Information Systems and Economics (2024)
Clausius Scientific Press, Canada
DOI: 10.23977/infse.2024.050414
ISSN 2523-6407 Vol. 5 Num. 4

Research on the application of gold price prediction based on LSTM model

Yuquan Guo*

School of Software, Shanxi Agricultural University, Jinzhong, 030801, China
**Corresponding author: guoyuquan2013@126.com*

Keywords: Gold Price Prediction, LSTM Model, Forgetting Mechanisms, Time Series

Abstract: A gold price prediction model with LSTM (Long Short Term Memory Network) is proposed. The data from 2013 to 2023 are evaluated. The results show that the model has an excellent prediction effect with an accuracy of 96.9%. This research provides new effective methods and ideas in the field of gold price prediction, provides highly valuable and important reference bases for the majority of gold industry practitioners in investment decision-making, risk control, etc., and is of great significance to the stable and efficient development of the gold trading market. This paper uses an LSTM network to predict the future price of gold, and verifies the performance and accuracy of the model through a series of data processing, model training and evaluation steps. The application research of gold price prediction based on LSTM model can provide a reliable price reference for the gold trading market, promote the stable operation of the market, reduce transaction costs, and improve the trading efficiency of the market.

1. Introduction

As an important precious metal, gold has a unique position in the global financial market. Gold can be used as a craft, as a collectible, or as an investment, and is often used in financial markets to preserve and appreciate value. Despite losing its position as the world's currency, gold still has the dual attributes of a commodity and a currency[1], and its value cannot be ignored. With the deepening of the impact of economic globalization, people's ownership and investment in gold are deepening, and there is an urgent need for accurate prediction of gold prices. Over the course of the year, AI has continued to evolve to provide us with many ways to predict the price of gold. At the same time, the international situation is characterized by frequent conflicts and economic turmoil, making it difficult to predict the trend of gold prices.

From statistical models to deep learning, from offline learning to online learning [2]: Juddhar et al. used traditional statistical models to combine the hidden Markov model and the sparrow search algorithm to predict the price of gold futures[3], and Yang Chen et al. established the EMD-LSTM model to improve the accuracy of price prediction[4]. Liang et al. used CEEMDAN-LSTM model for short-term prediction [5], He Linyun et al. used ICEEMDAN-SE-SSA-ELM algorithm to improve the prediction accuracy [6]. LSTM significantly improved the goodness-of-fit of nonlinear data [7].

2. Establishment of the model

2.1 An introduction to the basic concepts of the model

LSTM is a recurrent neural network with a forgetting mechanism that excels at working with time series data, often seen in tasks such as gold price prediction. It plays an important role in the field of application of gold price prediction. Gold prices form data series with sequential and time-dependent correlations over time, and LSTMs have the ability to capture medium- and long-term and short-term dependency patterns.

The LSTM contains a memory unit, which can remember or forget past information in a targeted manner, and control the flow of information with input gates, forgotten gates, and output gates. When learning from the past data of gold prices, it strives to understand the patterns and laws of price fluctuations, and extracts valuable characteristics on its own, such as the trend of prices and the magnitude of fluctuations, etc., so as to provide investors and traders with predictive information and provide auxiliary support for their decision-making.

2.2 Establishment of LSTM model

2.2.1 Introduction to the structure of the LSTM model

LSTM is a recurrent neural network (RNN) architecture with unique properties. It is no coincidence that it was born, but is specifically designed to deal with the gradient vanishing and gradient explosion problems that traditional RNNs frequently encounter when working with long series data.

ForgetGate [8]: It plays an important role in determining the information that should be discarded in the state of the cell.

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \tag{1}$$

InputGate: It has the key mission of clarifying what new information should be added to the cellular state.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{c} = tanh(W_c[h_{t-1}, x_t] + b_c) \tag{3}$$

$$c_t = f_c \circ c_{t-1} + i_t \circ \tilde{c} \tag{4}$$

OutputGate: Its core function is to closely control the output of the cell state.

$$o_t = \sigma(W_o[h_{h-1}, x_t] + b_o) \tag{5}$$

Among them, W_f, W_i, W_o and W_t are the weight matrix of forgetting gate, input gate, output gate, and input unit state, respectively. b_f, b_i, b_o and b_t are forgetfulness gates, input gates, output gates, and input cell state bias matrices, respectively [9]. The hidden layer structure of the LSTM is shown in Figure 1.

