The 6th International Conference on Economic Management and Green Development (ICEMGD 2022)
DOI: 10.54254/2754-1169/3/2022849

A Model for Quantifying Investment Decision-making Using Deep Reinforcement Learning (PPO Algorithm)

Xiaochen Xiao[1,a,*], Weifeng Chen[2]

[1]*PUniversity ofElectronic Science andTechnology of China, Zhongshan Guangdong 528402, China*
a. 971851309@qq.com
**corresponding author*

Abstract: Bitcoin has nowadays shown its importance in finance systems. With great interest to all kinds of investors in the financial sector, it is important to analyse the relationship between Bitcoin and gold. We consider bitcoin and gold as two stocks and calculate the correlation between bitcoin and gold. By introducing the calculation of dynamic penalty coefficient, the problem of dual-stock portfolio investment is transformed into the problem of single-stock purchase investment, which greatly reduces the difficulty of feature engineering and model application. In terms of decision-making model, deep reinforcement learning (PPO algorithm) is used to make quantitative investment decisions. Therefore, we use the expected data in SLTM as the input data of deep reinforcement learning, and combine it with deep reinforcement learning for training. Compared with machine learning to quantify investment decisions, after a period of training, the accuracy rate has improved by 10.038%.

Keywords: Deep reinforcement learning; Quantitative investment; Time series analysis; PPO (Proximal policy optimization); (Recurrent Neural Network)LSTM algorithm; Gold; Bitcoin.

1. Introduction

It is estimated that by 2023, China's per capita disposable income will reach 210 trillion Chinese Yuan (CNY). At the same time, the domestic economic downturn, the devaluation of CNY and low-interest rates have led the residents to diversify their wealth allocation and turn to financial assets. China's asset management industry will achieve a larger market size and rapid development. As an important participant in global wealth management, we believe that we will help our customers achieve in-depth development of wealth management in the near future. Improvement of the asset management level in China to meet the diversified investment needs of customers is an urgent target to reach. With its regularity, systematicness, accuracy, dispersion and timeliness, quantitative investment plays an increasingly important role in China's investment and financing system.

In this case, the application of quantitative investment is significant. Quantitative investment is a trading mode with quantitative statistical analysis tools as its core and programmed trading. Chincarini pointed out that quantitative investment follows the following concepts. First, the market is effective. Secondly, it is statistically significant to quantify arbitrage opportunities based on investment strategies. Thirdly, quantitative investment analysis should be supported by solid logic and theoretical foundation. Fourth, the model should be sustainable and stable. Fifth, the risk must be kept at a low enough level so that the excess return can be meaningful.

© 2023 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/3/2022849

Trend trading is to judge whether a trend is completed by calculating the price evolution, but the specific point when the trend occurs is unknown. Based on the tracking changes in the market price system, the machine-controlled accounts will systematically track the trend above the intermediate level. When the trend is formed, the trading strategy will adjust and change around the trend. Trend trading is based on the trend theory, according to which the corresponding trend line is drawn.

1.1 Research Purpose and Significance

This paper aims to provide an automatic trading scheme for Bitcoin and Gold portfolio investment.

With the continuous development of artificial intelligence, the fields involved in the AI industry are gradually expanding. At the same time, due to the development of China's financial market and people's increasing demand for wealth management, financial investment strategies, financial measurement models, and other essential tools have gradually become people's wealth management investment, making the artificial intelligence industry become the critical direction of asset investment. Asset evaluation is the inevitable result of the development of the market economy to a particular stage. Asset evaluation of listed companies helps make an accurate judgment on investment decisions. The analysis of assets evaluation in the artificial intelligence industry is helping to make up for the deficiency of assets evaluation regulations in China. From a macro point of view, the analysis of AI concept stocks portfolio in the AI industry can find the shortcomings of the current AI industry and improve it, which is helpful to the development of the AI industry.

