2023 Asia-Europe Conference on Electronics, Data Processing and Informatics (ACEDPI) | 979-8-3503-0115-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ACEDPI58926.2023.00040

Trading strategy prediction model based on quadratic programming and XGBoost

Shuaikai Ding*
Henan Polytechnic University
Jiaozuo, China
3417617790@qq.com

Shaobo Ding
Harbin University of Science and Technology
Harbin, China

Tianshuo Ding
Henan University of Science and Technology
Luoyang, China

***Abstract*—In this paper, the quadratic programming is established on the basis of mean-variance evaluation method, and the commission is taken into account to achieve the effect of large investment return and small risk. The error analysis between the predicted data and the actual data is made. Then, the obtained data are introduced into the quadratic programming model to constrain the constraint conditions, and the percentage of daily investment in gold and bitcoin in 5 years is solved, which is the trading strategy. XGBoost regression model and BP neural network model are used for prediction. The error analysis of the two strategies is carried out, and the XGBoost model with high prediction accuracy is selected. The prediction accuracy can prove that this strategy is the best strategy.**

Keywords—XGBoost regression model, BP neural network model, Quadratic programming model

I. INTRODUCTION

In the past four decades, the financial markets at home and abroad have been developing and innovating. Especially with the rapid development of computer technology, the financial environment has undergone great changes. At the same time, the risks of financial markets have been increasing[1]. Many scholars have conducted research on this, and now the methods and theories of stock market risk measurement have been more mature. However, due to its own hotness and challenge, it is still the object of research for many scholars.

Since the Second World War, the investment management industry in developed Western countries has enjoyed a rapid growth in development. Many prominent investors have studied the trading system approach, and many other scholars have quantified trading strategies and risks.

In 1952, Harry Markowitz, a famous American financier,in his academic paper "Asset Selection: Effective Diversification", first applied the mathematical concepts of mean and variance of asset portfolio payoffs to quantify benefits and risks, providing a fundamental approach to risk measurement.For the first time, the principles of marginal analysis were applied to the analytical study of asset portfolios. In the same year, Markowitz published the article "Asset Portfolio Selection - Effective Diversification of Investments" in the Journal of Finance. This article was the first to use the mean of risky assets and risk represented by variance to study the problem of asset portfolio and selection, which laid the theoretical foundation for modern finance and was a milestone in the history of modern financial theory

Later, Sharp simplified Markowitz's model and proposed the capital asset pricing model and beta theory, which were both based on Markowitz's model and solved the drawback that Markowitz's proposed model did not utilize the measurement of risk. However, there are still computational complexities and difficulties in effectively distinguishing between systematic and unsystematic risks.

In 1993,the VaR risk measurement method was proposed by large financial institutions represented by J.R. Morgon[2], G30 Group.The method is simple in meaning and intuitive in value judgment,and measures investment risk by calculating the maximum possible loss of a risky asset at a given confidence level over a holding time interval, taking into account the investment environment. Its calculation methods mainly include: normal distribution method, historical simulation method, stress test method, and Monte Carlo method.

II. XGBOOST REGRESSION MODEL

A. XGBoost regression model building

First we make some assumptions about the problem:

Assuming there are no other sources of investment or lending.

Assume that all the available cash is used to invest in gold and bitcoin.

It is assumed that the only additional cost of the trading process is the trading commission.

Assuming all investments are converted tocash at the close of trading on 9/ 10/2021

We should make a reasonable investment based on the data as of the day of the investment, so first we need the data we have looked up and the data in the attachment to make a prediction of the price of gold and bitcoin on the day of the investment and then determine the investment plan. The data collected is shown as Figure 1 and Figure 2.