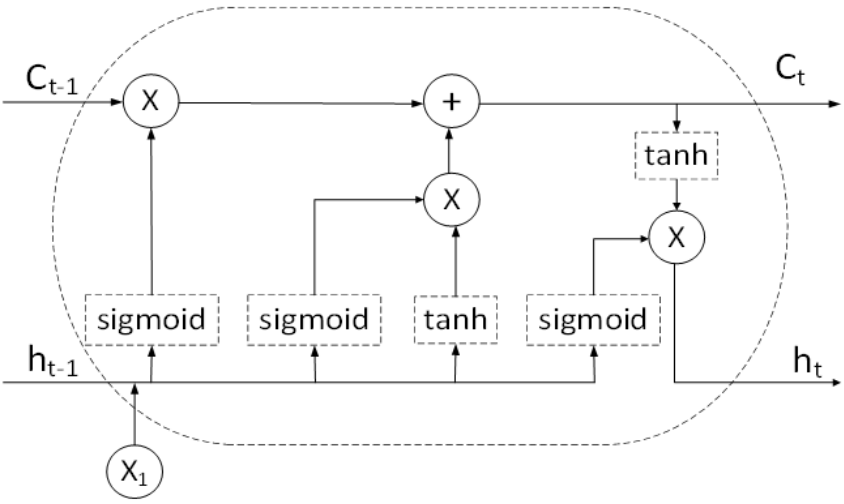

Figure 1: Hidden layer structure of an LSTM

2.2.2 LSTM-based gold price prediction method

The price of gold is markedly volatile and cyclical, while the data sets we collect clearly show a strong quality of following a time series distribution. In the learning process, the LSTM unit has the ability to remember long-term dependencies, which is critical for capturing cyclical and trending movements in gold prices. The model continuously adjusts its internal weights and parameters to minimize the error between the predicted and actual values, thereby improving its performance over time. After training, the gold price data can be imported into the trained LSTM model for prediction and profiling. The program flow diagram is shown in Figure 2.

Step1. Data preprocessing

After reading the file, delete the feature columns that you don't need. Sort in chronological order, resetting the index. Ensure that there are no duplicate samples and missing values, providing accurate and usable data.

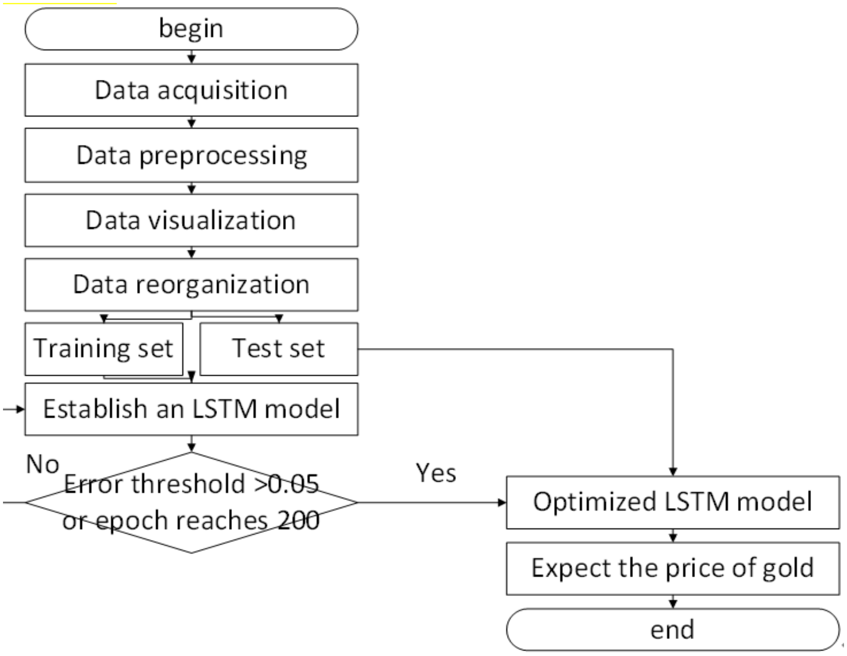

Figure 2: Procedure Flow Chart

Step2.Data visualization

In order to be more convenient and efficient to be able to better see the direction of the gold price. The line chart is a clear and useful form of chart that shows how the price of gold changes at each point in time. With its sleek lines, we are able to get a particularly direct and clear picture of the upward and downward movement of the gold price. In order to conduct a more comprehensive and in-depth exploration of the law of gold price changes, a training and test set of gold prices is also specially drawn up. In this diagram, it is illustrated by using polylines of different colors. The visualization is shown in Figure 3

Step3 Data reorganization

Future data in a time series cannot be directly trained and time series data cannot be randomly divided. In time series classification, the test set always follows the training set. We spend the last year on testing and all the rest for training. We used the last year of 2022 as the test set, and 2013-2021 as the training set.

