Transactions on Computational and Scientific Methods | Vo. 4, No. 12, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Dynamic Risk Control and Asset Allocation Using Q-Learning in Financial Markets

Mohan Jiang[1], Zhen Xu [2], Zhenghao Lin[3]

[1]New York University, New York, USA
[2]Independent Researcher, Shanghai, China
[3]Northeastern University, Boston, USA

*Correspondence should be addressed to Mohan Jiang; mj2589@nyu.edu

Abstract:

This study proposes an asset management risk control algorithm based on Q-learning, which aims to optimize asset allocation decisions through reinforcement learning, maximize investment returns, and effectively control risks. We used a real financial market dataset from Yahoo Finance to train and verify the algorithm. The dataset contains historical closing prices, trading volumes, volatility, and other information of multiple assets, with a time span of 5 years and a data frequency of daily. Experimental results show that Q-learning outperforms traditional models such as mean-variance optimization (MVO), genetic algorithm (GA), deep Q network (DQN), and support vector machine (SVM) in multiple evaluation indicators. Specifically, Q-learning achieved the best results in indicators such as cumulative return, Sharpe ratio, and maximum drawdown, demonstrating its adaptability and efficiency in a dynamic market environment. By simulating investment decisions under different market conditions, the Q-learning algorithm can adaptively adjust investment strategies, effectively respond to market fluctuations, and achieve optimal risk control. Despite this, reinforcement learning algorithms still have certain challenges in computational complexity and training time. In the future, the computational efficiency of the model can be improved by introducing more efficient algorithms and optimization strategies.

Keywords:

Q-learning, asset management, risk control, reinforcement learning

1. Introduction

In the field of asset management, risk control has always been a crucial topic. With the increasing complexity and uncertainty of the financial market, traditional risk management methods can no longer meet the ever-changing market needs. Therefore, studying how to achieve more accurate and dynamic risk control through advanced algorithms has become an important challenge in the field of asset management. In recent years, deep reinforcement learning, as a self-learning and highly adaptable intelligent algorithm, has gradually shown its great potential in the financial field. Q-learning, as a typical reinforcement learning algorithm, has broad application prospects, especially in risk control tasks in asset management, which can effectively optimize the decision-making process.

The Q-learning algorithm continuously updates its value function through interaction with the environment to achieve strategy optimization. Specifically, Q-learning can estimate future returns based on current status and behavior without clear supervision signals, thereby gradually learning the optimal strategy. In the risk control of asset management, investors are faced with a large amount of uncertain market information. Traditional strategies often rely on static models and are difficult to adapt to the ever-changing market situation. The Q-learning algorithm can select the optimal investment behavior under different market conditions by dynamically adjusting the decision-making strategy, thereby achieving effective risk control.

In addition, Q-learning has strong exploratory and adaptive capabilities. In the actual asset management process, investors often need to choose between different assets, and market volatility and external factors may lead to uncertainty in future returns. Q-learning can balance the relationship between exploration and utilization to a certain extent, allowing investors to flexibly adjust their investment portfolios in an unknown environment. By constantly interacting with the market environment, Q-learning can continuously update its strategy according to market changes, thereby effectively responding to complex market risks and uncertainties.

In practical applications, the Q-learning algorithm can be used to build an intelligent asset allocation model. The model can predict the future performance of different assets by learning historical market data, and automatically adjust the asset portfolio according to risk tolerance and expected returns. In this process, Q-learning can not only consider the independent risks of each asset but also comprehensively consider the correlation between assets and the overall trend of the market, so as to achieve better risk diversification in diversified investment strategies. This feature makes the application of Q-learning in asset management have great advantages, especially when the market changes rapidly, Q-learning can quickly adapt and adjust strategies.

However, Q-learning also faces certain challenges when applied to asset management. First, the decision-making problem in asset management is usually a high-dimensional, complex multi-objective optimization problem involving many variables. How to effectively search in the high-dimensional state space has become a difficulty in the application of Q-learning. Secondly, the data noise in the financial market is large, and traditional Q-learning may not be able to cope with a large amount of market volatility and uncertainty. To this end, researchers have proposed a variety of improvement methods, such as deep Q network (DQN) and double Q-learning (Double Q-learning), to improve the stability and efficiency of Q-learning in complex environments.