979-8-3503-0115-1/23/$31.00 ©2023 IEEE
DOI 10.1109/ACEDPI58926.2023.00040

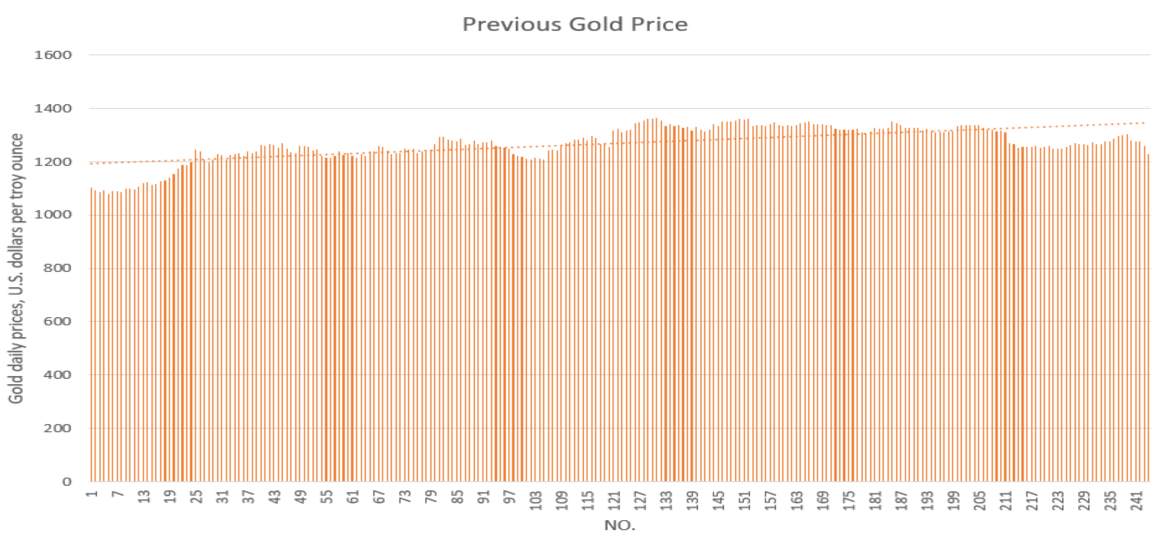

Figure 1 Price trend of gold in the 240 days before 9/ 11/2016\

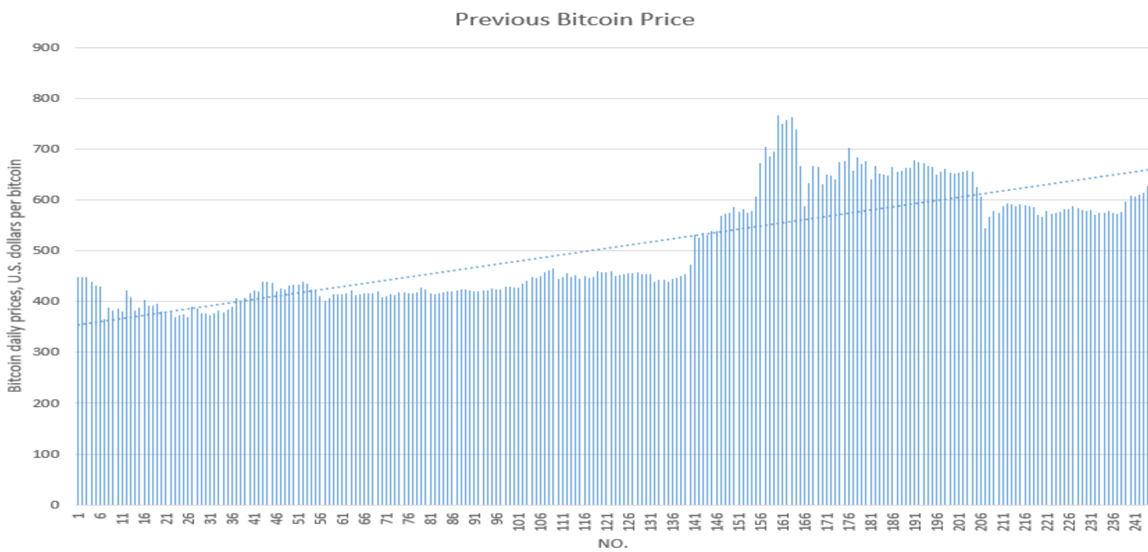

Figure 2 Price trend of bitcoin for 240 days before 9/ 112016

We should make a reasonable investment based on the data as of the day of the investment, so first we need the data we have looked up and the data in the attachment to make a prediction of the price of gold and bitcoin on the day of the investment and then determine the investment plan.[3]The XGBoost algorithm is an extension of the gradient boosting machine (GBM) algorithm, which is an optimization model with the characteristics of both linear and tree models, capable of performing both regression and classification tasks. [4] The XGBoost algorithm consists of multiple decision trees (CART), and machine learning is achieved through decision tree integration, where the predicted values of all the decision trees are summed to the model predictions and trained by the gradient boosting Decision Tree (GBDT) algorithm for simulation training.Unlike the GBDT algorithm, the XGBoost algorithm is fast in training and high in prediction accuracy by performing a second-order Taylor expansion on the loss function to quickly approximate the objective function and adding a regular term adjustment parameter to avoid overfitting. Therefore, in terms of prediction, the XGBoost regression model has more obvious advantages due to its characteristics of stronger generalization ability, higher expansion ability, and faster computing speed.

