A Comparative Study on the Effects of Transfer Learning in Reinforcement Learning

Ryan Sherby
Computer Science
Columbia University
NY, USA
ryan.sherby@columbia.edu

Ishan Kumar
Computer Science
Columbia University
NY, USA
ik2592@columbia.edu

I. Introduction

We explore the potential for different deep reinforcement learning (DRL) algorithms to empirically demonstrate transfer learning in a quantitative trading scenario. Transfer learning, the process of leveraging knowledge gained from solving one task to improve performance on a related task, is particularly valuable in finance, where data collection fees can be prohibitive and collinear signals often result in overfitting.

Specifically, we evaluate how effectively agents trained in the equity market environment generalize their learned trading policies to the crypto market environment, thus quantifying both the benefits and limitations of transfer learning in quantitative trading. Through this we aim to answer the following research questions:

RQ1: How does transfer learning from a related financial domain affect the convergence rate and final performance of RL agents in quantitative trading compared to training from scratch?

RQ2: Are RL agents able to learn fundamental trading knowledge that is invariant across different financial markets?

Within our state space, we differentiate between static observations, those observed by the agent, and dynamic observations, those controlled by the agent. Our static observations consist of selected financial metrics, temporal information, historical observations, and unstructured detail. Our dynamic observations consist of available units for liquidation, unit costs, and purchasing power.

We use a discrete action space consisting of three actions: "Buy", "Hold", and "Sell", where each represents a decision to acquire, stand, or dispose of 1 or more units at a particular time. The state transition effects of each action are as follows:

- Buy: Affects the dynamic observation space and returns no reward.
- Hold: Affects the static observation space and returns no reward.
- Sell: Affects the dynamic observation space and returns a reward.

We define a reward function as follows:

$$R(s_t, a_t) := \mathbb{I}[a_t = a^S, s_t = s^{VD}]\left(\frac{P_t - P_k}{P_k} - \lambda\sigma_t - \mu\frac{P_{max} - P_t}{P_{max}}\right)$$

where a^S is the action to "Sell", s^{VD} reflects a state with a valid dynamic observation space, P_i is the price at time i, σ_i is the volatility at time i, and P_{max} is the largest price observed; λ and μ are tunable hyperparameters.

II. Methods

A. Dataset for Target Domain

We plan to use the FinRL competition dataset [1] for testing our two approaches. The competition provides a training set of Limit Order Book data of Bitcoin as well as a hidden test set.

B. Models and Evaluation

We aim to compare transfer learning from source domain and training from scratch for different traditional RL agents like DQN, PPO, and A2C implemented using neural networks. In addition, we plan to use an LLM-controlled agent to see if it has a better transfer because of larger parameter space. We plan to apply the policies learned on the equity dataset to our crypto test dataset and evaluate the earned rewards against the crypto-specific policies.

References

[1] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," *CoRR*, vol. abs/2106.09685, 2021. [Online]. Available: https://arxiv.org/abs/2106.09685

[2] S. Yun, "Pretrained llm adapted with lora as a decision transformer for offline rl in quantitative trading," 2024. [Online]. Available: https://arxiv.org/abs/2411.17900

[3] H. Yang, X.-Y. Liu, and C. D. Wang, "Fingpt: Open-source financial large language models," 2023. [Online]. Available: https://arxiv.org/abs/2306.06031

[4] X.-Y. Liu, H. Yang, J. Gao, and C. D. Wang, "FinRL: Deep reinforcement learning framework to automate trading in quantitative finance," *ACM International Conference on AI in Finance (ICAIF)*, 2021.

[5] J. Hong, A. Dragan, and S. Levine, "Q-sft: Q-learning for language models via supervised fine-tuning," 2024. [Online]. Available: https://arxiv.org/abs/2411.05193

[1] https://open-finance-lab.github.io/FinRL_Contest_2025/