Stock Portfolio Optimization Based on Reinforcement Learning

Jinglong Li

School of economics, Beijing International Studies University, Beijing, 100024, China

*Corresponding author Email:1459343592@qq.com

ABSTRACT. This paper made a profound study of the application of reinforcement learning in portfolio optimization, using deep learning algorithm, combine various indicators, and analyze the explanatory variables that can effectively improve portfolio risk control through multi-dimensional financial indicators and statistical indicators. Designing a reasonable and effective value function from the reward and punishment mechanism to achieve the optimization goal of income maximization and risk control, mining problems from the perspective of practice, and the research results is of great significance for portfolio management.

Keywords: Stock portfolio optimization, Reinforcement Learning, financial indicators and statistical indicators, value function.

1 Introduction

1.1 Research Background

Deep learning is mainly a method of analyzing and modeling research problems by using deep neural network, which was first proposed by G.E. Hinton of the University of Toronto in 2006.[1] Deep learning (DL) in the field of machine learning has realized the functions of image recognition, audio recognition, natural language processing and so on, which reflects the ability of deep learning in information perception. Reinforcement learning (RL) is another development achievement of artificial intelligence. [2] At present, it is widely used in machine control, robot and other fields. The development goal of artificial intelligence is to realize an agent that can observe environmental information and think and make decisions independently. Portfolio optimization and option pricing are the basic problems of financial mathematics, which have extensive and important applications in the field of asset management and investment decision-making. The classical mean variance optimization model and Black Scholes Merton option pricing model depend on the special probability distribution sensitive to the model parameters. The application of deep learning in the financial field is actually a hot topic in recent years.

F. Cao et al. (eds.), *Proceedings of the 2023 5th International Conference on Economic Management and Cultural Industry (ICEMCI 2023)*, Advances in Economics, Business and Management Research 276,
https://doi.org/10.2991/978-94-6463-368-9_16

1.2 Research Significance

This project explores the optimization of financial product portfolio by deep learning, which belongs to the category of the combination of asset portfolio and deep learning technology. The main method is to build a new factor combination based on the factor model to achieve higher yield and less pullback. The frontier of factor investment lies in mining new factors and optimizing asset portfolio. Using machine learning and deep learning for data mining and feature engineering is a relatively direct application at present. It can help solve the problems of nonlinear and high-dimensional data. Neural network has stronger information extraction ability and can better complete the tasks of pattern recognition and signal extraction. In the academic community, the more promising direction is to integrate deep learning into the framework of factor investment, resulting in directions such as deep factor model and deep portfolio.[3]

1.3 Previous Successful Research Results

The leading company in the application of deep reinforcement learning model is Google's deep mind company, which lays the foundation of deep reinforcement learning model. The paper published by mnih in 2013 laid a theoretical foundation for deep reinforcement learning. The paper mainly overcomes the problems of data correlation and non-stationary distribution. Its approach is to randomly sample from the previous state transition matrix for training. This has at least two advantages: 1 Data utilization is high because a sample is used multiple times. 2. The correlation of continuous samples will make the variance of parameter update larger, and this mechanism can reduce this correlation In 2015, mnih continued to improve its 2013 article.[4] The main change in algorithm is the introduction of a separate q-function network. This paper proposes an iterative update method, which makes the parameters of Q function update only after a certain number of steps, which is equivalent to delaying the update to reduce the correlation between Q function and Q function objectives. In 2016, silver disclosed the research on the application of reinforcement learning algorithm combined with artificial neural network to go program. The most well-known is that alpha go robot challenged world champion Li Shishi and finally won.

1.4 Application of Reinforcement Learning in Financial Field

Stock price analysis and prediction is the most challenging problem in financial problems. Because the financial market risk is high, and it is often difficult to simply predict, and there are too many dynamic and uncertain factors. Fundamental analysis and technical analysis are the two most mainstream analysis modes to study the stock market. In addition to these two analysis models, stock prices are affected by many factors, such as economic policy, current news, political events and investor sentiment. Each form of analysis has its own attributes to analyze or predict stock prices. Technical analysis is a study of past market data through the analysis method of market forecast price. Technical analysis will use several attributes of historical market data, such as date, opening price, closing price, highest price, lowest price, quantity and other technical data.[5] Fundamental analysis includes studying the overall economic and social conditions, industry conditions, financial conditions and company management to predict the price fluctuation of stocks. In addition, the combination of two analysis methods can be used in the prediction process, such as technical analysis and investor sentiment analysis, which will produce a more effective prediction method.[6]