Using q to denote the number of samples mapped to leaf nodes, T to denote the number of leaf nodes, and w to denote the real fraction of leaf nodes, the set of decision tree structures of CART is obtained as follows

$$F = \{f(x) = \omega_{q(x)}\}(q : R^3 \to \{1,2,\cdots,T\}, \omega \in R^T) \quad (1)$$

Using the Rit dataset and cash, gold, and bitcoin weights representations x_i to obtain

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F(i = 1,2,\cdots,1826) \tag{2}$$

Using L to denote the error function term and Ω to denote the complexity function term of the model, the objective function of the XGBoost regression model is obtained as follows.

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3}$$

$$\Omega = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{4}$$

$$Obj = L + \Omega \tag{5}$$

Then, we can substitute the training data into the model, and in this process, we have to optimize the training so that the objective function is as small as possible. The new objective function is obtained by $\hat{y}_i^{(t)}$. $f_t(x_i)$ denoting the predicted value of the model at time t and the new function added at time .

$$Obj^{(t)} = \sum_{i=1}^{n} (y_i - (\hat{y}_i^{(t-1)} + f_t(x_i)))^2 + \Omega \tag{6}$$

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{7}$$

B. Solution of XGBoost regression model

The second-order Taylor expansion of the obtained objective function yields the following objective function

$$Obj^{(t)} \approx \sum_{i=1}^{n} [(y_i - \hat{y}^{(t-1)})^2 + 2(y_i - \hat{y}_i^{(t-1)}) f_t(x_i) - h_i f_t^2(x_i)] + \Omega \tag{8}$$

Then remove the constant term to obtain the relationship between the objective function and the first and second order derivatives of the error function.

$$Obj^{(t)} \approx \sum_{i=1}^{n} [g_i \omega_{q(x_i)} + \frac{1}{2} h_i \omega_{q(x_i)}^2] + \gamma T + \frac{1}{2} \sum_{j=1}^{T} \omega_j^2 = \sum_{j=1}^{T} [(\sum_{i \in I_j} g_i)\omega_j + \frac{1}{2}(\sum_{i \in I_j} h_i + \lambda)\omega_j^2] + \gamma T \tag{9}$$

Finally, the quadratic function is used to solve for the minimum value to obtain the optimal solution [5].

$$x_j = \frac{-\sum_{i \in I_j} g_j}{\sum_{i \in I_j} h_i + \lambda} \tag{10}$$

$$Obj = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \lambda T \tag{11}$$

The graph of the predicted data brought into Matlab software and run is as Figure 3 and Figure 4.

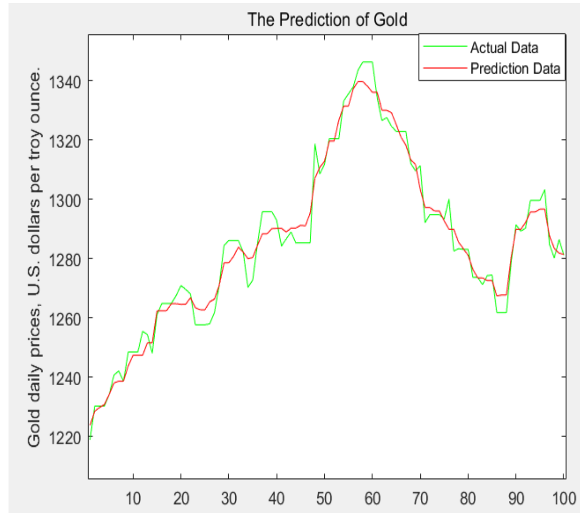

Figure 3 Forecast of gold returns for XGBoost regression model(7/ 13/2017-10/20/2017)