Nevertheless, the application prospects of Q-learning in asset management risk control are still very broad. With the continuous advancement of computing power and data processing technology, Q-learning algorithms can handle more complex and high-dimensional decision-making problems, thereby providing more accurate and real-time risk control solutions for asset management. In addition, combined with other deep learning models and optimization algorithms, Q-learning can achieve more efficient decision-making in scenarios such as big data and high-frequency trading, thereby further improving the overall effect of asset management. In the future, with the continuous development of intelligent algorithms, Q-learning is expected to become an indispensable part of the asset management field, providing investors with more accurate, intelligent and efficient risk control tools.

2. Method

In this study, we proposed an asset management risk control algorithm based on Q-learning, which gradually learns the optimal risk control strategy by simulating the interaction between investors and the market. Q-learning is a model-free reinforcement learning algorithm that estimates the value function of each state-action pair through interaction with the environment and optimizes the decision-making strategy by maximizing the expected return. In asset management applications, investors need to make investment decisions based on the state of the market and their own risk preferences. The Q-learning algorithm can effectively achieve risk control by continuously updating its strategy. Its network architecture is shown in Figure 1.

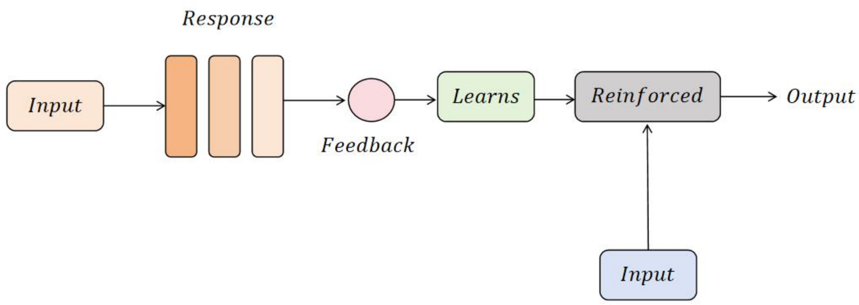

Figure 1. Overall network architecture diagram

First of all, the core of Q-learning lies in the value iteration process. In asset management, we abstract the state space and behavior space of the market into a discrete lattice model. Assuming the current state is s_t, which represents the state of the market, investors can choose a series of investment behaviors a_t, such as buying, selling or holding. The goal of Q-learning is to measure the long-term return of taking a certain behavior in a given state by learning a state-behavior value function $Q(s_t, a_t)$. This value function is updated by the following formula:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \\ \gamma \max Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$$

Among them, α is the learning rate, which controls the impact of new information on existing values; r_{t+1} is the immediate return after taking action s_t from state a_t; γ is the discount factor, which determines the weight of future returns; and $\max Q(s_{t+1}, a_{t+1})$ represents the expected return of choosing the best action in the next state s_{t+1}. Through continuous iterative updates, Q-learning can eventually learn an optimal decision-making strategy, allowing investors to take the most favorable investment actions in different market conditions.

In order to better adapt to the actual asset management environment, we have improved the standard Q-learning. First, the state space and behavior space of the market are continuous, while Q-learning usually assumes that these spaces are discrete. To solve this problem, we use a deep Q network (DQN) to approximate the value function. DQN uses a deep neural network to approximate the state-behavior value function $Q(s, a)$, making it possible to handle high-dimensional state and action spaces. By using experience replay and target networks, DQN is able to reduce the estimation bias of Q values, thereby improving the stability and efficiency of learning.