Since COVID-19 ravaged the world in 2019, the "black swan" incidents in the neo-liberal world have emerged. It can be predicted that the future world will present a more turbulent scene-no matter where you are. The military will usher in a local and low-intensity war, but it is enough to put normal countries into an emergency. Commercial development will usher in a low ebb. As an emergency safe-haven asset, gold will become the darling of investors. However, the high-risk and high-yield features of Bitcoin are still full of allure. Compared with gold, Bitcoin has the characteristics of high yield, high volatility, no supervision, no taxation, etc. It will have huge development space in the financial sector. Bitcoin is sometimes referred to as the new type of gold, which can replace gold to hedge against inflation and become a new type of safe-haven asset that can complement gold for hedging. It is of great significance to provide a solution for the quantitative transactions between Bitcoin and gold.

1.2 Domestic and Foreign Similar Research

With the development of deep learning technology, the related research in quantitative investment is also continuously developing. We witnessed the research process from simple models such as BP and SVM neural network to convolution neural network, LSTM neural network and embedded model.

The earliest application of neural network theory in stock forecasting was White, who made a forecast of IBM's earnings per share but failed. Diler and others tried to apply the neural network to forecast the stock index price and achieved good results.[1].

According to Zhou Xu's research (the application of Bollinger Band trend breakthrough strategy in the digital currency market), because the trading mechanism of the foreign exchange market is similar to that of the digital currency market, the Hurst index analysis shows that digital currency's foreign exchange market is more trend-oriented. The trend trading strategy has application prospects in the digital currency market. It is completely feasible to make a trend trading on the trend of bitcoin through the algorithm [2].

According to Su Chunlin and others (the application research of the multi-factor model in the digital currency market), the traditional multi-factor back-calculation method is effective in Bitcoin transactions [3].

DOI: 10.54254/2754-1169/3/2022849

Zhang Jing proposed in applying the multi-factor model for quantitative investment that the single model strategy is challenging to achieve both the objectives of increasing excess returns and reducing retracement. By trying to operate the multi-factor model simultaneously, it is easier to catch different investment opportunities in the market [4].

The research scope of the above research is still the traditional quantitative investment without using a machine learning algorithm. Although it is still valid, it is slightly outdated for academic research.

Meng Ye et al. used technical analysis and fundamental analysis to construct a fundamental data classifier by combining affinity propagation in data mining with Adaboost, an integrated learning algorithm, in a safe quantitative stock selection strategy—an empirical study from the A-share market, and iteratively screened out the stock portfolios with rising potential to formulate a stock selection strategy. Simulated transactions have tested this strategy from 2015 to 2017. The results show that the annualized income of this strategy outperforms the extensive market index, and the success rate in-stock selection is high, which provides a new idea for quantitative stock selection [5]. This research provides a theoretical basis for introducing machine learning algorithms into a quantitative investment, but it does not use the neural network. Although it is practical, it is somewhat outdated for academic research.

According to the research of Huang Zepeng and others (modelling and application of gold price prediction based on deep learning), the gold price is a non-stationary time series which is extremely sensitive to the external market, and the general time series model and shallow learning model cannot accurately grasp its changing characteristics[5]. A better way to describe the complex change in gold price is to establish Elman neural network with a multi-layer network structure such as delay operator acceptance layer structure and double hidden layers. This research provides a theoretical basis for adding the neural network to quantitative investment. We introduce the BP neural network on this basis and compare it to the PPO.

It can be seen that the application of neural network in the field of quantitative investment has made considerable progress. We have further introduced deep reinforcement learning into quantitative investment on this basis.

1.3 Machine Learning Overview

Machine learning is a multi-disciplinary cross-discipline which covers the knowledge of probability theory, statistics, approximation theory and complex algorithms. It uses computers to simulate human learning in real-time, decompose existing content into knowledge structures, and improve learning efficiency.

Stochastic parallel gradient descent algorithm, or SPGD algorithm for short. As a model-free optimization algorithm, it is more suitable for the optimization control process with more control variables and complicated controlled systems which cannot establish an accurate mathematical model. The principle of the SPGD control algorithm is mainly developed from the stochastic approximation (SA) theory and artificial neural network technology, and the adaptive optical correction technology developed by Vorontsov of the US Army Research Laboratory based on the simultaneous perturbation stochastic approximation control algorithm.

