Reinforcement Learning-Based AI for Sustainable Stock Trading System in the Stock Exchange of Thailand

Pannawish Tanthawichian, Than Rattanakij, Prompong Pakawanwong, Kanes Sumetpipat
Kamnoetvidya Science Academy, THAILAND
pannawish.t@kvis.ac.th, than.r@kvis.ac.th, prompong.p@kvis.ac.th

Abstract

This research proposes a reinforcement learning (RL)-based automated stock trading system tailored for the Stock Exchange of Thailand (SET). The study explores the application of RL algorithms—Advantage Actor-Critic (A2C), Soft Actor-Critic (SAC), Proximal Policy Optimization (PPO), Deep Deterministic Policy Gradient (DDPG), and Twin Delayed DDPG (TD3)—combined with Bayesian hyperparameter optimization. Experiments were conducted using two datasets with 6 and 9 input variables, focusing on 70 and 130 stocks. Results indicate that hyperparameter tuning significantly improves returns, while the number of input variables has minimal impact. PPO emerged as the best-performing model, achieving a cumulative return of 214.59% with 70 stocks and 6 input variables. The study also explores ensemble methods and stop-loss strategies, demonstrating their effectiveness in stabilizing returns and managing risk. These findings highlight the potential of RL-based systems for sustainable stock trading in Thailand's market.

Introduction

The stock market is a dynamic and complex environment where investors seek to maximize profits while managing risk. Traditional trading strategies often struggle to adapt to rapidly changing market conditions. Artificial intelligence (AI)-based automated trading systems have emerged as a solution, leveraging machine learning techniques to analyze data and execute trades. Reinforcement learning (RL), a subset of machine learning, has shown promise in developing adaptive trading systems by learning optimal strategies through interaction with the environment.

While RL-based trading systems have been extensively researched in global markets, their application in Thailand's stock market remains underexplored. This study aims to fill this gap by developing an RL-based trading system for the Stock Exchange of Thailand (SET). The research focuses on optimizing RL algorithms (A2C, SAC, PPO, DDPG, TD3) through hyperparameter tuning and evaluates their performance using key metrics such as cumulative returns, Sharpe ratio, and maximum drawdown. The findings contribute to the growing body of knowledge on AI-driven trading systems and provide insights into their applicability in emerging markets like Thailand.

Literature Review

Machine Learning in Trading

Machine learning has revolutionized stock trading by enabling the analysis of large datasets to identify patterns and predict price movements. Techniques such as Long Short-Term Memory (LSTM) networks are used for time-series forecasting (Silva et al., 2020), while reinforcement learning (RL) algorithms are employed to develop adaptive trading strategies (Li et al., 2020). RL-based systems learn by interacting with the market environment, maximizing cumulative rewards through trial and error.

Reinforcement Learning Algorithms

This study utilizes five RL algorithms:

1. **Advantage Actor-Critic (A2C):** Combines policy gradient methods with a value function to stabilize training (Yoon, 2019).

2. **Soft Actor-Critic (SAC):** Maximizes both returns and policy entropy, encouraging exploration (Liu, 2023).

3. **Proximal Policy Optimization (PPO):** Uses a clipped objective function to ensure stable policy updates (Liu, 2023).

4. **Deep Deterministic Policy Gradient (DDPG):** Extends Q-learning to continuous action spaces (Liu, 2023).

5. **Twin Delayed DDPG (TD3):** Improves DDPG by reducing overestimation bias and adding noise to target policies (Liu, 2023).

Hyperparameter Optimization

Hyperparameter tuning is critical for optimizing model performance. This study employs Bayesian optimization, a probabilistic approach that efficiently searches the hyperparameter space to maximize returns (Wu et al., 2019).

Methodology

The research methodology involves the following steps:

1. **Data Collection:** Historical stock data from 2014 to 2024 was collected using the Yahoo Finance API.

2. **Preprocessing:** Technical indicators such as MACD, RSI, and moving averages were added to the dataset. Training Set: 2014 to 2019, Validation Set: 2020 to 2021, Test Set: 2022 to 2024.