In addition, we introduced a risk control mechanism in Q-learning. The goal of asset management is not only to maximize investment returns but also to ensure that risks are controlled while achieving returns. In order to quantify the risk of investment behavior, we define the risk-return ratio R_{risk}, which represents the ratio of the volatility σ of the portfolio to the return given the return r. The formula is as follows:

$$R_{risk} = \frac{r}{\sigma}$$

In the objective function of Q-learning, we incorporate this risk control indicator to ensure that the algorithm can balance the relationship between risk and reward during the learning process. Specifically, we modified the reward function of Q-learning and added a penalty term for the risk-reward ratio so that the risk level remains within an acceptable range while the reward is high. The final objective function can be expressed as:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} - \lambda R_{risk} + \gamma \max Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$$

Among them, λ is the weight parameter of risk control, which is used to balance the relationship between return and risk. By adjusting λ, we can control the risk tolerance and ensure that the algorithm can effectively learn the optimal strategy under different risk preferences.

Through the above improvements, our Q-learning algorithm can dynamically adjust the investment portfolio in an uncertain market environment to achieve a balance between risk control and return maximization. In practical applications, the algorithm can continuously optimize investment strategies based on historical market data and real-time market changes, thereby providing investors with an intelligent, automated asset management tool. As the algorithm training deepens, it can gradually adapt to market changes and capture potential laws in the market, thereby achieving more accurate risk control and return optimization.

3. Experiment

3.1 Datasets

In this study, we used a real historical dataset provided by Yahoo Finance to train and evaluate the Q-learning algorithm. The dataset used contains price information of multiple assets, including historical closing prices, trading volumes, volatility, and other data of stocks, bonds, and other financial instruments. The time span of the dataset is 5 years, covering different economic cycles of the market, and the data frequency is daily, ensuring sufficient sample size and diversity of time series. By analyzing these data, we can construct the market state space and associate it with investment decision-making behavior, so as to achieve effective training of the Q-learning algorithm.

To ensure the quality and accuracy of the data, we preprocessed the dataset in detail, including missing value filling, outlier detection, and standardization. Considering the risk control needs in asset management, we also pay special attention to indicators related to market volatility, such as price volatility and volume changes, which are of great reference value for risk assessment and return prediction. In addition, the asset types and market characteristics in the dataset are highly representative, so they can effectively reflect the investment behavior and risk control needs in the actual market. Through these real market data, we can simulate the decision-making process of investors in different market environments and apply the Q-learning algorithm to actual asset management.

This dataset is not only highly timely and market representative, but also provides a rich experimental scenario for our research. Through in-depth analysis of the data, we can test the performance of the Q-learning algorithm under different market conditions and verify its adaptability to risk control in a complex environment. In short, using this real financial market dataset can not only ensure the authenticity and practicality of the research, but also improve the reliability and effectiveness of the algorithm in practical applications.

3.2 Experimental Results

In this study, in order to comprehensively evaluate the performance of the Q-learning-based asset management risk control algorithm, we conducted a comparative experiment. By comparing with the existing classic asset management algorithms, we can fully verify the performance advantages of the Q-learning algorithm we proposed in different market environments. We selected four commonly used models as comparison objects: the traditional mean-variance optimization model (MVO)[6], the asset allocation model based on the genetic algorithm (GA)[7], the deep reinforcement learning model (DQN)[8], and the risk control model based on the support vector machine (SVM)[9]. These models each represent different portfolio optimization and risk control methods, which can provide strong comparative support for the advantages of the Q-learning algorithm.

In the experiment, we used three reinforcement learning evaluation indicators to evaluate the performance of the model. The first evaluation indicator is the cumulative return, which represents the total return of the portfolio during the entire test period. The cumulative return can intuitively reflect the comprehensive performance of the model in risk control and return optimization. The second indicator is the Sharpe Ratio, which is used to measure the excess return per unit of risk. The higher the Sharpe ratio, the higher the return the model can achieve while controlling risk. Therefore, it is a commonly used standard for measuring portfolio performance. Finally, drawdown is an indicator of the maximum loss of an asset portfolio, which is mainly used to evaluate the risk level of the portfolio in the worst case. A smaller drawdown means that the model can effectively control risk and avoid large capital losses.

By comparing with these models, we can comprehensively evaluate the performance of the Q-learning algorithm in actual asset management and verify its adaptability under different risk preferences and market conditions. At the same time, using these reinforcement learning evaluation indicators can help us more objectively compare the advantages and disadvantages of different models and provide a basis for further optimizing the Q-learning algorithm. The experimental results are shown in Table 1.