A loss function or cost function is a function that maps the value of a random event or its associated random variable to a nonnegative actual number to represent the "risk" or "loss" of that random event. In applications, the loss function is usually associated with the optimization problem as a learning criterion, i.e. solving and evaluating the model by minimizing the loss function. For example, it is used for parameter estimation of statistics and machine learning models, risk management and decision-making in macroeconomics, and optimal control theory.

DOI: 10.54254/2754-1169/3/2022849

LSTM is a time-recurrent neural network specially used to solve the long-term dependence problem of general RNN (Recurrent Neural Network). All RNNs have a chained form of repetitive neural network modules. In the standard RNN, this repetitive structure module only has a straightforward structure, such as a hyperbolic tangent function (tanh) layer.

2. Model construction

2.1 Assumptions

To introduce deep reinforcement learning into a quantitative investment, we must first understand the assumptions of quantitative investment.

The Efficient Market Hypothesis (EMH) is a popular theory with a long history. Its main content is that "market price reflects all known information." In an efficient market, no investor can consistently get a return higher than the market average for a long time. "

Specifically, the EMH is divided into three levels. The lowest "Weak-Form Market Efficiency" holds that investors cannot predict the future stock price trend with known publicly traded data. However, they can still use fundamental data to estimate the stock's intrinsic value and make profits or obtain stable excess profits through undisclosed inside information.

However, the strongest "Strong-Form Market Efficiency" holds that investors can not obtain excess profits stably regardless of whether they use public information or inside information.

The EMH is divided into three levels. Among them, the lowest "Weak-Form Market Efficiency" holds that investors cannot predict the future stock price trend with known publicly traded data but can still use the fundamental data to estimate the intrinsic value of the stock and make a profit, or obtain a stable excess profit through undisclosed inside information.

The most convincing "Strong-Form Market Efficiency" holds that investors cannot obtain excess profits stably, whether by using public information or inside information.

We want to make the following assumptions, where α is the transaction tax, and bitcoin does not have a tax but will generate a corresponding handling fee to support the miners' operation during the transaction.

1 α Gold =1%, α Bitcoin =2%.

2 No cost of assets held

3. The price fluctuation of USD notes in recent years is not considered.

2.2 Feature engineering

Figure 1: Bitcoin Data.

DOI: 10.54254/2754-1169/3/2022849

To treat Bitcoin and gold as one stock and reduce the difficulty of applying and calculating the model, we put forward the concept of penalty coefficient and regard Bitcoin and gold as two stocks to calculate the correlation between Bitcoin and gold. By introducing the dynamic penalty coefficient calculation, the double-stock portfolio investment problem is transformed into the single-stock subscription investment problem, significantly reducing the difficulty of feature engineering and model application.

2.2.1 Correlation analysis

According to the literature [6], for the correlation between gold and bitcoin, when one party is in an extreme market, the risk of contagion to the other party's market will be strengthened. In the downward market, gold is vulnerable to the impact of the bitcoin market. In the upward market, the bitcoin market is more vulnerable to the impact of the gold market. At a specific point in time, there is an exchange relationship between the correlations.

Figure 2: Gold Data.

According to the statistics of SPSS on the existing data, gold and Bitcoin are highly positively correlated in the upward market, with the correlation coefficient as below.

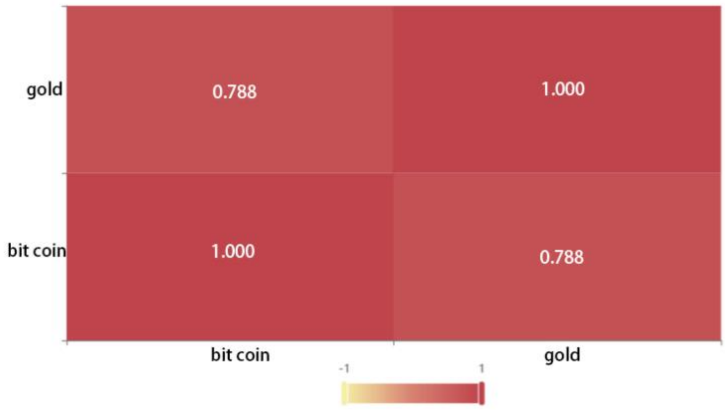

Figure 3: Bitcoin Correlation Heat Map.

