

포트폴리오 최적화를 위한 위험 회피 강화 학습

Bayaraa Enkhsaikhan, *조오 현

충북대학교, 컴퓨터과학 전공

bayaraa@chungbuk.ac.kr, *ohyunjo@chungbuk.ac.kr

Risk Averse Reinforcement Learning for Portfolio Optimization

Bayaraa Enkhsaikhan, *Ohyun Jo

Chungbuk National University, Department of Computer Science.

Abstract

In this paper, we investigate investment portfolio optimization using Reinforcement Learning (RL) with risk assessment. Due to market friction, the reaction of other market participants and uncertainties, it is challenging to trade and optimize investment portfolios dynamically. The financial market is sophisticated and complex to model. Moreover, regulatory requirement and internal risk policy require investors to make risk-averse decisions for preventing catastrophic results, which is hard to recover later. One way to solve the problem is to set a high enough penalty to reward for the risk. As the experiment result suggests, the proposed Value at Risk (VaR) technique using Actor-Critic reinforcement learning could benefit faster learning and reward.

I. Introduction

According to the Fundamental Asset Pricing theory, an investor cannot profit without making a risky investment. Moreover, the investor needs to assess the portfolio's risks invested by ensuring that the potential loss of the portfolio may not result in a catastrophic outcome. The Modern Portfolio Theory (Markowitz portfolio theory) suggests that the optimal portfolios with minimum risk given the constant return or maximum return given the constant risk can be computed. Those optimal portfolios formulate "efficient frontier", which can be illustrated by a line showing risk and returns are positively correlated. Therefore, if the investor wants a higher return, then riskier will be the portfolio. The ultimate optimum portfolio to choose depends on the investor's risk appetite (how much risk the investor ready to accept) and capacity (how much risk the investor can bear).

Risk appetite is usually pre-defined for sophisticated investors, and sometimes regulatory bodies restrict some financial institutions (such as banks, pension funds) to prevent high risk. Hence, the investment decision always constrained by the risk of the portfolio. The expected risk and return of the portfolios are always dynamic and can change time by time. Hence, investors always need to trade in the market actively and adjust to the optimal portfolio. However, continuous trading incurs fees and cost associated with the trade can outweigh the efficiency of the optimal portfolio.

This paper suggests the Machine-Learning decision-making technique called Reinforcement Learning to optimize portfolio selection with VaR constraint. The machine learning model learns possible risk and returns corresponding to trading decision and suggests the best possible policy to follow. We will use a statistical technique called Value at Risk (VaR), to calculate maximum potential loss given confidence. Moreover, VaR is used ML field with substantial risks such as autonomous driving. [1]

II. Model

Value at Risk.

Let X be a return distribution. The VaR at $\alpha \in (0, 1)$ confidence level is:

$$VaR_{\alpha}(X) = -\inf \{x \in \mathbb{R} : F_X(x) > \alpha\}$$

VaR can be computed through well-defines probability distribution, but the extreme event can impact market and it can become illiquid. We assume that the distribution of returns can be long tailed. Therefore, historical simulation based on the historical data to assess VaR is calculated and used for predicting portfolio VaR.

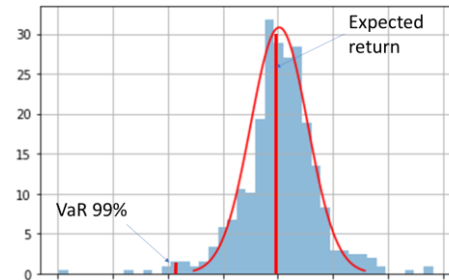


Figure2. Historical simulation method for VaR

The portfolio VaR is:

$$VaR_{\alpha}(X) = \sqrt{\sum_{i,j} w_i * VaR_{\alpha}(X_i) * w_j * VaR_{\alpha}(X_j) * cor(X_i, X_j)}$$

where X_i, X_j belong to X , w_i weight and $cor()$ is correlation function.

From above equation, the risk of the portfolio also depends on the correlation coefficient. If the coefficient is low, then optimization will diversify away risk.

Modern Portfolio theory.

The collection of optimal portfolios are those having minimal risk and maximal returns. In Figure1, the portfolios on the "efficient frontier" are better compared to the possible portfolios under the curve. The curve is called the efficient frontier. Investing in a

portfolio underneath the curve is not desirable because it does not maximize returns for a given level of risk.

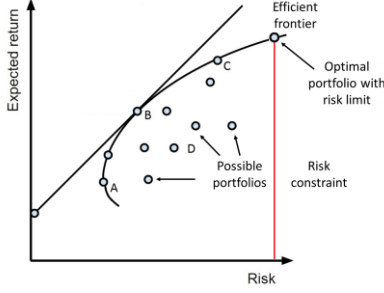


Figure3. Optimal portfolios

Reinforcement Learning.

We used Actor-Critic reinforcement learning model.