For the financial field, reinforcement learning is also involved, but generally speaking, there is not much relevant literature. For financial problems, uncertainty and dynamics are the difficulties, and reinforcement learning has great advantages in solving such problems. Reinforcement learning model can be applied to securities trading, especially high-frequency trading and portfolio management. Many scholars use deep learning algorithm to design stock investment framework, conduct simulation trading, and use the real market situation for back testing, which shows that this method is feasible. Practice shows that applying deep reinforcement learning to digital currency investment, CNN, RNN and LSTM networks are used as Q-value networks of deep reinforcement learning respectively, and good results are achieved by training and investment with minute K-line data. The above methods provide new ideas and vision for the research and development of this paper. [7-8]

2 Contribution of This Paper

It is an important trend of financial market to use algorithm to carry out stock quantitative trading. In many complex games such as chess and go, deep reinforcement learning (DRL) agents have made amazing achievements. The theory of deep reinforcement learning is also applicable to the quantitative decision-making of stock market. However, the financial market is unpredictable and uncertain. At present, no deterministic model can accurately describe the changes of the financial market, and no strategy can win forever in the financial market. Facing the complex market, asset managers need to constantly learn, summarize and optimize their investment strategies in the market. The deep reinforcement learning model can learn continuously in the continuous interaction with the market, and adjusting its own strategy is one of the feasible methods to deal with this changeable market. Deep mind creatively combines deep learning (DL) and RL to form a new research hot-spot in the field of artificial intelligence, namely deep reinforcement learning (DRL). By using the decision-making ability of DL and RL, the agent trained by DRL can be comparable to or even surpass human beings. DRL has recently been applied to solve challenging problems, including Atari games, man-machine confrontations, and mobile crowd sensing. In DRL, the agent we train must interact with the environment to adapt to the environment, so that the agent takes a long time to achieve the best expectation. In this process, the agent learns how to adjust constantly in the environment according to the reward function we give. In order to accelerate the learning process, we always hope that the reward function can record and guide the state of the agent timely and accurately. Therefore, the design of reward function has become a key aspect of RL. Through in-depth study of the application of reinforcement learning in portfolio, this paper designs a reasonable and effective value function from the reward and punishment mechanism to achieve the optimization goal of risk control. For stock price forecasting, the fundamental research of a company is essential, and financial indicators are the top priority. It is a relative index to evaluate the operation status of the company and the current situation of the enterprise. In China's A-share market, we study the explanatory ability of financial indicators for stock excess return, and find that the financial indicators reflecting the company's development ability have strong guiding significance for stock investors. However, there are still some problems, such as unclear selection range of financial indicators and deviation in the definition of investment time. Therefore, we use deep learning algorithm combined with various indicators to analyze the explanatory variables that can effectively improve portfolio risk control through multi-dimensional financial indicators and statistical indicators. Solve the above problems. From the perspective of practice, the research results will have guiding significance for portfolio management.

3 Math and Equations (Model)

3.1 Reinforcement Learning Algorithm

We prescribe the environment as the Markov decision-making process: M = {S, A, P, γ, R}, where

(1) $S \in R^m$ is the space of **observed states**. At each step the exchange-agent system is in $s_t \in S$.

(2) A =$[a_1, a_2]$ is the space of **actions**. In the trading problem, $a_1 = \{1,2,3\}$, $a_2 \in R^+$. When $a_1 = 1$, the agent buys a_2 shares of stocks; when $a_1 = 2$, the agent sells a_2 shares of stocks; while $a_1 = 3$, the agent takes non actions.

At each step we choose the action $\mathrm{a_t} \in \mathrm{A}$ from the developed politics $\pi(\mathrm{a}|\mathrm{s})$, that is, the probability to choose a in s.

(3) $P(s'|s, a)$ is the **transition probability** of the assumed Markov process.

(4) $\boldsymbol{R(s, a)}$ is the reward function. At each step the agent becomes the reward depending, not only on the current, but also on the previous actions $r_t = \boldsymbol{R}(s_t, a_t)$.

(5) $\gamma \in [0, 1]$ is the decay multiplier with which the next reward is summed up into the total reward for an action $R_t = \sum_{i=0}^{T} \gamma^i r_{t+i}$.