Figure 3: Visualization diagram

Step4 Create LSTM model

Gold price prediction is essentially a comprehensive analysis and forecasting of complex and dynamically changing time series data. We selected the mean squared error to act as a loss function to evaluate the predicted effect. On this basis, the specific parameters of the input dimension, the detailed setting of the output dimension, the precise number of neurons in the fully connected layer, and the specific form of the excitation function of the fully connected layer are further clarified.

Next, the weights and biases are initialized, and the training process of the LSTM network begins. Closely monitor the change of the loss value of the validation set, and if the verification loss does not show significant improvement within 50 consecutive epochs, then the training is stopped. If the training is not stopped early due to the callback function before 200 epochs, then the entire model iteration process will end after 200 epochs.

Step5 Predict the price of gold

With the help of LSTM models to predict gold prices, we expect to be able to grasp the general direction of the price as well as cyclical movements. The model will give a forecast range for gold prices in 2022, as well as an assessment of the likelihood of rising or falling prices. The LSTM model is suitable for forecasting gold prices in a relatively stable and predictable market environment.

3. Solving the model

3.1 Preprocessing of data

This article selects 2,583 data samples from January 2, 2013 to December 30, 2022 for weekends and holidays. The data used are all from the authoritative platform: Investing.com Gold Historical Data. The data is divided into training sets and there are two parts of the test set, 2013-2021 as the training set, and the whole year of 2022 as the training set. The data field introduction table is shown in Table 1.

Table 1: Introduction to data fields

field	Description of the field	Extraction instructions
Date	The date of the transaction	It is recorded in the format of "month/day/year" to clarify the specific point in time of the data.
Price	The closing price of the day	It reflects the final transaction price at the end of the day's trading and is a key indicator to measure the day's trading results.
Open	The opening price of the day	It represents the initial price at the beginning of the day's trading, and can be compared with the closing price to analyze the price movement.
High	The highest price of the day	Shows the highest level reached by the price increase during the trading session of the day.
Low	The lowest price of the day	Reflects the lowest level touched by the price drop during the day's trading.
Vol.	Volume	Indicates the size of the number of transactions on the day and is used to assess the activity of trading in the market.
Change%	Percentage change for the day	By calculating the difference between the current day's closing price and the previous day's closing price, the price rises and falls as a percentage.

Date features are stored as objects in a data frame. To speed up the calculation, we convert its data type to datetime and then sort the feature in ascending order. Remove some unnecessary symbols, check for duplicates, missing items.

3.2 Evaluation indicators

In this paper, MSE, MAPE and Accuracy were used as evaluation indicators.
MSE is a measure of the mean squared difference between the predicted and true values [10].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{6}$$

MAPE is a measure of the relative error between the predicted value and the true value [10].

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y_i}}{y_i}\right| \tag{7}$$

The accuracy formula is shown below

$$Accuracy = 1 - MAPE \tag{8}$$

Among them, the variables y_i and $\hat{y_i}$ represent the true and predicted values of the gold price at time i, respectively.

3.3 Analysis of model solving

The chart of the gold price forecast is shown in Figure 4, and it is very clear that the actual situation of the gold price in 2022 is very similar to the previous forecast. This phenomenon shows that the results of the LSTM model are consistent with the actual situation in reality after a series of training. For example, in March 2022, the volatility of the gold price showed a high degree of agreement between the forecast and the actual value, which undoubtedly proves that the training effect of the model is quite good

As shown in Table 2, we can see more clearly and unambiguously that the MSE value is relatively low, and the deviation between the predicted gold price and the actual gold price is extremely small. The MAPE value is 0.03023, which is also at a relatively low level. This further confirms the accuracy of the forecast.

