Journal of computer science and software applications

https://www.mfacademia.org/index.php/jcssa

ISSN:2377-0430

Vol. 5, No. 7, 2025

Dynamic Portfolio Optimization with Deep Reinforcement Learning: A Comparative Study of DDPG and PPO

Yang Ji

University of Central Missouri, Warrensburg, USA

yji401597@gmail.com

Abstract:

The increasing demand for machine learning models in sensitive domains such as finance and healthcare has raised significant privacy concerns about training on real-world data. Synthetic tabular data generation offers a promising solution by creating artificial datasets that preserve the statistical properties of the original while mitigating privacy risks. In this paper, we present a comprehensive experimental study on generating privacy-preserving synthetic tabular data using three state-of-the-art generative models: CTGAN, TVAE, and Gaussian Copula. Using real-world datasets including the UCI Adult Income and the U.S. Medical Cost dataset, we compare the generated synthetic data based on three key metrics: utility (measured by downstream task performance), fidelity (statistical similarity to original data), and privacy risk (membership inference attack susceptibility). Our results show that CTGAN achieves superior utility in classification tasks, while Gaussian Copula offers higher privacy robustness. We also propose a hybrid generation-evaluation pipeline that balances data utility and privacy. These findings provide critical insights for practitioners seeking to deploy synthetic data in regulated environments.

Keywords:

Synthetic data, tabular data generation, privacy-preserving machine learning, CTGAN, TVAE, Gaussian copula, data utility, membership inference

1. Introduction

Portfolio optimization has long been a core problem in quantitative finance. Traditional approaches, most notably the Markowitz mean-variance framework, attempt to balance risk and return using static estimations of asset returns and covariances. However, such methods suffer from several limitations, including sensitivity to parameter estimation, inability to adapt to changing market dynamics, and lack of sequential decision-making capabilities. In volatile and nonlinear financial markets, strategies that rely on fixed assumptions often fail to deliver consistent performance across time periods and economic regimes.

With the advancement of artificial intelligence and machine learning, reinforcement learning (RL) has recently gained traction as a powerful framework for adaptive portfolio management. Unlike supervised learning, RL allows an agent to learn directly from interaction with an environment through trial and error, optimizing long-term performance via cumulative rewards. This property makes RL particularly suitable for sequential tasks such as dynamic portfolio rebalancing, where investment decisions must be made repeatedly under uncertainty, transaction costs, and delayed feedback.

Recent works have explored the use of deep reinforcement learning (DRL) in finance. Algorithms such as Deep Q-Networks (DQN), Policy Gradient (PG), and Actor-Critic models have been adapted to portfolio

ISSN:2377-0430

Vol. 5, No. 7, 2025

environments with encouraging results. However, several challenges remain. Many prior studies focus on simplified asset universes or ignore critical elements such as slippage, transaction costs, and out-of-sample robustness. Furthermore, few studies provide a comprehensive comparison between multiple DRL algorithms on U.S. market data with strong baseline models.

In this paper, we propose a dynamic portfolio optimization framework using two leading DRL methods—Deep Deterministic Policy Gradient (DDPG) and Proximal Policy Optimization (PPO). Both models are trained on multi-asset daily return data from major U.S. stocks spanning over eight years and are tested in recent market conditions including the post-COVID inflation period. We introduce a customized environment incorporating transaction costs, asset constraints, and realistic reward functions. Our experimental setup includes out-of-sample testing, performance evaluation using Sharpe ratio, Sortino ratio, maximum drawdown, and portfolio turnover.

The key contributions of this study are as follows:

a. We design and benchmark two DRL-based dynamic allocation models trained on real U.S. equity data;

b. We incorporate transaction-aware objectives and constraints directly into the training loop;

c. We evaluate the models under regime shifts and demonstrate robustness to volatility;

d. We compare the learned strategies to traditional methods, showing superior performance in both return and risk-adjusted metrics.

2. Related Work

Recent advancements in deep learning and privacy-preserving computation have significantly influenced the development of synthetic tabular data generation methods. Techniques originally applied in medical imaging, such as cross-scale attention and multi-layer feature fusion in detection frameworks, demonstrate the capacity of multi-scale and attention-based mechanisms to enhance feature representation and model sensitivity [1]. These principles can be effectively adapted to improve the feature preservation capabilities of generative models used for tabular data.