Then the task of the algorithm is to find the strategy $\pi: S \rightarrow A$ that maximizes the mathematical expectation of reward ρ^π:

$$\rho^\pi = \rho^\pi(\theta) = E[R|\pi] \rightarrow max_\pi$$

$\rho^\pi = E[R|\pi(\theta)] = \int_T p(\tau|\pi(\theta))R(\tau)d\tau \rightarrow max_\theta$, where the track $\tau = \{s_t, a_t\}_{t=0}^{T}$ is the realization of one game and $R(\tau)$ is the total reward.[9]

3.2 Index Construction

First, we download the daily data of all stocks since the establishment of A shares on February 10, 1991 from Ruisi database. In order to facilitate calculation and operation, we convert it into separate monthly data of each stock, with month as the minimum operating unit. Then,we simply select a group of financial indicators randomly as parameters to test the impact of value function on agent. We set three different groups of value functions and test its accuracy based on ACC(table 1).[10] The specific results are shown in the table below. We can see that simple reward and punishment changes have no essential impact on the accuracy of agent processing data, and the change of ACC is only about 0.2%.

Table 1. Accuracy of value function

Value function	ACC
(1,1)	19.33%
(1,10)	19.21%
(1,100)	19.40%

Next, we use the financial indicators to construct five different groups of financial indicator parameters(table 2), which are arranged and combined to find the optimal solution.[11] At the same time, ACC is still used as the benchmark to test the accuracy of agent. In the continuous test, we find that the accuracy of a group of parameters is quite outstanding, so we consider increasing the number of parameters on this basis to make its accuracy higher. However, in fact, continuing to stack the number of parameters can not improve the accuracy, which may be due to the limitation of the operation time of monthly data and the autocorrelation between parameters.[12]

Table 2. Build optimal parameter array

parameters	outputA	outputB	outputC	outputD	outputE
OpeprfPS[1]	√	-	-	-	-
Opeprfrt[2]	√	-	-	√	-
TORGrRt[3]	√	-	-	-	-
Susgrrt[4]	√	-	-	√	-
EPSTTM[5]	-	√	-	-	-
Qckrt[6]	-	√	-	-	-
Netassgrrt[7]	-	√	-	-	-
AvgROE[8]	-	√	-	-	-
ROIC[9]	-	√	-	-	-
ROAEBIT[10]	-	-	√	√	√
Currt[11]	-	-	√	√	√
EntireliaEBITDA[12]	-	-	√	√	√
OpeCPSgrrt[13]	-	-	√	√	√
EqumulDP[14]	-	-	√	√	√
NAPSgrrt[15]	-	-	-	-	√
TNonCurLiaSEWMI[16]	-	-	-	-	√
ACC	19.40%	9.46%	39.14%	19.05%	30.17%

Notes:

1. OpeprfPS[1]=Operating profit=(main business profit + other business profit - operating expenses - administrative expenses - financial expenses) / Total share capital
2. Opeprfrt[2]=Operating profit margin=Operating profit / revenue (sales of goods)*100%.
3. TORGrRt[3]=Year on year growth rate of total operating revenue=(current operating income - previous operating income) / previous operating income*100%.
4. Susgrrt[4]=Sustainable growth rate=Net interest rate of equity at the end of the period*Current profit retention rate / (1 - net interest rate of equity at the end of the period)*Current profit retention rate).
5. EPSTTM[5]=Earnings per share=(current gross profit - preferred stock dividend) / total share capital at the end of the period.
6. Qckrt[6]=Quick ratio=quick assets / current liabilities.
7. Netassgrrt[7]=Growth rate of net assets= increase in net assets in the current period / total net assets in the previous period * 100%.
8. AvgROE[8]=Return on net assets = net profit / shareholders' equity at the end of the year * 100%.
9. ROIC[9]=Return on investment capital=after tax operating income / (total property - Excess Cash - interest free current liabilities).
10. ROAEBIT[10]=Return on total assets=(total profit + interest expense) / average total assets*100%.
11. Currt[11]=Current ratio=current assets / current liabilities *100%.
12. EntireliaEBITDA[12]=Total debt / EBITDA.
13. OpeCPSgrrt[13]=Growth rate of operating cash flow per share = (operating cash flow per share in the current period - operating cash flow per share in the previous period) / operating cash flow per share in the previous period * 100%.
14. EqumulDP[14]=Equity multiplier = total assets / total shareholders' equity.
15. NAPSgrrt[15]=Growth rate of net assets per share = annual growth of net assets per share / net assets per share at the beginning of the year* 100%.
16. TNonCurLiaSEWMI[16]=Long term capital liability ratio = [non current liabilities / (non current liabilities + shareholders' equity)]*100%.