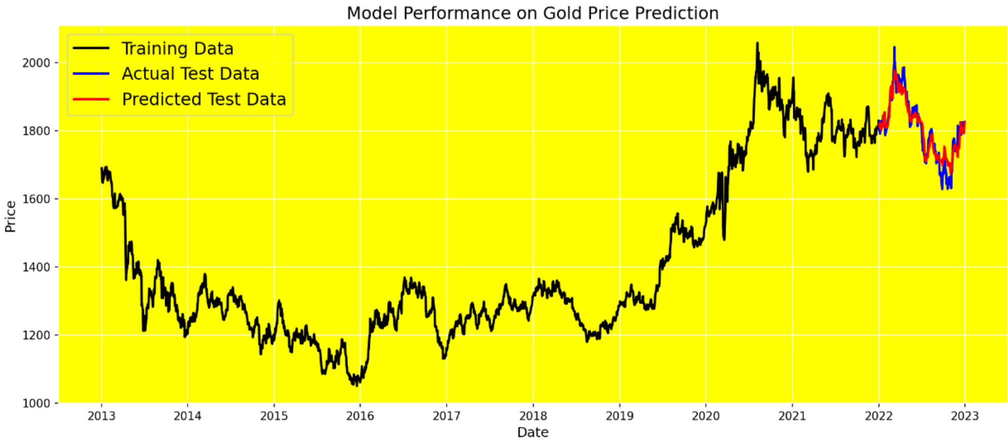

Figure 4: Gold Price Forecast Results

Table 2: Results of gold price prediction evaluation indicators

Test MSE	0.00081
Test MAPE	0.03023
Test Accuracy	0.96976

From the perspective of accuracy, the accuracy of the LSTM model is as high as 96.9% after a series of training. The accuracy fully demonstrates the excellent performance of the LSTM model in the training process of gold price prediction. The high accuracy and stability of its forecasts provide a solid and reliable basis for research and analysis of the gold market.

3.4 Conclusions of the experimental part

This paper pre-processes data for the decade from 2013 to 2023. In this process, irrelevant feature columns are decisively removed, and possible duplicates are screened and possible missing values are identified. Due to the obvious time series characteristics of these data, after comprehensive consideration and analysis, it was finally decided to select LSTM neural network for subsequent processing and analysis.

In the course of this training, the response of various evaluation indicators was excellent. In particular, the accuracy rate is a key indicator, which has reached an impressive 96.9%. This figure fully demonstrates the remarkable results achieved in this training, and also strongly proves the scientific and effective training method.

LSTM neural networks have strong predictive capabilities for gold prices and are able to delve into extremely large amounts of historical data. This is clearly an extremely valuable tool for those working in gold-related industries. Relying on the model's analysis of these historical data can help them predict the future development of gold prices more accurately. There are also significant benefits for gold buyers, as the use of the model can provide valuable decision-making aids to help them more effectively determine the best time to buy and sell.

4. Conclusions

This paper successfully constructed a gold price prediction model based on the LSTM model, which effectively solved the problem that gold prices are difficult to accurately predict due to the influence of multiple complex factors. The research results show that the built model has excellent prediction performance, with an accuracy rate of up to 96.9%. Through in-depth analysis and processing of data from 2013 to 2023, the model can accurately predict the movement and trend of gold prices, providing a valuable decision-making basis for gold industry practitioners. However, the model currently does not take into account quantitative indicators of the global political and economic situation. In subsequent research, the article will expand the sources and time span of data, enhance the generalization ability of the model, and add more influencing factors to make the model more comprehensive. In addition, the application of the model in different market environments and financial products will also be explored to provide broader support for prediction and decision-making in the financial field.

References

[1] Kang Yizhi. Analysis of historical fluctuation trajectory characteristics and influencing factors of gold price [J]. Modern Enterprise, 2022, (09): 94-96.

[2] Shi Yansong, Yang Bo. Prediction of gold futures price by CNN-LSTM model under RNN online learning framework [J]. Modern Information Science & Technology, 2024, 8(11): 141-144+152.

[3] Jia Degao, Liu Chunyu, Liu Jiapeng. Research on gold futures price trend based on hidden Markov model [J]. China Business Review, 2024, (11): 100-103.

[4] Su M , Nie Y , Li J ,et al.Futures markets and the baltic dry index: A prediction study based on deep learning[J]. Research in International Business and Finance, 2024, 71.DOI:10.1016/j.ribaf.2024.102447.

[5] Liang Longyue, Huang Ying. Research on short-term prediction method of gold futures price—COMEX gold futures price data analysis based on CEEMDAN and LSTM model [J]. Price Theory & Practice, 2023, (09): 164-168.

[6] He Linyun. Gold futures price prediction based on ICEEMDAN-SE-SSA-ELM algorithm [J]. Journal of Lanzhou University of Arts and Sciences (Natural Science Edition), 2023, 37(01): 35-39.

[7] Fang Qian, Yang Nan. Construction of nonlinear error correction learning model based on LSTM network [J]. Statistics and Decision, 2022, 38(24): 5-10.

[8] Man Ke, Wu Liwen, Liu Xiaoli, et al. TBM tunneling parameters and rockburst grade prediction based on CNN-LSTM model [J/OL]. Coal Science and Technology, 1-19 [2024-07-19].

[9] Li Fen, Sun Ling, Wang Yawei, et al. Short-term interval prediction of photovoltaic power based on CEEMDAN-GSA-LSTM and SVR [J]. Journal of Shanghai Jiao Tong University, 2024, 58(06): 806-818.

[10] Kou Yuxuan, Li Sihan, Liu Dunhu, et al. Research on Container Throughput Prediction Model Based on DKELM Selective Deep Integration [J/OL]. CMS, 1-25 [2024-07-20].