Construction and application of gold price prediction model based on historical data

Xuexia Xing*, *Qingqing* Yan, and *Yutong* Wang

Shandong Xiehe University, Jinan, Shandong, China, 250100

Abstract. This article delves into the methods and practices of building gold price prediction models based on historical data. Predictive models were constructed using machine learning methods such as random forests and BP neural networks by collecting and preprocessing data on gold prices and related economic indicators from September 2022 to December 2023. Trained and validated, BP neural network models excel in prediction accuracy and stability due to their strong nonlinear fitting capabilities. It can provide investors with accurate decision-making basis and help to preserve and increase the value of assets.

Keywords: Gold price prediction, Time series analysis, Machine learning; BP neural network.

1 Introduction

Gold, as a special asset with both monetary and commodity attributes, occupies an important position in the financial market. Its price fluctuations are interactively influenced by multiple factors such as the US dollar exchange rate, inflation rate and interest rate. In the current complex and volatile economic situation, accurately predicting the gold price is of extremely crucial significance for investors to optimize asset allocation, implement risk management, as well as for enterprises to formulate business strategies and for the government to improve policy regulation. This article focuses on how to utilize historical data to mine the changing patterns of gold prices, aiming to construct an efficient and accurate prediction model.

2 The current state of gold price forecasting research

Dooley Isard and Taylor(1995) [1]The impact of exchange rate changes on gold prices is studied, and an empirical analysis of data from 1976 to 1990 shows that the impact of exchange rate changes on gold prices is significant; Christie-David et al.（2000）[2]. analyzed the monthly data of gold futures prices from 1992 to 1995, and the experimental results showed that GDP had a significant impact on gold prices; Capie, Mills & Wood(2004) [3] The relationship between gold prices and major currency exchange rates is studied, and the

* Corresponding author: 653392026@qq.com

results show that the two move in the opposite direction.;Yang Liuyong and Shi Zhentao(2004) [4] Empirical studies show that the Dow Jones price index, the U.S. inflation rate, the U.S. dollar nominal effective exchange rate, and the U.S. federal funds rate have a significant impact on gold prices in the long run. These research status clearly verify that these indicators do affect the fluctuation of gold prices, which provides extremely important evidence for the selection of research indicators in this paper. Antonino Parisi(2008) [5] This paper mainly investigates the prediction ability of the neural network model on the gold price, and the comparative results show that the feedback neural network model achieves a good prediction effect.

3 Data processing

3.1 Data sources and variable definitions

The independent variables cover GDP, inflation rate, daily open, close, high, low prices that affect the price of gold. The dependent variable is the gold price, and the data is taken from the authoritative gold investment network, which has been strictly screened and verified to ensure the authenticity and reliability of the data.

3.2 Missing values are filled

Using the function of replacing missing values in SPSS26.0 software, the sequence average method is adopted, that is, the sum of all elements in the sequence is divided by the number of elements, so as to avoid a large loss of effective data and the degradation of model performance, and the filled data can more truly reflect the actual situation.

3.3 Data standardization

To eliminate dimensional differences between different data items, Z-score normalization was used. SPSS26.0 is used to calculate the mean and standard deviation, then calculate the difference between each data point and the mean value, and finally divide by the standard deviation to make the mean value of the data 0 and the standard deviation to 1, which follows the normal distribution law, improves the convergence speed and accuracy of the algorithm, and enhances the stability of the model.

4 Model building

4.1 BP neural network predicts the price of gold

Implementation principle: The BP neural network is trained based on the error backpropagation algorithm, which is composed of an input layer, a hidden layer and an output layer. The learning process includes signal forward propagation and error backpropagation, and the complex mapping relationship between input and output is learned by iteratively adjusting the weight.

Implementation steps: A total of 438 Chow Tai Fook gold price and related economic indicator data were collected from the second half of 2022 to 2023, and divided into training sets and test sets after preprocessing. The BP neural network structure was designed, the number of neurons in each layer was determined, and the appropriate activation function and loss function were selected. The training dataset was used to train the model, the weights were obtained, and the mean square error (MSE), coefficient of determination (R^2), mean absolute error (MAE) and mean relative error (MRE) of the prediction results were calculated to evaluate the model performance.

Multiple iterations are required during training until the prediction performance of the network reaches a preset requirement or the maximum number of iterations is reached. In fact, there were 12 iterations, and the best validation performance on the training set was 8.1588e-07 for 6 rounds. (as shown in Figure 1)

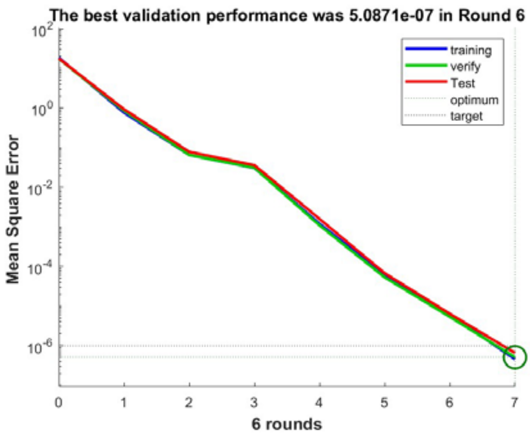

Fig. 1. Diagram of the sixth round of optimal performance verification.