3. **Experiments:**

Experiment 1: Trading with 70 stocks and 6 input variables.

Experiment 2: Trading with 70 stocks and 9 input variables.

Experiment 3: Hyperparameter tuning with 70 stocks and 6 input variables.

Experiment 4: Ensemble methods combining multiple models.

Experiment 5: Hyperparameter tuning with 130 stocks and 9 input variables.

Experiment 6: Stop-loss implementation to manage risk.

4. **Performance Metrics:** Cumulative returns, annual return, Sharpe ratio, Sortino ratio, and max drawdown were used to evaluate performance.

Results and Discussion

Key Findings

Experiment 1 (70 Stocks, 6 Inputs): SAC achieved the highest cumulative return (60.32%).

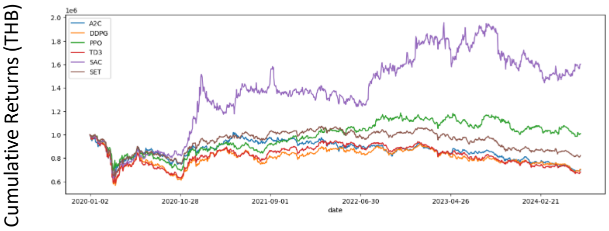

Figure 1: Cumulative Returns vs. Time with 70 stocks using 6 Input Variables.

Experiment 2 (70 Stocks, 9 Inputs): DDPG performed best (50.04%), but increasing input variables did not significantly improve returns.

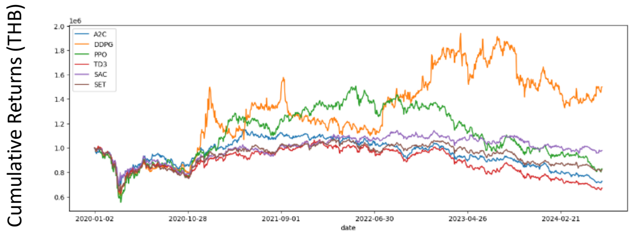

Figure 2: Cumulative Returns vs. Time with 70 stocks using 9 Input Variables.

Experiment 3 (Hyperparameter Tuning): PPO outperformed other models, achieving a cumulative return of 214.59%.

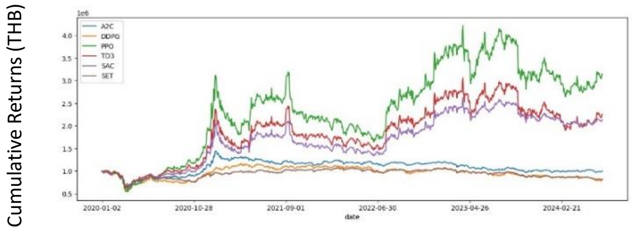

Figure 3: Cumulative Returns vs. Time with Hyperparameter Tuning and 70 stocks using 6 Input Variables.

Experiment 4 (Ensemble Methods): The ensemble approach achieved 117.53% cumulative returns, demonstrating improved stability.

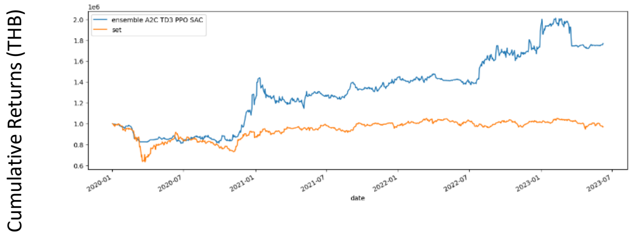

Figure 4: Cumulative Returns vs. Time with Ensemble Methods.

Experiment 5 (130 Stocks, 9 Inputs): PPO maintained consistent positive annual returns from 2020 to 2024.

