Risk-Aware RL for Dynamic Portfolio Optimization in Financial Markets

Sitong Tie

August 2025

1 Background

Modern Portfolio Theory (MPT), established by Harry Markowitz, continues to serve as a fundamental framework for portfolio selection and asset allocation. Nevertheless, MPT presupposes unchanging market conditions, fixed asset correlations, and returns that follow a normal distribution. Financial markets demonstrate non-stationarity, volatility clustering, and regime shifts that necessitate adaptive decision-making.

Reinforcement Learning (RL) provides a data-centric methodology for sequential decision-making in uncertain environments. In contrast to conventional optimization techniques that address a static allocation issue, reinforcement learning can derive insights from market interactions and adjust allocations in real-time. This adaptive ability is especially beneficial in contexts characterized by frequent regime shifts, such as equities, commodities, or cryptocurrencies.

By incorporating transaction costs and drawdown constraints into the reward function, the proposed project seeks to use Deep Reinforcement Learning (DRL) to optimize portfolio allocations in a risk-aware manner. In volatile markets, it is hypothesized that a well-designed RL agent can outperform static strategies on risk-adjusted metrics.

2 Problem Statement

The goal is to build a RL-based portfolio management system that:

1. Allocates weights dynamically across multiple assets.
2. Optimizes for risk-adjusted returns e.g. Sharpe ratio.
3. Demonstrates statistically significant improvements over traditional benchmarks.

3 Methodology

3.1 data

We are going to use 5 to 10 highly liquid assets from like S&P500, the data will be from Yahoo finance with daily resolution, includes high, low, start, end price for the past 10 years.

3.2 State Representation

Each state vector at time t will include recent returns, rolling volatility, momentum indicators, the weight of portfolio.

3.3 Reward function

$R_t = \frac{E[r_p]}{\sigma_p} - \lambda_c \cdot TC_t - \lambda_d \cdot DD_t$.
Where r_t =portfolio return; σ_p =portfolio volatility; TC_t =transaction cost at time t; DD_t = drawdown penalty; and λ_c, λ_d are penalty coefficients.

3.4 RL Algorithms and strategies

We will use Deep Q-Network (DQN) with discretized allocations for now, and with strategies of buy and hold equal weights, Markowitz mean-variance optimization(will rebalanced monthly); I'm still considering if need to involve short and long.

3.5 Evaluation metrics

We will have serval metrics to analysis our result. such as Annualized return, Sharpe ratio, Maximum drawdown, Sortino ratio, Calmar ratio, etc.

3.6 Expected Outcomes

We expect to have an improved Sharpe ratio and drawdown control compared to static benchmarks. And be able to identify conditions where RL provides the largest benefit. (Provide an open-source implementation of a risk-aware RL portfolio optimizer.)