Table 1. Experimental results

Model	**Cumulative Return**	**Sharpe Ratio**	**Drawdown**
MVO	12.45%	0.89	-22.3
GA	15.32%	1.01	-18.7
DQN	18.25%	1.15	-16.4
SVM	14.87%	0.95	-20.2
Q-learning(Ours)	22.78%	1.28	-12.1

From the experimental results, the asset management risk control algorithm based on Q-learning has shown superior performance in all evaluation indicators. First, from the perspective of cumulative return, the return of the Q-learning model reached 22.78%, which is significantly higher than other comparison models. This shows that under the same market conditions, the Q-learning algorithm can achieve higher returns. This result proves the advantage of the Q-learning model in asset allocation, which can better capture market opportunities and optimize portfolio returns. In contrast, the return of the traditional mean-variance optimization model (MVO) is only 12.45%, which is the lowest level among all models, showing its limitations in a complex market environment.

Secondly, the Sharpe ratio is an important indicator for measuring risk-adjusted returns. The Sharpe ratio of the Q-learning model is 1.28, the highest among all models. This means that the Q-learning algorithm can not only obtain higher returns, but also perform well in risk control and can obtain more excess returns under unit risk. In comparison, the Sharpe ratio of the DQN model is 1.15, which is also good, but still lower than the Q-learning model. This shows that although DQN also performs well in terms of returns, it is slightly inferior to Q-learning in terms of risk control. The Sharpe ratios of the traditional MVO and SVM models are 0.89 and 0.95, respectively, which are significantly lower than those of Q-learning and DQN, indicating that the risk-adjusted returns of these two traditional methods are low, especially when dealing with complex risk control tasks, they are not as effective as models based on reinforcement learning.

In terms of drawdown, the drawdown of Q-learning is -12.1%, which is much lower than all other models. Drawdown reflects the maximum loss of an investment portfolio during market fluctuations. A smaller drawdown value means that the model can better protect investors' capital when encountering a market downturn. Q-learning can effectively avoid extreme market risks and avoid large capital losses by dynamically adjusting the market status through the reinforcement learning algorithm. Other models, such as MVO and SVM, have drawdowns of -22.3% and -20.2%, respectively, indicating that these traditional methods have poor risk control capabilities during market fluctuations and are prone to large losses. Especially MVO, due to its over-reliance on the mean and variance of historical data, often cannot effectively cope with the sharp fluctuations in the market, resulting in large drawdowns.

In contrast, the genetic algorithm (GA) showed a strong return ability, with a cumulative return of 15.32%, higher than the MVO and SVM models, but still lower than DQN and Q-learning. However, the drawdown of the genetic algorithm was -18.7%, which was better than MVO and SVM, but still much higher than Q-learning. This shows that the genetic algorithm can optimize asset allocation to a certain extent, thereby improving returns, but there are still certain defects in risk control, especially when the market fluctuates violently, the drawdown is still large. Therefore, although GA has achieved certain advantages in terms of returns, its risk control effect is still not comparable to Q-learning based on reinforcement learning.

As a deep reinforcement learning algorithm, DQN achieved a return of 18.25%, ranking third among all models. DQN approximates the Q-value function through a deep neural network, can better adapt to the complex asset management environment, and performs well in return optimization. However, despite DQN's good performance in terms of returns, its drawdown is -16.4%, which is still higher than the Q-learning model. This shows that although DQN can learn better investment decision-making strategies during training, it still has certain shortcomings in controlling market risks and avoiding large losses. In contrast, Q-learning, through its state-behavior based learning mechanism, pays more attention to risk avoidance, so it can effectively control drawdowns while improving returns.