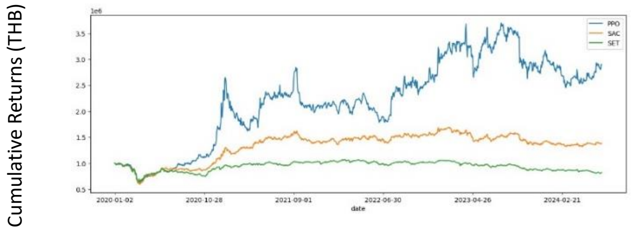

Figure 5: Cumulative Returns vs. Time with Hyperparameter Tuning with 130 Stocks and 9 Input Variables.

Experiment 6 (Stop Loss): Stop-loss strategies were effective for PPO, reducing losses and improving profitability.

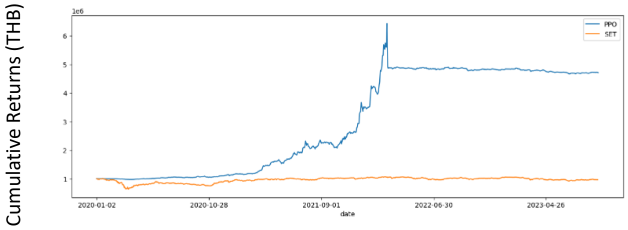

Figure 6: Cumulative Returns vs. Time with Stop Loss Implementation for PPO.

Discussion

Hyperparameter tuning had the most significant impact on performance, with PPO emerging as the best-performing model. Ensemble methods and stop-loss strategies further enhanced stability and risk management. The results highlight the importance of optimizing RL algorithms for sustainable trading in Thailand’s market.

Conclusion

This study demonstrates the effectiveness of reinforcement learning-based trading systems in the Stock Exchange of Thailand. PPO, combined with hyperparameter tuning, delivered the highest returns and best risk-adjusted performance (214.59% from 2020 to 2024). Future research could explore integrating LSTM networks with RL models to further improve prediction accuracy and decision-making. The findings provide a foundation for developing sustainable AI-driven trading systems in emerging markets.

Experiment	Cumulative Returns	Sharpe Ratio	Max Drawdown	Best Performing Model	Notable Observations
1)70 Stocks, 6 Inputs	60.32%	0.53	-37.30%	SAC	Moderate returns, some input variables more effective
2)70 Stocks, 9 Inputs	50.04%	0.46	-39.24%	DDPG	the returns were moderate, and increasing the number of inputs did not improve the returns.
3)Hyperparameter Tuning, 70 Stocks, 6 Inputs	214.59%	0.8	-48.59%	PPO	Significant improvement with tuning, better risk management
4)Ensemble Methods	117.53%	0.76	-36.42%	A2C, PPO, TD3, SAC	Stable returns, reduced variance, effective in managing volatility
5)Hyperparameter Tuning, 130 Stocks, 9 Inputs	189.92%	0.81	-38.56%	PPO	The annual returns were consistently positive and stable each year.
6)Stop Loss Implementation	370.72%	1.91	-27.44%	PPO	Effective in reducing losses for PPO

Table 1: Presents a summary of the results of all experiments.

References

Li, Y., Ni, P., & Chang, V. (2020). Application of deep reinforcement learning in stock trading strategies and stock forecasting. *Computing*, *102*(6), 1305-1322.

Liu, S. (2023). An evaluation of DDPG, TD3, SAC, and PPO: Deep reinforcement learning algorithms for controlling continuous systems. In *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*. https://doi.org/10.2991/978-94-6463-370-2_3

Silva, T. R., Li, A. W., & Pamplona, E. O. (2020, July). Automated trading system for stock index using LSTM neural networks and risk management. In *2020 international joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.

Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, *17*(1), 26-40.

Yoon, C. (2019, July 17). Understanding actor critic methods. *Medium*. Retrieved August 9, 2023, from https://towardsdatascience.com/understanding-actor-critic-methods-931b97b6df3f

Zhang, H. (n.d.). FinRL Ecosystem Tutorial and Introduction (Part I). *Zhihu*. Retrieved August 9, 2023,
from https://zhuanlan.zhihu.com/p/490682397