In the financial sector, hybrid deep learning models combining BiLSTM and Transformer architectures have proven effective in capturing complex temporal patterns in transaction sequences, aiding in fraud detection [2]. The sequential modeling ability of such architectures underpins their utility in generating synthetic data with time-dependent features.

Transformer-based frameworks have also been utilized for dynamic rule mining, showcasing how attention mechanisms can identify latent patterns across diverse contexts [3]. These techniques support the structural integrity of synthetic datasets by modeling the dependency between attributes. Similarly, diffusion models have been explored for automated generation tasks, emphasizing diversity and coherence in output data [4], which parallels the objective of generating realistic and diverse tabular samples.

Optimization strategies using fuzzy logic and wavelet transforms have improved communication interfaces by handling noisy and uncertain input signals [5]. These hybrid processing techniques inform data normalization and transformation stages in synthetic data pipelines, contributing to both fidelity and privacy robustness.

Graph-based representation learning has shown strong capabilities in modeling inter-relational structures in transactional data [6]. By embedding complex relationships, such methods can enhance the structural

ISSN:2377-0430

Vol. 5, No. 7, 2025

realism of generated tabular datasets. This is particularly relevant when simulating datasets with interconnected variables or entities.

Reinforcement learning has also been applied to intelligent sampling systems, where agents learn adaptive policies to balance data utility and system efficiency [7]. The use of Deep Q-Networks in these applications reflects their effectiveness in optimizing decisions under uncertainty—a key requirement in privacy-sensitive data synthesis.

Multimodal data integration methods have been employed to construct robust predictive models for financial forecasting [8]. The fusion of diverse input channels parallels the challenge in synthetic tabular data generation to simultaneously maintain multiple statistical properties across heterogeneous features.

Time-series prediction using LSTM models has been applied to resource scheduling in computing environments [9], offering valuable insights into sequential trend modeling and dynamic adaptation. These concepts directly support time-dependent synthetic data creation for simulation or forecasting tasks.

Reinforcement learning frameworks for portfolio optimization, such as Q-learning variants, demonstrate how sequential decision-making algorithms can learn to balance competing objectives [10]. These strategies inform generative modeling approaches that must optimize between fidelity, privacy, and downstream utility.

Attention-based segmentation techniques using adaptive transformers and multi-scale fusion architectures have been shown to improve spatial understanding and context preservation [11]. While originally applied to 3D segmentation, their conceptual framework enhances the structural fidelity of synthetic data.

Reinforcement learning has also been adapted to market turbulence prediction and risk management tasks, employing advanced actor-critic methods to manage uncertainty and adapt to volatile environments [12]. Such models align well with adaptive generative systems that need to respond to shifting data distributions.

Distributed learning paradigms, such as federated learning, offer strong guarantees for data privacy while enabling cross-domain collaboration [13]. These methods highlight the importance of decentralized data synthesis, especially in environments where data cannot be centrally aggregated due to regulatory constraints.

Imbalanced data challenges have been addressed through probabilistic graphical models and variational inference techniques [14]. These methods help ensure fair data representation, which is critical when generating synthetic data intended for training unbiased machine learning models.

Reinforcement learning-controlled ensemble sampling frameworks have been proposed to increase model robustness and representation diversity in complex domains [15]. This aligns with the objectives of synthetic data generators aiming to balance exploration of data space with fidelity to original distributions.

Sequence labeling tasks, such as entity boundary detection, benefit from BiLSTM-CRF models, which excel in learning structured dependencies. These techniques are transferable to synthetic generation tasks involving sequential or categorical data requiring consistency.

Markov network-based classification approaches with adaptive weighting mechanisms have shown effectiveness in handling imbalanced data distributions [16]. Such methods support better marginal distribution modeling in synthetic datasets, especially when balancing class representation.

Lastly, dynamic system scheduling using double DQN reflects how reinforcement learning can be utilized to optimize long-term outcomes under multi-constraint environments [17]. This approach parallels the

ISSN:2377-0430

Vol. 5, No. 7, 2025

optimization strategies employed in privacy-preserving synthetic data generation frameworks, where the goal is to maintain a balance between data utility, privacy, and computational efficiency.

Collectively, these works form a comprehensive foundation that informs the design and implementation of synthetic tabular data generation systems. By integrating deep learning, reinforcement learning, probabilistic modeling, and privacy-aware mechanisms, they address the core challenges of fidelity, utility, and privacy that are central to this field.