DOI: 10.54254/2754-1169/3/2022849

Table 1: Correlation Analysis of Bitcoin and Gold.

correlation			
		gold	BTC
gold	Pearson correlationg	1	.299**
	Sig(Two-tail)		.000
	Number of cases	477	477
BTC	Pearson correlationg	.299**	1
	Sig(Two-tail)		.000
	Number of cases	477	477

The correlation coefficient is as high as 0.788, which indicates that Bitcoin is highly correlated with gold.

We regard Bitcoin and gold as two stocks. To reduce the difficulty of calculation, we need a way to transform the combination problem of the two stocks into the purchase problem of a single stock.

There is a particular dynamic relation between the two stocks. The combination investment problem of the two stocks is converted into the purchase investment problem of a single stock, which significantly reduces the application difficulty of the feature engineering and the model. The model's generalisation ability is improved from the perspective of data.

Concerning the penalty factor:

One is the penalty for the gold incident, and the other is the penalty for the overall international situation.

The penalty for the gold incident is an incident that affects the fluctuation of gold. Positive incidents are set to 1, invalid incidents to 0.7, the penalty for the international situation, positive incidents to 1 and negative incidents to 0.5.

When the penalty for the gold incident * the penalty for the overall international situation equals 1, it is determined to be an upward trend. At this time, all bitcoins will be bought if the gold price is stable.

When it is not equal to 1, it is determined to be a falling market. At this time, we buy gold.

We combined this operation with the following model to make it a completely automated trading system.

2.3 Model overview

Artificial Neural Networks (ANNs) are also called neural networks (NNs) or connection models. It is an algorithmic mathematical model to simulate the behavioral characteristics of animal neural networks and to perform distributed parallel information processing. This kind of network relies on the complexity of the system, and achieves the purpose of processing information by adjusting the interconnection relationship between a large number of internal nodes.

This is essentially a combination of multiple linear algorithm. For example, the logistic regression model is binary classification.

DOI: 10.54254/2754-1169/3/2022849

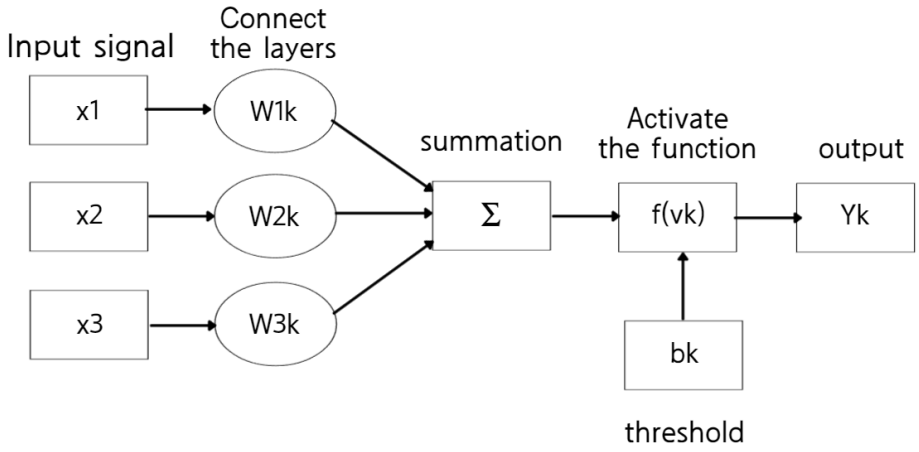

Figure 4 : Neural network structure.

2.4 Deep reinforcement learning (PPO) algorithm

PPO algorithm is a new policy gradient algorithm. The policy gradient algorithm is very sensitive to the step size, but it is difficult to select the appropriate step size. If there is too much difference between the old and new strategies in the training process, it is not conducive to learning. PPO proposes a new objective function, which can be updated in small batches in multiple training steps, thus solving the problem that the step size in the policy gradient algorithm is difficult to determine.

The policy gradient algorithm is reviewed again. The policy gradient does not use error back propagation. It selects an action by observing the information and transmits it back directly. Of course, what is surprising is that it is not wrong. Instead, rewards are used to directly increase and decrease the probability of selecting behaviors. Good behaviors will increase the probability of being selected next time, while bad behaviors will decrease the probability of being selected next time.