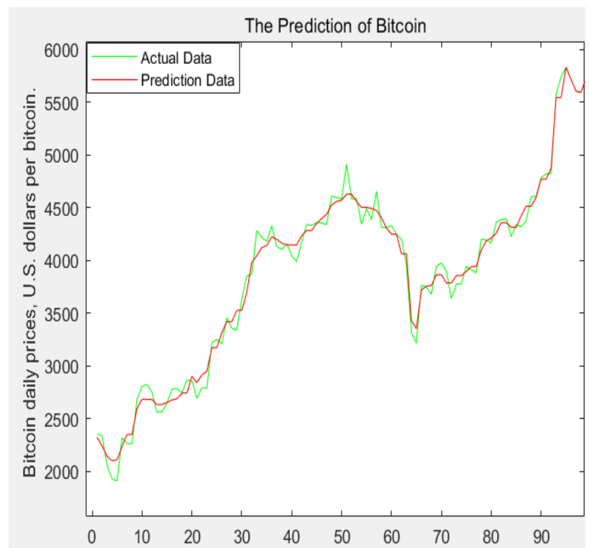

Figure 4 XGBoost regression model 7/ 13/2017- 10/20/2017 BP's bitcoin return prediction graph

III. BP NEURAL NETWORK MODEL

Neural networks are a newly emerging research method in artificial intelligence with massively parallel, distributed storage and processing, self-organization, self-adaptation and self-learning capabilities, especially for dealing with imprecise and ambiguous information processing problems that require many factors and conditions to be considered simultaneously· The present problem is nonlinear and has a large amount of data with its extreme fitness, so we try to use BP neural networks for its solution.

A. BP neural network modeling

The closing price of troy ounces of gold in US dollars on the specified date and the dollar price of a single bitcoin.

This is very tedious to handle directly, so we process the data for the closing price of troy ounces of gold in US dollars on the specified date and the dollar price of a single bitcoin into the yield of gold on that date and the daily yield of bitcoin [6]. The processing is expressed as follows.

$$R_{2j} = \frac{G_j}{G_{j-1}} \tag{12}$$

$$R_{3j} = \frac{B_j}{B_{j-1}} \tag{13}$$

Where equation, at the $j = 5$, 10, 15, …, 1825 time, uses the following equation

$$R_{3j} = \frac{B_j}{B_{j-3}} \tag{14}$$

The model takes as input the number of days to start and as output the return of gold on that date and the return of bitcoin on each day [7].

B. Solution of BP neural network model

We use Matlab to create, train and simulate the BP neural network, and then inverse normalize the data to obtain the real data, the trading strategy and the final return.

The predicted returns for day one gold and bitcoin were obtained using the collected data forecasts as Figure 5 and Figure 6.

After that, each day we add its previous data to the forecast again to get the next predicted returns of gold and bitcoin [8]. Below is a partial image of the forecast results for gold and bitcoin for the 100-day period from 7/ 13/2017 to 10/20/2017, as follows in Figure 7 and Figure 8.

