requiring sequential decision-making [12]. Here, $\mathcal{S}$, $\mathcal{A}$, $\mathcal{P}$, $\mathcal{R}$, and γ represent the state set, action set, transition probability matrix, reward function, and discount factor, respectively [13]. Assuming T is the total time step, each agent aims to maximize its total expected return $G_t = \sum_{k=0}^{T} \gamma^k \mathcal{R}_k$. The probability of acting in a given state s, denoted as π, can be written as follows: $\pi(a|s) = P[\mathcal{S}_t = s | \mathcal{A}_t = a]$. By applying the Bellman equation, the optimal policy π^* may be found as follows: $Q^*(s, a) = \mathcal{R} + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}^a_{ss'} \max_{a'} * [Q^*(s', a')]$, where $Q(s, a)$ is the state-action value function [14]. The optimal policy π^* can be directly found by using the formula $\pi^* = \underset{a}{\mathrm{argmax}} Q^*(s, a)$ once $Q^*(s, a)$ has been acquired through interactions. Fig. 1 presents the general architecture of RL [15].

III. Reinforcement Learning-based Stock Prediction

As the use of AI technology has increased rapidly in the financial market in recent years, various algorithms and models are being introduced in the field of stock prediction [16]. Previously, machine learning was introduced to predict stocks. However, this method had limitations such as overfitting the data and not being able to sufficiently reflect the complex nonlinearity of the market. To overcome these limitations, stock prediction methods using RL have recently been attracting attention. RL allows agents to learn optimal policies through interaction with the environment and has the potential to effectively model the complex characteristics of the stock market. These characteristics of RL are considered a promising forecasting tool in volatile and complex stock markets, and research on this is actively underway.

[17] proposes the cascaded long short-term memory (CLSTM-PPO) model to address performance limitations in deep reinforcement learning (DRL)-based stock trading strategies due to low signal-to-noise ratio and imbalance in financial data. The proposed model adopts a cascade structure using a two-stage deep long short-term memory (LSTM) network to capture hidden information in daily stock data. This study conducts experiments on four major stock indices: the Dow Jones Industrial Average (DJI) of the United States, the Shanghai Stock Exchange 50 (SSE50) of China, the S&P BSE Sensex Index (SENSEX) of India, and the FTSE100 Index of the United Kingdom.

In [18], a new model for stock price prediction in the oil and gas sector is proposed and verified. This study proposes the DLQL (Deep Long Short-Term Memory Q-Learning) model and the DLAQL (Deep Long Short-Term Memory Attention Q-Learning) model, with the goal of improving the accuracy of stock price predictions by comparing them with the LSTM model. Additionally, the proposed model is created and verified using historical stock price data of Cenovus Energy Inc. (CVE), MPLX LP (MPLX), Cheniere Energy Inc. (LNG), and Suncor Energy Inc. (SU).

For managing customized stock pools (CSPs), [19] proposes EarnMore, an RL system that uses maskable stock representation to manage portfolio management (PM) with CSPs via one-shot training in a global stock pool (GSP). This model can solve the problem of traditional RL methods, which require retraining RL agents even when the stock pool is slightly changed, resulting in high computational cost and performance instability. Comparing EarnMore to 14 cutting-edge baseline models, the study discovered that it provides more than 40% revenue improvement for 6 popular financial KPIs. Fig. 2 shows a schematic diagram of EarnMore's portfolio management in CSPs.

Research has also been conducted to develop a model that integrates graph functions with RL. In [20], the GraphSAGE and DRL coupled model (GRL), which integrates a GraphSAGE-based feature extractor into the existing Proximal Policy Optimization (PPO) model, is proposed to effectively process non-Euclidean relationships in complex financial data. GraphSAGE-based feature extractors capture complex non-Euclidean relationships between market indices, industry indices, and stocks. Through experiments, it can be seen that the

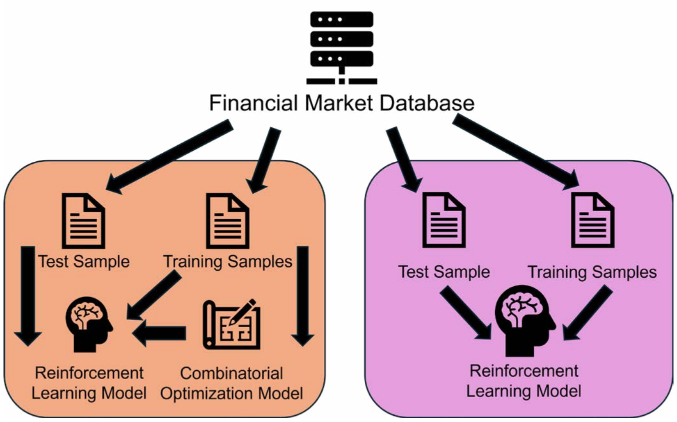

Fig. 3. Overall structure of an approach that integrates a RL model and a combinatorial optimization model.

GRL model shows better performance than the existing PPO model in five indicators: Return on Investment (ROI), Sharpe Ratio, Sortino Ratio, Maximum Drawdown, and Calmar Ratio.