In the gradient policy, we need to continuously generate the number of samples:

$$E_{x\sim P}[f(x)] = \int_x p(x)f(x)d_x = E_{x\sim P'}\left[\frac{P}{P'}f(x)\right] \approx \frac{1}{N}\sum_{x_{i\sim P',i=1}}^{N}\frac{P(x)}{P(x)'}f(x) \tag{1}$$

Continuously sample x from the distribution of p, change x to f(X), and then find the desired value of f(X). Use the formula: f(X)p(X)DX=f(X)p(x))Q(X)DX, where p(x) can be used as the weight term.

$$D_{KL}(\pi_{\theta'}\|\pi_\theta) = E_{\pi_{\theta'}}[log\pi_\theta - log\pi_{\theta'}] \tag{2}$$

Importance sampling:

The data are from P samples, which are used for t training θθ2e. In practice, they predict the results as much as possible through the network. We can directly take the previous parameter iteration of the training model:

$$clip\left(\frac{p_\theta(a_t|s_t)}{p_{\theta^k}(a_t|s_t)},1-\varepsilon,1+\varepsilon\right)(RestrictionsinPPO2version) \tag{3}$$

$$\nabla_{\theta'}J(\theta') = E_{\tau\sim\pi_\theta(\tau)}\left[\frac{\pi_{\theta'}(\tau)}{\pi_\theta(\tau)}\nabla_{\theta'}log\pi_{\theta'}(\tau)r(\tau)\right] \tag{4}$$

DOI: 10.54254/2754-1169/3/2022849

LSTM is used to predict the future trend of bitcoin and gold, and then these data are used as input data to PPO algorithm for reinforcement learning

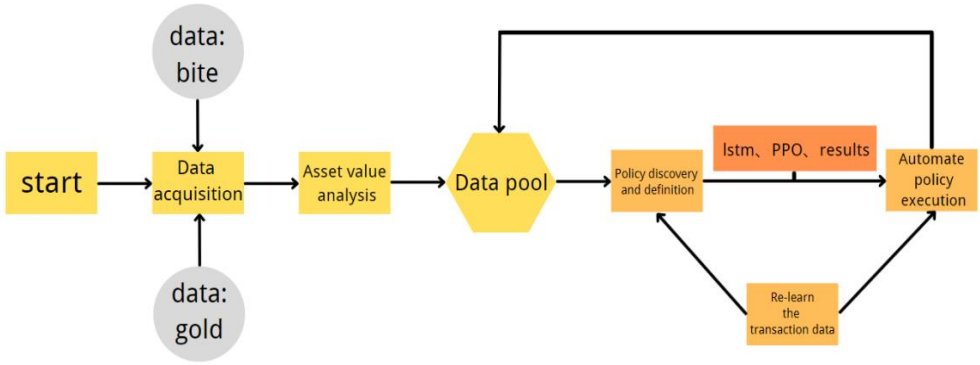

Figure 5 : Overall process flow chart.

2.5 Model Implementation

Environment parameter

MAX ACCOUNT BALANCE=2147483647 MAX NUM SHARES =2147483647 MAX SHARE PRICE = 5000 MAX VOLUME = 1000e8 MAX AMOUNT = 3e10

MAX OPEN POSITIONS=5 MAX STEPS = 20000 MAX DAY CHANGE =1

INITIAL ACCOUNT BALANCE=10000

We put bitcoin and bitcoin time series data into LSTM algorithm and the results are as follows:

Figure 6 : LSTM-Bitcoin.

We put the time series data of gold and gold into LSTM algorithm and the results are as follows:

DOI: 10.54254/2754-1169/3/2022849

Figure 7 : LSTM-GOL.

As the expectation of gold and bitcoin, it is put into PPO.

The design of reward function is very important to reinforce the learning goal. In the stock trading environment, we are most concerned about the current profit, so we use the current profit as the reward function.