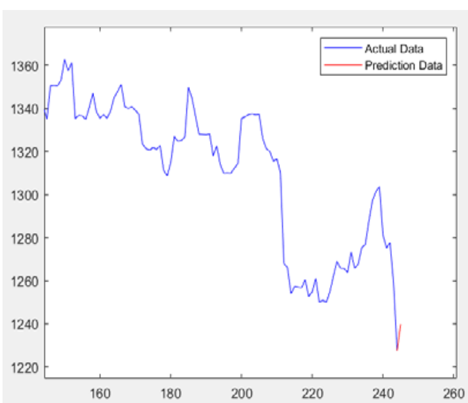

Figure 5 Yield forecast chart for gold on 9/ 11/2016

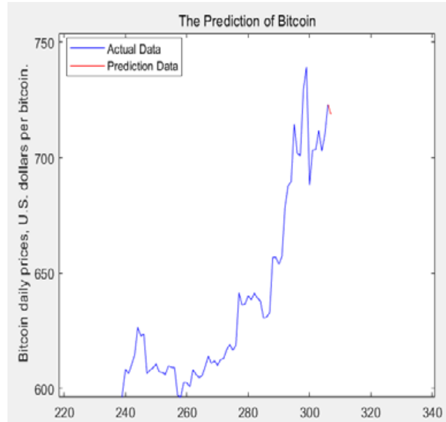

Figure 6 Yield Forecast for Bitcoin on 9/ 11/2016

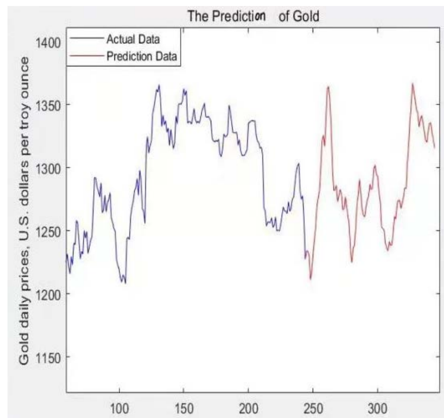

Figure 7 BP Neural Network Model Gold Return Forecast for 7/ 13/2017-10/20/2017

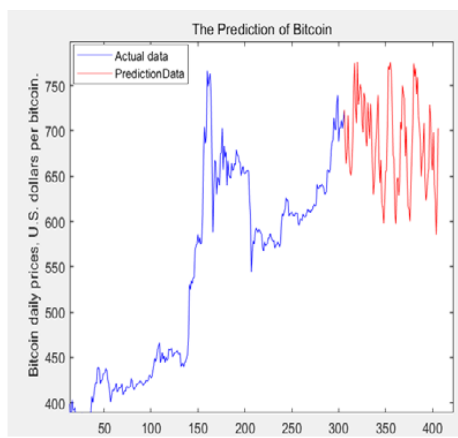

Figure 8 BP neural network model 2017.7. 13-2017. 10.20 BP's bitcoin return prediction graph

IV. Comparison of Predictive Models

From the BP neural network prediction model we can get the error between its predicted obtained gold and bitcoin returns and the actual data, i.e. Error analysis of BP neural network model prediction data and real data is shown as Figure 9.

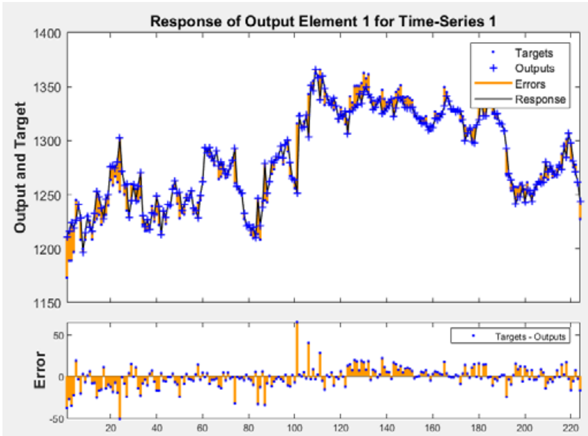

Figure 9 Error analysis of BP neural network model prediction data and real data

From the XGBoost regression prediction model we can get the error between its predicted obtained gold and bitcoin returns and the actual data, i.e.

Error analysis of XGBoost model gold prediction data and real data is shown as table I. Error analysis of XGBoost model bitcoin prediction data and real data is shown as table II.

TABLE I ERROR ANALYSIS OF XGBOOST MODEL GOLD PREDICTION DATA AND REAL DATA

Gold	MSE	RMSE	MAE	MAPE	R^2
Training sets	33.036	5.748	4.591	0.375	0.98
Test sets	20.099	4.483	3.562	0.277	0.975

TABLE II ERROR ANALYSIS OF XGBOOST MODEL BITCOIN PREDICTION DATA AND REAL DATA

Bitcoin	MSE	RMSE	MAE	MAPE	R^2
Training sets	2786.5	52.787	36.076	2.519	0.993
Test sets	9162.7	95.722	77.7.7	2.219	0.987

Observing the error analysis graph of the predicted data and the real data of the BP neural network model, we can see that the error obtained by the BP neural network prediction is large and its accuracy cannot be guaranteed, which will make a large error in the investment ratio. In the error table of gold and bitcoin data and real data obtained from the XGBoost regression model, we can see that the mean square error - MSE, root mean square error – RMSE [9], mean absolute error - MAE, and mean absolute percentage error - MAPE all take small values, and R^2 are all very close to 1, indicating that the data obtained from the prediction is very close to the real value. Therefore, we used the data obtained from the XGBoost regression model for the determination of daily scenarios and the final value calculation [10].

V. PROOF OF OPTIMAL STRATEGY

Before rationalizing the investment plan in part2, we first make a forecast of the data. In the prediction process, because of its large amount of data and the demanding nature of the accuracy, we chose the BP neural network prediction model and XGBoost regression prediction model for the prediction of the data, and then we compared it with the true value to get to conduct an error analysis, and used the one with a smaller error to determine the transaction strategy and the final value, this improves the accuracy of the final results. The quadratic programming model allows us to measure the relationship between the variables gold price and bitcoin price and returns for different scale changes during the calculation and is a good choice for solving our problem.