4.2 Random forest predicts the price of gold

The implementation principle of random forest prediction of gold price mainly includes Bagging and feature random selection. Bagging uses the self-service aggregation method to build a large number of decision trees, and the training set of each tree is randomly sampled, allowing for duplication, which increases the diversity of the model, helps to reduce variance and prevent overfitting. At the same time, when the random forest is constructed, a part of the features will be randomly selected for splitting, which further introduces randomness, reduces the variance of the model, and improves the generalization ability. Random forests also provide a mechanism for assessing the importance of variables, and by calculating the importance score of each feature across all decision trees, it is possible to understand which features have a greater impact on the prediction outcome.

The steps are: First of all, load the data, here we use the randomly generated data as an example, extract the sample number m=100 and the number of features n=10, randomly generate the feature values and target variables to refer to the Chow Tai Fook gold price data table, load out 444 data, and then divide the dataset into a training set and a test set, keep 30% of the data as the test set, and then use the fitctree function to train the random forest model, and then use the trained model to predict, The model performance was evaluated, the prediction accuracy and error rate were calculated, and finally a tree in the random forest was visualized, as shown in Figure 2.

5 Conclusion

From the above figure, the above theoretical basis and the real error data, it can be seen that the coefficient of determination of the BP neural network R^2=0.999 is more close to 1, so the fitting effect of the BP neural network model is better, the accuracy is higher and better.

Secondly, the mean square error of BP nerve is 1.5711e-06 and the mean absolute error is 0.0012278, which is much smaller than the error of random forest (3.4242), which is close to 0, so the prediction error of the model is smaller and the accuracy is higher.

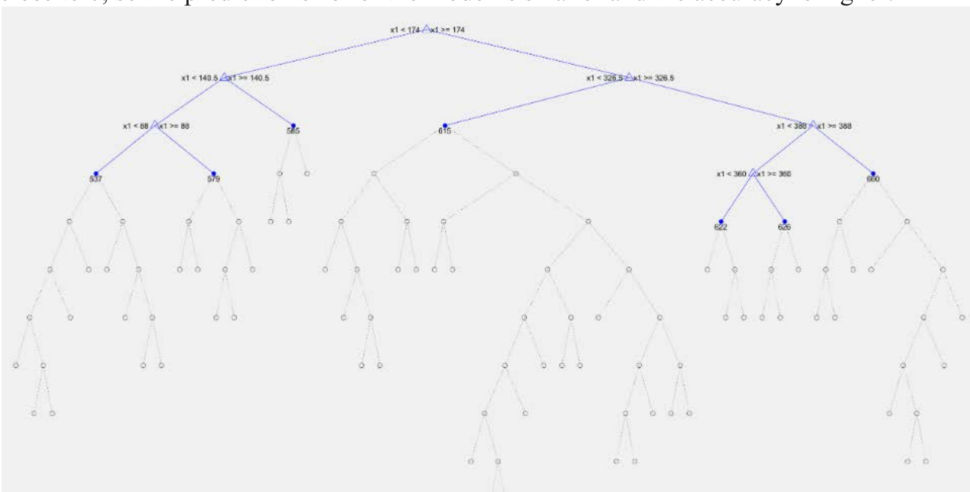

Fig. 2. Visualize a random tree plot.

Finally, after changing and adjusting the input layer dataset and setting the training parameters, the average relative error of the BP neural network has reached the minimum value of all iterations, and the optimal fitting effect is completed.

Considering the goodness of fit of the model, prediction error and other factors, the BP neural network was selected as the optimal prediction model. Strong learning and adaptability in predicting gold prices, the ability to deal with complex problems, good generalization ability, and the ability to improve forecast accuracy and efficiency through optimization methods. This makes BP neural networks an effective tool for predicting the price of gold. Its learning rules can automatically adjust internal parameters to adapt to complex nonlinear problems. Its generalization capabilities enable it to respond to unpredictable factors in the gold market and provide investors with accurate forecasts.

In the future, we can strengthen interdisciplinary cooperation and exchanges, and combine time series analysis and machine learning algorithms with knowledge in economics, finance, psychology and other fields to jointly promote the development and application of gold price prediction technology. This will help us better understand the nature and laws of gold prices, and provide investors with more accurate and reliable predictions.

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Mengxia Wang, whose meticulous guidance and selfless support have enabled me to successfully complete this research. Her rigorous academic attitude and profound scholarly achievements have had a profound impact on my academic growth.

I would also like to thank all the members of the Artificial Intelligence Laboratory, who have provided valuable suggestions and assistance during the research process. Their support in data analysis, paper writing, and other aspects has been invaluable to me.

Furthermore, I am grateful for the financial and equipment support provided by the school for this research. These resources are crucial for the smooth progress of experiments and the collection and analysis of data.

I would also like to express my gratitude to my family and friends, who have offered endless understanding, support, and encouragement throughout my academic journey. It is their companionship and dedication that have allowed me to immerse myself fully in academic research.

References

1. Dooley Isard,Taylor.Exchange rates,country specific shocks and gold[J].Applied financial economics,1995(5): 121-129.
2. Christie-David,R. Chaudhry,M. and Koch, T.W. Do macroeconomic news releases affect gold and silver prices [J]. Journal of Economics and Business, 2000(52), 405-421.
3. Forrest Capie, Terence C Mills, and Geoffrey Wood. Goid as a hedge against the US dollar[R]. The World Gold Council, 2004.
4. YANG Liuyong,SHI Zhentao. Analysis of the long-term determinants of the price of gold[J],Statistical research,2004(6):21-24
5. Antonino Parisi, Franco Parisi, David Diaz. Forecasting gold price changes: recursive neural network models [J]. Journal of Multinational management,2008,18(5):477-487