In summary, Q-learning has shown obvious advantages in all evaluation indicators, especially in terms of return and drawdown control. The Q-learning model continuously adjusts its decision-making strategy through reinforcement learning, and can find the optimal investment behavior in a changing market environment, thereby obtaining higher returns and effectively controlling risks. Other comparative models, such as MVO, GA, DQN and SVM, although they perform differently in some indicators, cannot compare with Q-learning in terms of risk control and drawdown. The experimental results fully verify the application potential of the Q-learning algorithm in asset management and prove its superiority in asset allocation optimization and risk control. Future research can further optimize the Q-learning model and combine more market data and risk control strategies to improve its performance in more complex market environments.

4. Conclusion

This study proposed an asset management risk control algorithm based on Q-learning, and verified its superiority through comparative experiments with traditional models. The experimental results show that Q-learning not only performs well in return optimization but also shows obvious advantages in risk control, especially in the control of maximum drawdown, Q-learning can effectively reduce the risk of investment portfolios. This provides a new idea for the field of asset management, especially in complex and volatile market environments, Q-learning can adaptively adjust investment strategies and provide more stable and efficient asset allocation.

Although this study has achieved certain results, there are still some aspects that need further improvement. First, the training process of the Q-learning algorithm may be limited by data quality and data volume, and may face more complex market environments in practical applications. Future research can further improve the accuracy and robustness of the model by introducing more market characteristics and external macroeconomic data. In addition, the computational complexity of the reinforcement learning algorithm is high, especially in the training stage, which may require a lot of computing resources and time. In the future, more algorithm optimization methods can be explored, such as deep Q-learning (DQN) and hierarchical Q-learning, to improve the efficiency and adaptability of the algorithm.

In future work, the performance of Q-learning may be further enhanced by combining it with other machine learning methods, such as deep learning and evolutionary algorithms. For example, one can try to use models such as convolutional neural networks (CNN) or long short-term memory networks (LSTM) to extract more complex market features, thereby improving the decision-making ability of Q-learning under multi-dimensional data. In addition, considering the uncertainty and dynamics of the market, studying how to achieve online learning and incremental learning of the Q-learning algorithm will also be an important research direction.

Finally, future research can also consider combining the Q-learning algorithm with modern financial instruments and asset management strategies, such as quantitative trading strategies and investment decision-making systems of hedge funds. This can not only further expand the scope of application of Q-learning in the financial field but also verify its feasibility and effectiveness in real transactions through back testing and optimization in the actual market. With the development of technology and the continuous changes in the market environment, Q-learning and its extended methods are expected to become important tools in the field of asset management and risk control.

References

[1] Jiang G, Zhao S, Yang H, et al. Research on Finance Risk Management based on Combination Optimization and Reinforcement Learning[J]. 2024.

[2] Cannelli L, Nuti G, Sala M, et al. Hedging using reinforcement learning: Contextual k-armed bandit versus Q-learning[J]. The Journal of Finance and Data Science, 2023, 9: 100101.

[3] Massahi M, Mahootchi M. A deep Q-learning based algorithmic trading system for commodity futures markets[J]. Expert Systems with Applications, 2024, 237: 121711.

[4] Oyewola D O, Akinwunmi S A, Omotehinwa T O. Deep LSTM and LSTM-Attention Q-learning based reinforcement learning in oil and gas sector prediction[J]. Knowledge-Based Systems, 2024, 284: 111290.

[5] Xu S, Zhang G, Yuan X. An Enterprise Multi-agent Model with Game Q-Learning Based on a Single Decision Factor[J]. Computational Economics, 2023: 1-40.

[6] Pham L. Stocks portfolio optimization based on forecasting through the integration of the MVO method and LSTM model[J]. 2023.

[7] Yi X. Application of data mining in enterprise financial risk prediction based on genetic algorithm and linear adaptive optimization[J]. Soft Computing-A Fusion of Foundations, Methodologies & Applications, 2023, 27(14).

[8] Sahu S K, Mokhade A, Bokde N D. An overview of machine learning, deep learning, and reinforcement learning-based techniques in quantitative finance: recent progress and challenges[J]. Applied Sciences, 2023, 13(3): 1956.

[9] Kurani A, Doshi P, Vakharia A, et al. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting[J]. Annals of Data Science, 2023, 10(1): 183-208.