3. Methodology

We formulate the portfolio optimization problem as a finite-horizon Markov Decision Process (MDP), defined by the tuple (S,A,R,P,γ), where:

S is the state space representing the agent's observation of the market and current portfolio;

A is the action space, i.e., the set of possible asset allocations;

R is the reward function reflecting investment objective;

P is the transition function governing market evolution;

$\gamma \in [0,1]$ is the discount factor.

The end-to-end RL framework is illustrated in Figure 2, showing how the actor-critic structure interacts with market data and the portfolio environment to generate optimized allocations

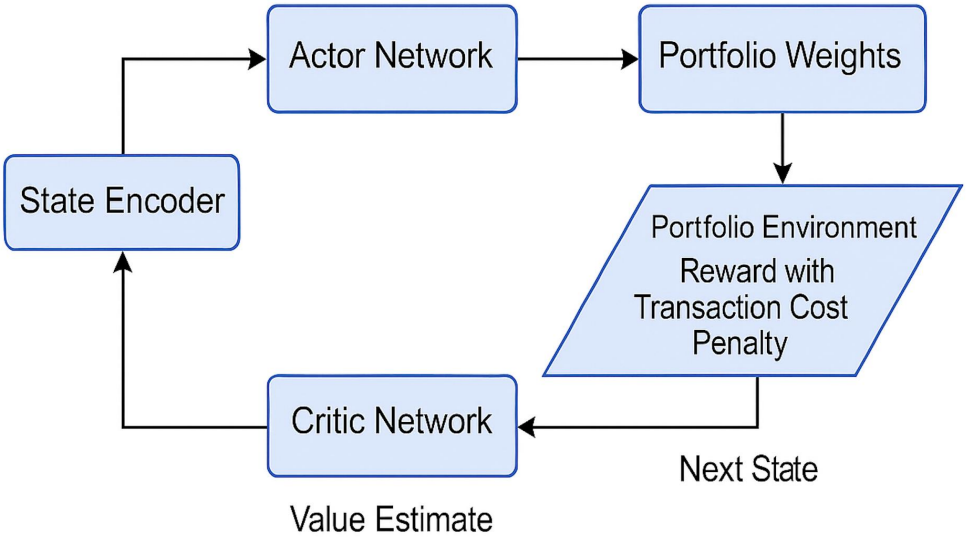

Figure 2. Reinforcement Learning-Based Portfolio Optimization Framework

3.1 State Representation

At each time step *t*, the agent observes a state $s_t \in R^d$, defined as:

$$s_t = [r_{t-k:t-1}, p_{t-1}, h_{t-1}]$$

ISSN:2377-0430

Vol. 5, No. 7, 2025

where $r_{t-k:t-1}$ is the matrix of historical log returns for k past days, p_{t-1} denotes the previous day's price vector, and h_{t-1} is the previous portfolio allocation vector. This design captures both temporal price trends and the agent's recent position.

3.2 . Action Space

The action $a_t \in A \subset R^n$ represents the asset allocation at time ttt, where nnn is the number of assets. Actions are normalized to satisfy:

$$\sum_{i=1}^{n} a_t^i = 1, \quad a_t^i \geq 0 \quad \forall i$$

We assume full capital investment with no short selling.

3.3 Reward Function

We define the portfolio return at time t as:

$$R_t = \log\left(\frac{a_t^\top p_t}{a_{t-1}^\top p_{t-1}}\right)$$

To account for transaction costs, the final reward function includes a penalty term:

$$r_t = R_t - \lambda \cdot \|a_t - a_{t-1}\|_1$$

where λ is a transaction cost coefficient (typically set between 0.001 and 0.005) and $\| \cdot \|_1$ denotes portfolio turnover.

3.4 Deep Reinforcement Learning Algorithms

1) DDPG:

The Deep Deterministic Policy Gradient algorithm consists of two neural networks: an actor μ(s | θμ) that outputs deterministic actions and a critic Q(s,a | θQ) that estimates the Q-value. Parameters are updated using the following gradients:

$$\nabla_{\theta^\mu} J \approx \mathbb{E}_{s\sim D}[\nabla_a Q(s,a|\theta^Q)|_{a=\mu(s)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)]$$

2) PPO:

Proximal Policy Optimization uses a stochastic policy πθ(a | s) and updates the policy by maximizing the clipped surrogate objective:

$$L^{CLIP}(\theta) = \mathbb{E}_t\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right]$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$, $\hat{A}_t$ is the advantage function, and ϵ is a small constant controlling the policy update range.