Furthermore, [21] investigates a method that combines combinatorial optimization and RL to enhance financial asset trading. This approach compensates for the lack of objectivity in determining the optimal reward, as RL relies on a subjective reward function. In the combinatorial optimization model, the maximum profit achievable during a given trading period is determined, and based on this information, the RL model adjusts the reward function and learns. An experiment is conducted using stocks of Petrobras, a Brazilian oil company, and the profit maximized through a combinatorial optimization model demonstrates the effectiveness of the proposed approach. Fig. 3 shows the overall structure of the approach that combines an RL model and a combinatorial optimization model.

IV. Conclusion

The fundamental ideas of portfolio theory, RL, and RL-based stock prediction research are presented in this paper. Recently, RL algorithms have been applied to various fields, and research is especially active in the field of stock prediction, where real-time is important. Every scenario requires a different algorithm to be used, as there are evident variations in accuracy and learning speed based on the RL method. RL also has the ability to model complicated environments and create predictions that consider several elements. This means that when applied to the field of stock prediction, RL may learn more effectively by reflecting more variables and interactions. However, RL-based stock prediction has the problem that calculation costs can be high depending on the situation, and learning time can be long. Therefore, research should be focused on improving the learning speed to solve this problem through efficient algorithm design, use of parallel processing technology, etc. Through a work that applies RL to the field of stock prediction, we introduce recent research outcomes in this paper. An overview of portfolio theory, RL, and its applications to stock prediction is given in this document. In closing, this work presents issues that require further improvement in RL and makes future research direction.

V. Acknowledgement

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government [MSIT (Ministry of Science and ICT (Information and Communications Technology))] (RS-2024-00439803, SW Star Lab) for Quantum AI Empowered Second-Life Platform Technology. J. Kim is the corresponding author of this paper(joongheon@korea.ac.kr).

References

[1] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios," *Applied Stochastic Models in Business and Industry*, vol. 33, no. 1, pp. 3–12, October 2017.

[2] T. G. Fischer, "Reinforcement learning in financial markets-a survey," FAU discussion papers in economics, Tech. Rep., 2018.

[3] Y. Gao, K. Y. C. Lui, and P. Hernandez-Leal, "Robust risk-sensitive reinforcement learning agents for trading markets," *CoRR*, vol. abs/2107.08083, July 2021.

[4] J. E. Moody and M. Saffell, "Learning to trade via direct reinforcement," *IEEE Trans. Neural Networks*, vol. 12, no. 4, pp. 875–889, July 2001.

[5] S. G. Dimmock, N. Wang, and J. Yang, "The endowment model and modern portfolio theory," *Management Science*, vol. 70, no. 3, pp. 1554–1579, April 2024.

[6] H. MarNowitz, "Portfolio selection," *Journal of Finance*, vol. 7, no. 1, pp. 77–91, March 1952.

[7] Z. Bodie, A. Kane, and A. Marcus, *Ebook: Investments-global edition*. McGraw Hill, 2014.

[8] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The journal of finance*, vol. 19, no. 3, pp. 425–442, September 1964.

[9] J. Lintner, "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," *The Review of Economics and Statistics*, vol. 47, no. 1, pp. 13–37, February 1965.

[10] E. F. Fama and K. R. French, "The capital asset pricing model: Theory and evidence," *Journal of economic perspectives*, vol. 18, no. 3, pp. 25–46, August 2004.

[11] R. Nian, J. Liu, and B. Huang, "A review on reinforcement learning: Introduction and applications in industrial process control," *Computers & Chemical Engineering*, vol. 139, no. 106886, August 2020.

[12] R. S. Sutton and A. G. Barto, *Reinforcement learning - an introduction*, ser. Adaptive computation and machine learning. MIT Press, 1998.

[13] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, ser. Wiley Series in Probability and Statistics. Wiley, 1994.

[14] S. Park, J. Chung, C. Park, S. Jung, M. Choi, S. Cho, and J. Kim, "Joint quantum reinforcement learning and stabilized control for spatio-temporal coordination in metaverse," *IEEE Transactions on Mobile Computing*, May 2024 (Early Access).

[15] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. A. Riedmiller, A. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, February 2015.

[16] W. Jiang, "Applications of deep learning in stock market prediction: recent progress," *Expert Systems with Applications*, vol. 184, p. 115537, December 2021.

[17] J. Zou, J. Lou, B. Wang, and S. Liu, "A novel deep reinforcement learning based automated stock trading system using cascaded LSTM networks," *Expert Systems with Applications*, vol. 242, p. 122801, May 2024.

[18] D. O. Oyewola, S. A. Akinwunmi, and T. O. Omotehinwa, "Deep LSTM and LSTM-Attention Q-learning based reinforcement learning in oil and gas sector prediction," *Knowledge-Based Systems*, vol. 284, p. 111290, January 2024.

[19] W. Zhang, Y. Zhao, S. Sun, J. Ying, Y. Xie, Z. Song, X. Wang, and B. An, "Reinforcement learning with maskable stock representation for portfolio management in customizable stock pools," in *Proc. of the ACM on Web Conference*, Singapore, May 2024, pp. 187–198.
[20] Q. Sun, X. Wei, and X. Yang, "GraphSAGE with deep reinforcement learning for financial portfolio optimization," *Expert Systems with Applications*, vol. 238, no. Part C, p. 122027, March 2024.
[21] G. D. Santos and K. R. Lima, "Impact of combinatorial optimization on reinforcement learning for stock trading in financial markets," in *Proc. of the Brazilian Symposium on Information Systems (SBSI)*, Juiz de Fora, Brazil, May 2024, pp. 1–8.