VI. SENSITIVITY ANALYSIS

We should to assess the sensitivity of the strategy to transaction costs, i.e., to perform a sensitivity analysis of the quadratic programming model. We use the final benefit to respond to the strategy and the commission rate in the transaction to respond to the transaction cost, and discuss the impact of changes in the commission rate on the final benefit to screen the sensitivity of the strategy to the transaction.

We borrowed the process of generating random numbers from the Sobol method and set the range of variation of the trading rate to about 0.35%, so replace αG and αB in the objective function with the range of $\alpha \pm 0.35\%$

The error rate of the trading strategy is shown in the chart below. It can be seen that an error rate of less than 1% is a better ideal when the trading commission changes. Trading strategy error rate is shown as Figure 10.

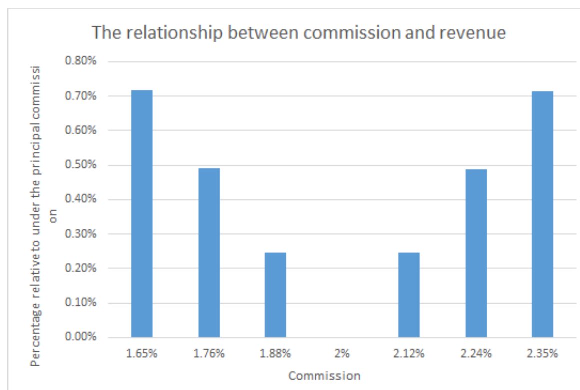

Figure 10 Trading strategy error rate

Regarding how transaction costs affect trading strategies and outcomes, we vary transaction costs and by comparing trading strategies we can know: When transaction costs become large, the number of transactions is reduced to minimize costsWhen the cost of trading becomes small, the number of trades will increase in pursuit of higher returns

VII. ADVANTAGES

Pre- processing the data before solving simplifies the calculation and converting all variables to the same units avoids confusion in the calculation process.

When predicting the data, BP neural network model and XGBoost regression model are used, and then the real values are compared, and the data obtained from the XGBoost regression model with high accuracy is selected, making the results accurate and improved.

The sensitivity of the strategy to transaction costs is clearly expressed through the data graph.

In the quadratic programming model, risk and transaction costs are innovatively combined to ensure a large rate of return with minimal risk.

REFERENCES

[1] Jirapongpan R , Phumchusri N . Prediction of the Profitability of Pairs Trading Strategy Using Machine Learning[J]. IEEE, 2020.

[2] https://www.cngold.org/quote/.2/22/2022,21:09:01

[3] https://www.528btc.com/zhuanti/533577.html.2/22/2022,21:09:01

[4] Chen JQ, Yang JJ, Lou ZB. Research on a new short-term load forecasting model based on XGBoost algorithm[J]. Electrical Measurement and Instrumentation,2019,56(21):23-29.

[5] Han T, Yin WS, Fang M. Sentiment analysis model based on convolutional neural network and XGBoost [J]. Journal of Fudan (Natural Science Edition,2019,58(5):560-564.

[6] Wen X, Chen Dy, Qiao Yu. Domestic neural network fault techniques and their applications in spacecraft[J]. Journal of Shenyang University of Aeronautics and Astronautics,2018,35(3):17-25.

[7] Liu Hao-Ran, Zhao Cui-Xiang, Li Xuan et al. A research on a neural network optimization algorithm based on improved genetic algorithm[J]. Journal of Instrumentation,2016,37(7):1573- 1580.

[8] Fu Clan, Li Hui, Zhang Xia. BP neural network prediction model and application[J]. Information and Computer(Theory Edition), 2016,(11):47-48.

[9] Shi Q . Prediction of O2O Coupon Usage Based on XGBoost Model[C]// ICEME '20: 2020 The 11th International Conference on E-business, Management and Economics. 2020.

[10] Chong GuoDeben CaoYiran QiaoZhenyao YangZhenzhong ChangDi ZhaoZhuoran Hou. Energy Management Strategy of Extended-Range Electric Bus Based on Model Predictive Control[J]. SAE International Journal of Commercial Vehicles, 2021, 14(2).