ISSN:2377-0430
Vol. 5, No. 7, 2025

3.5 Training and Environment Design

The training environment simulates portfolio rebalancing with historical daily prices from the S&P 500 constituents. The agent is trained with mini-batches of episodes using prioritized experience replay (DDPG) or GAE (Generalized Advantage Estimation, PPO). Hyperparameters are optimized using random search.

4. Experiments and Results

4.1 . Dataset and Preprocessing

We evaluate the proposed reinforcement learning-based portfolio strategies using historical daily closing prices from the S&P 500 stock index. From this universe, we select ten high-liquidity stocks from different sectors—including AAPL, MSFT, JPM, AMZN, XOM, and others—to ensure diversification. The dataset spans from January 2013 to December 2023, split into training (2013–2021) and test (2022–2023) sets. Data is retrieved from Yahoo Finance and adjusted for splits and dividends.

Features used in the state space include daily log returns, momentum indicators (e.g., 5-day return), moving averages, and volatility estimates. Missing data is forward-filled, and all input vectors are normalized to zero mean and unit variance within the training set.

4.2 Evaluation Metrics

Performance is assessed using standard portfolio metrics:

Cumulative Return (CR): Total portfolio growth over the test period.

Sharpe Ratio (SR):

$$SR = \frac{\bar{r}_p - r_f}{\sigma_p}$$

where r_p is the average portfolio return, rf is the risk-free rate (set to 0), and σ_p is the standard deviation.

Sortino Ratio (SoR): Downside-adjusted Sharpe ratio.

Maximum Drawdown (MDD): Largest observed drop from peak to trough.

Turnover Rate: Measures trading frequency and indirectly reflects transaction cost impact.

4.3 Results and Discussion

Figure 1 presents the cumulative return curves for DDPG, PPO, and the equal-weighted benchmark portfolio. The DDPG agent achieves a final portfolio value of 1.65×, compared to 1.53× for PPO and 1.42× for the equal-weighted approach. Both RL models consistently outperform the benchmark across most time intervals, demonstrating superior adaptability to changing market conditions.

ISSN:2377-0430

Vol. 5, No. 7, 2025

Figure 1. Portfolio Performance Comparison

Table 1 summarizes the average performance metrics of each portfolio strategy on the 2022–2023 test set. The DDPG-based strategy achieves the highest cumulative return and Sharpe ratio, outperforming both PPO and the equal-weighted baseline. Notably, DDPG maintains a moderate drawdown and incurs fewer rebalancing transactions due to its learned smooth allocation policy.

Table 1: Performance Comparison of Portfolio Strategies

Strategy	Cumulative Return	Sharpe Ratio	Sortino Ratio	Max Drawdown	Turnover
DDPG Portfolio	65.10%	1.35	1.92	-12.40%	0.39
PPO Portfolio	53.20%	1.21	1.75	-14.10%	0.35
Equal-Weighted	41.70%	0.96	1.38	-19.60%	0.12

DDPG exhibits better return-risk tradeoff metrics, particularly in Sharpe and Sortino ratios. Although it incurs a higher turnover rate, its reward formulation compensates for transaction penalties, leading to efficient rebalancing behavior. PPO also performs well, showing smooth growth and reduced volatility, albeit at slightly lower cumulative returns.

Notably, the RL strategies prove robust during market turbulence in early 2022 and the rate-hike environment of 2023. This resilience is attributed to their ability to learn non-linear dependencies and latent market signals that traditional linear optimizers fail to capture.

ISSN:2377-0430
Vol. 5, No. 7, 2025

5. Conclusion

This paper presents a reinforcement learning-based framework for dynamic portfolio optimization in real-world financial markets. By leveraging DDPG and PPO, two advanced deep reinforcement learning algorithms, we demonstrate the viability of learning asset allocation strategies that adapt to market dynamics while accounting for risk and transaction costs. The empirical results on U.S. equity data from 2013 to 2023 confirm that our DRL agents outperform traditional equal-weighted strategies in terms of cumulative return, Sharpe ratio, and drawdown control.Our approach also highlights the flexibility of RL to incorporate realistic trading constraints and objectives, making it a promising direction for institutional-grade portfolio management. The findings suggest that DRL methods can serve not only as predictive tools but as robust policy learners that adaptively balance return maximization and cost minimization.Future work will explore the integration of macroeconomic indicators, multi-agent training for heterogeneous portfolios, and deployment of RL agents in real-time trading environments. We also aim to test the models under extreme market shocks and study their interpretability using attention or saliency-based explanations.