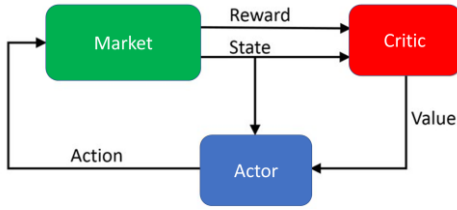


Figure4. Architecture of actor-critic reinforcement learning

Action: Our action space is weights for each asset in our portfolio. The weight corresponds to i -th asset at time t is $w_{i,t}$. The weights are constrained in interval $[0:1]$. Total value of the portfolio is v_t (episode length 1000).

Reward: The reward is calculated by the movement of values at each time. Company shares can have dividends; thus, we took adjusted price of the shares with dividend. The cost of transaction:

$$\beta_t = \sum_i |w_{i,t} * v_t * p_{i,t-1}/p_{i,t} - w_{i,t-1} * v_{t-1}|$$

Where $p_{i,t}$ is price of asset i at time t .

The reward is:

$$return_t = v_t - v_{t-1} - fee * \beta_t$$

We use self-financing trade; therefore, discount factor is 1.

State: We have selected historical and current moving average, maximum, minimum, median, 1st and 3rd quantile, standard deviation and VaR at 99% confidence over historical 1 month, 3 months, half-year and yearly periods as observation values. Moreover, price movement and normalized trading volume are also included in the observation state.

Actor-Critic: Asset price is considered continuous stochastic process. Therefore, we used Deep deterministic policy gradient method for our Actor and Critic feed forward neural networks. As shown in Figure4, Critic network is fed with reward and state, then gives result to Actor network.

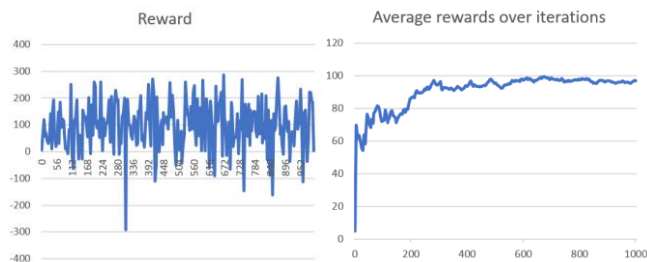


Figure5. Rewards after learning number of episodes.

In Figure5, we have illustrated the result of the experiment. The first figure shows the result of rewards over individual iterations. The second figure shows the average returns. On average, the model learns, but the result is still risky.

Then, we set a penalty reward to the model if it makes a risky decision. If the VaR of the asset exceeds the pre-defined risk limit, then the reward is penalized by a non-zero penalty.

$$penalty = \sum_i \gamma_i X_i \quad \text{if } VaR_{0.99}(X_i) > \epsilon, \text{ then } \gamma_i > 0, \text{ else } \gamma_i = 0$$

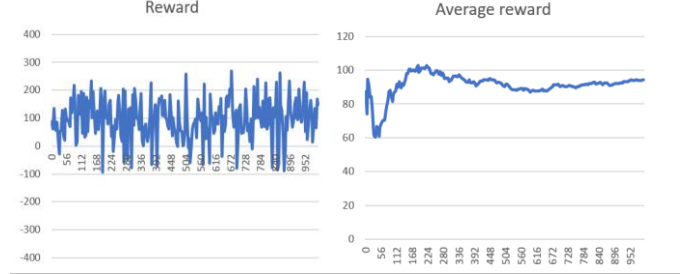


Figure6. Rewards after learning several episodes.

The result of the experiment with the risk penalty shown in Figure6. The standard deviation of the reward has been reduced by 18 per cent in that variant, but the average reward decreased slightly as expected.

III. Conclusion

A risk-averse reinforcement learning algorithm for optimal portfolio allocation has been proposed in this paper. We have successfully decreased the risk of dynamic portfolios for a change of a decrease in return. We conclude that implementing a risk-averse RL technique could prevent significant risk (a bankruptcy for extreme case), which cannot be recovered.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2020R1A6A1A12047945). In part, this work was supported by the Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2021-0-00165, Development of 5G+ Intelligent Basestation Software Modem).

References

- [1] Armengol Urp, N., Curi, S., and Krause, A., "Risk-Averse Offline Reinforcement Learning", <i>arXiv e-prints</i>, 2021.
- [2] Daping, Z., Yong, F., Chaoliang, Z., and Zongrun, W. "Portfolio Selection Based on Bayesian Theory" 2019, Volume 2019, (https://doi.org/10.1155/2019/4246903).
- [3] Martin, A et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467 (14 March 2016).
- [4] Yun, S., Ruihong, H., Chang, Y., and Klaus, O., "Risk-averse reinforcement learning for algorithmic trading," 2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr), 2014, pp. 391-398, doi: 10.1109/CIFEr.2014.6924100.