In the total 232 financial indicators, we selected the above 16 representative indicators and combined them according to their respective properties. For example, profit indicators, debt service indicators and development indicators form a group of parameters, and their ACC is calculated through deep machine learning.

Through comparison, we can find that the ROAEBIT[10] reflects the profitability of the owner's investment. Currt[11] and EntireliaEBITDA[12] can measure the company's ability to repay debts. OpeCPSgrrt[13] can help us analyze the real financial situation of the enterprise and avoid false prosperity from affecting our judgment to a certain extent. reflects the asset allocation of the company, EqumulDP[14] and can also reflect the long-term financial situation and solvency of the enterprise from another side. These five financial indicators are combined to build the parameter group with the best winning rate of 39.14%.

After we got a decent result as Output C, Whether we can further improve the accuracy by increasing the number of parameters becomes our primary purpose.So in the next two groups of parameters, we added a group of profit indicators and a group of sustainable operation indicators on the basis of Output C. However, we can clearly see from the data that the results obtained are not ideal, and even ACC has a significant retreat.

4 Conclusion

This paper explores the impact of financial indicators and financial data on stock prices, in order to find a set of optimal parameters to guide the construction of portfolio with the support of deep machine learning algorithm. In the process of practice, we find that simply changing the upper and lower limits of value function can not effectively affect the final accuracy. The key point is to build a reasonable and effective parameter group for agent learning. In the process of continuous trial and training, we found a set of ideal parameters. From the perspective of finance, this set of parameters also explains the profitability and solvency of a company in all aspects. The agent trained through this set of parameters has a winning rate of 39.14% in the transaction. Although it does not reach a very amazing accuracy rate, the construction of portfolio must not only focus on the financial indicators. Therefore, the methods and ideas of this paper have certain limitations. I hope we can improve these places in the follow-up research.

References

1. Y. Deng, F. Bao, Y. Kong, Z. Ren and Q. Dai, March 2017,"Deep Direct Reinforcement Learning for Financial Signal Representation and Trading," in IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 3, pp. 653-664.
2. Yuqin Dai, Chris Wang, Iris Wang, Yilun Xu, "Reinforcement Learning for FX trading"Chien Yi Huang. Financial trading as a game: A deep reinforcement learning approach. arXiv preprint arXiv:1807.02787, 2018.
3. Almahdi, S.; and Yang, S. Y. 2017. An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. Expert Systems with Applications 87: 267–279.
4. Deng, Y.; Bao, F.; Kong, Y.; Ren, Z.; and Dai, Q. 2016. Deep direct reinforcement learning for financial signal representation and trading. IEEE transactions on neural networks and learning systems 28(3): 653–664.
5. Grinblatt, M.; Titman, S.; and Wermers, R. 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. The American economic review 1088–1105.
6. Xu, L.; and Cheung, Y.-m. 1997. Adaptive supervised learning decision networks for traders and portfolios. In Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr), 206–212. IEEE.
7. Jegadeesh, N.; and Titman, S. 1993. Returns to buying winners and selling losers: Implications for stock market effificiency. The Journal of fifinance 48(1): 65–91.

8. Poterba, J. M.; and Summers, L. H. 1988. Mean reversion in stock prices: Evidence and implications. Journal of financial economics 22(1): 27–59.
9. D. Zhang and S. Lou, “The application research of neural network and BP algorithm in stock price pattern classification and prediction,” Futur. Gener. Comput. Syst., vol. 115, pp. 872– 879, Feb 2021, doi: 10.1016/j.future.2020.10.009.
10. M. Ahmad, H. Soeparno, and T. A. Napitupulu, “Stock trading alert: With fuzzy knowledgebased systems and technical analysis,” in 2020 International Conference on Information Technology Systems and Innovation, ICITSI 2020 - Proceedings, Okt 2020, pp. 155–160, doi: 10.1109/ICITSI50517.2020.9264914.
11. K. Kaczmarczyk and M. Hernes, “Financial decisions support using the supervised learning method based on random forests,” Procedia Comput. Sci., vol. 176, pp. 2802–2811, 2020, doi: 10.1016/j.procs.2020.09.276.
12. “Short Term Stock Price Prediction Using Deep Learning,” RTEICT-2017 2nd IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol. Proc. 19-20 May 2017, 2017.