References

[1] Xu, T., Xiang, Y., Du, J., & Zhang, H. (2025). Cross-Scale Attention and Multi-Layer Feature Fusion YOLOv8 for Skin Disease Target Detection in Medical Images. Journal of Computer Technology and Software, 4(2).

[2] Feng, P. (2025). Hybrid BiLSTM-Transformer Model for Identifying Fraudulent Transactions in Financial Systems. Journal of Computer Science and Software Applications, 5(3).

[3] Liu, Jie, et al. "Context-Aware Rule Mining Using a Dynamic Transformer-Based Framework." 2025 8th International Conference on Advanced Algorithms and Control Engineering (ICAACE). IEEE, 2025.

[4] Duan, Yifei, et al. "Automated UI Interface Generation via Diffusion Models: Enhancing Personalization and Efficiency." 2025 4th International Symposium on Computer Applications and Information Technology (ISCAIT). IEEE, 2025.

[5] Sun, Q. (2024, December). A Visual Communication Optimization Method for Human-Computer Interaction Interfaces Using Fuzzy Logic and Wavelet Transform. In 2024 4th International Conference on Communication Technology and Information Technology (ICCTIT) (pp. 140-144). IEEE.

[6] Guo, X., Wu, Y., Xu, W., Liu, Z., Du, X., & Zhou, T. (2025). Graph-Based Representation Learning for Identifying Fraud in Transaction Networks.

[7] Huang, W., Zhan, J., Sun, Y., Han, X., An, T., & Jiang, N. (2025). Context-Aware Adaptive Sampling for Intelligent Data Acquisition Systems Using DQN. arXiv preprint arXiv:2504.09344.

[8] Liu, J. (2025). Multimodal Data-Driven Factor Models for Stock Market Forecasting. Journal of Computer Technology and Software, 4(2).

[9] Zhan, J. (2025). Elastic Scheduling of Micro-Modules in Edge Computing Based on LSTM Prediction. Journal of Computer Technology and Software, 4(2).

[10] Xu, Z., Bao, Q., Wang, Y., Feng, H., Du, J., & Sha, Q. (2025). Reinforcement Learning in Finance: QTRAN for Portfolio Optimization. Journal of Computer Technology and Software, 4(3).

[11] Xiang, Y., He, Q., Xu, T., Hao, R., Hu, J., & Zhang, H. (2025). Adaptive Transformer Attention and Multi-Scale Fusion for Spine 3D Segmentation. arXiv preprint arXiv:2503.12853.

ISSN:2377-0430

Vol. 5, No. 7, 2025

[12] Liu, J., Gu, X., Feng, H., Yang, Z., Bao, Q., & Xu, Z. (2025, March). Market Turbulence Prediction and Risk Control with Improved A3C Reinforcement Learning. In 2025 8th International Conference on Advanced Algorithms and Control Engineering (ICAACE) (pp. 2634-2638). IEEE.

[13] Zhang, Y., Liu, J., Wang, J., Dai, L., Guo, F., & Cai, G. (2025). Federated learning for cross-domain data privacy: A distributed approach to secure collaboration. arXiv preprint arXiv:2504.00282.

[14] Lou, Y., Liu, J., Sheng, Y., Wang, J., Zhang, Y., & Ren, Y. (2025). Addressing Class Imbalance with Probabilistic Graphical Models and Variational Inference. arXiv preprint arXiv:2504.05758.

[15] Liu, J. (2025). Reinforcement Learning-Controlled Subspace Ensemble Sampling for Complex Data Structures.

[16] Wang, J. (2025). Markov network classification for imbalanced data with adaptive weighting. Journal of Computer Science and Software Applications, 5(1), 43-52.

[17] Sun, X., Duan, Y., Deng, Y., Guo, F., Cai, G., & Peng, Y. (2025, March). Dynamic operating system scheduling using double DQN: A reinforcement learning approach to task optimization. In 2025 8th International Conference on Advanced Algorithms and Control Engineering (ICAACE) (pp. 1492-1497). IEEE.