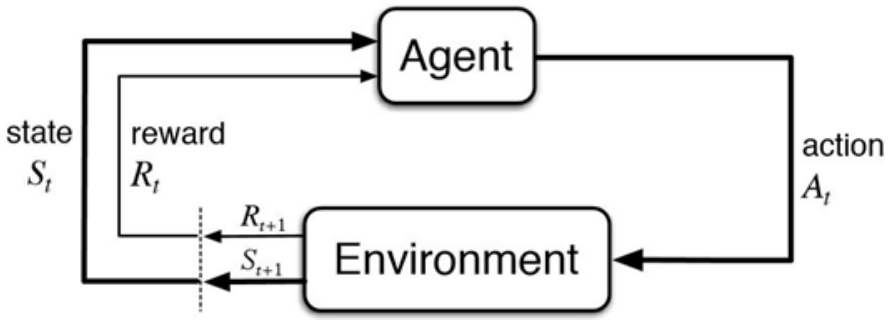

Figure 8 : Deep RL.

profit

award=net_worth−initial_account_balance, award =1

if award > 0 otherwise award = 100

In order to make the network learn the profit policy faster, when the profit is negative, the network is given a larger penalty (-100).

At the same time, the penalty coefficient λ is introduced (when the penalty for the gold incident * the penalty for the overall international situation equals 1, it is determined to be an upward trend). At this time, if the gold price is stable, buy all bitcoins.

In our deep reinforcement learning decision, bitcoin and gold revenue performance is as follows: The result of our model is: $ 157852.6585.

3. Model Advantages and Conclusions

This paper evaluates the performance of the model by calculating the accuracy and loss rate of the data sets of the price changes of bitcoin and gold.

DOI: 10.54254/2754-1169/3/2022849

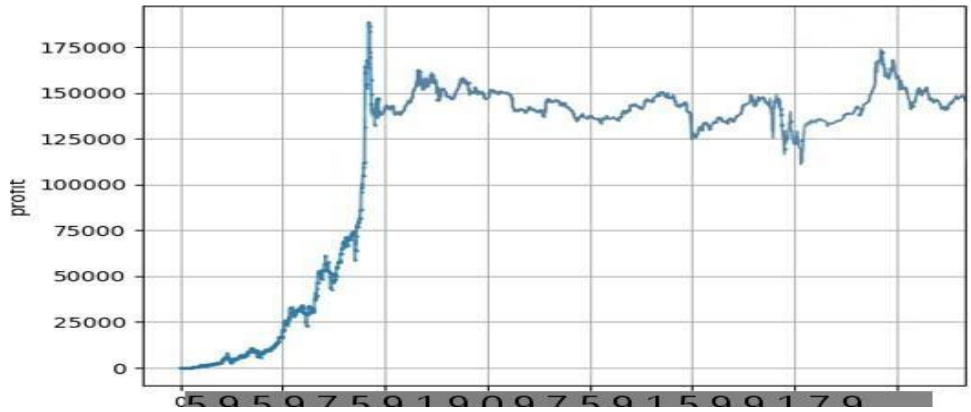

Figure 9: Bitcoin Performance.

Experiments will be based on PPO+LSTM, ordinary neural network, respectively forecast data sets. As can be seen from the table, the PPO+LSTM model is superior to the ordinary neural network model. PPO+LSTM has better prediction performance (Acc=0.95) and higher accuracy.

Table 2. Comparison with Traditional Quantified Investments (General Neural Network Version)

Model	Acc%	Loss%
PPO+LSTM	0.9562	0.2939
General neural network	0.8524	0.6548

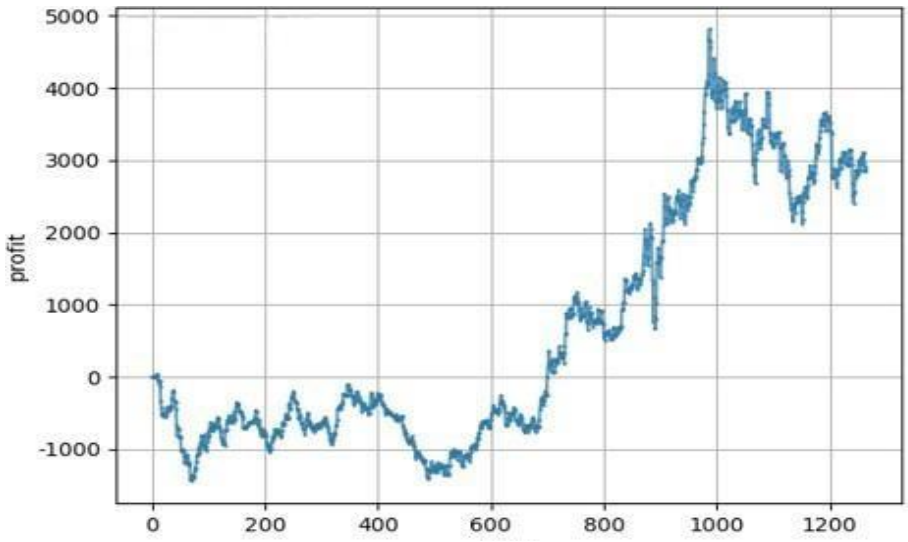

Figure 10 : Gold Performance.

The correlation between gold and bitcoin is that when one party is in an extreme market, the risk contagion to the other party's market is strengthened. When it is in the down market, gold is easily influenced by the bitcoin market, while when it is in the up market, the bitcoin market is more vulnerable to the impact of the gold market. That is to say, at a specific point in time, there is an

DOI: 10.54254/2754-1169/3/2022849

exchange relationship (either positive correlation or weak positive correlation) between these correlations. Considering the high revenue of bitcoin, we should purchase bitcoin at an appropriate time to get the maximum revenue. Then we need a penalty factor λ to judge this point in time. Bitcoin and gold are regarded as two stocks with certain dynamic correlation. By transforming the portfolio investment problem of two stocks into the purchase investment problem of a single stock, the generalization ability of the model is improved from the perspective of data. The existing quantitative investment strategy does not require much transformation. The penalty factor is a penalty for the gold incident and a penalty for the overall international situation. The penalty for the gold incident is an incident that affects the fluctuation of gold. The positive incident is set to 1, the negative incident is set to 0.7 as a penalty for the overall international situation, the positive incident is set to 1, and the negative incident is set to 0.5. When the penalty for the gold incident * the penalty for the overall international situation equals 1, it is determined to be an upward trend. At this point, if the price of gold is stable, all bitcoins will be bought. Therefore, with sufficient data support, the effect of deep reinforcement learning is better than that of traditional methods. We suggest that deep reinforcement learning should be introduced into the decision-making of quantitative investment.

Comparing the results of the traditional quantitative strategy with that of the artificial intelligence quantitative strategy, the strategic effects of the two strategies are compared under the same simple investment. According to the research results, the basic environment for the use of different strategic ideas and the risks to be avoided in the use process are explored, so as to reduce the trial and error rate of investors, help beginners who want to use quantitative ideas to correctly use different types of strategies, and reduce the cost of capital use and learning.

References

[1] *White H. Economic prediction using neural networks: the case of IBM daily stock returns[C]// IEEE International Conference on Neural Networks. 1988:451-458 vol.2 (1988).*

[2] *Zhou Xu. Application of Bolin Belt Trend Breakthrough Strategy in digital currency Market [D]. Zhejiang University of Technology and Industry, (2021).*

[3] *Su Chunlin. The application research of multi-factor model in digital currency market (master's degree thesis, University of Electronic Science and Technology). HTTPS://kns.cnki.net/kcms/detail/detail.aspx?dbname=CMFD202001&filename=1019853878.nh (2019)*

[4] *Zhang Jing. The application of multiple models to quantify investment. Dynamic analysis of stock market (32),87 (2013).*

[5] *Meng Ye, Yu Zhongqing & Zhou Qiang. A safe quantitative stock selection strategy-empirical evidence from A-share market. Financial theory and practice (08),102-107 (2018).*

[6] *Huang Zepeng. modeling and application of gold price forecast based on in-depth learning (master's degree thesis, Shanghai jiaotong university). https://kns.cnki.net/kcms/detail/detail.aspx?dbname=CMFD201902&filename=1019880198.nh 2017()*

[7] *Ye Wuyi, Sun Liping, Miao Baiqi. Dynamic cointegration study of gold and bitcoin-based on semiparametric MIDAS quantile regression model [J]. Systems Science and Mathematics, 40(07):1270-1285 (2020).*