DOI: 10.54254/2754-1169/61/20231255

Time Series Prediction Based on LSTM Neural Network for Gold Prices

Wang Yiying[1,a,*]

[1]*Nanjing University, Department of Computer Science and Technology, Nanjing, 210023, China*
a. yiyingwang82@gmail.com
**corresponding author*

Abstract: Metals like gold have been traded around the world for ages. Gold, the yellow shiny metal, can maintain the stability of the market, disperse the risk of transfer price, and stabilize the national economy. Therefore, the research and prediction of gold prices is not only of great significance to the countries but also of great significance to individuals. However, compared with other financial time series, research on gold prices is still relatively few and lacks models integrated with deep learning. The purpose of this paper is to establish a new model for the time series prediction of gold prices and at the same time to provide new ideas for future financial time series prediction problems. In this paper, a prediction model based on LSTM (long-term memory Neural Network) is proposed by using the time series of large amounts of financial data as experimental data for prediction. The RMSE (root mean square error), which acts as a loss function, is selected to evaluate the accuracy of our prediction model. The model proposed has a good prediction effect on financial time series.

Keywords: time series forecasting, gold prices, LSTM, neural network, RMSE

1. Introduction

Gold, the yellow shiny metal, has been the fancy of mankind for ages. From making jewellery to being used as a way of investment, gold covers a huge spectrum of use.

Gold has a risk-averse value. Gold itself is stable and does not fluctuate greatly, making it the preferred investment choice for human currency. Gold T+D, gold ETF, gold concept stocks and other investment forms have long-term appreciation potential. While the price of gold is volatile in the short term, its safe-haven function is particularly important in times of political and economic instability to preserve the value of the asset.

Gold can suppress inflation, act as an indirect means of international payments, and guarantee the solvency of countries. In addition, gold reserves are also an important resource for stabilizing the national economy and curbing inflation. Investment in spot gold can outperform inflation and hedging value.

Gold is not only seen as a symbol of beauty and wealth but also has a high value in application. With the development of modern industry, together with science and technology, gold plays a significant role in the fields of space navigation, medicine, electronics and so on. Its more and more extensive use and increasing consumption have aroused attention and strong interest around the world. The existence of gold has brought convenience and well-being to human life.

DOI: 10.54254/2754-1169/61/20231255

As an important investment asset, the price fluctuation of gold is affected by many factors. Investors need to understand these factors when participating in the gold market to better grasp market opportunities.

Gold, like other metals, is also traded in the commodity futures trading platform all over the world. For better understanding of time series in a real-life scenario, we will work with data set of gold prices collected historically and predict its variation trend of future value. In this paper, the gold price is used as the experimental data to predict the time series of large amounts of financial data, and a prediction model based on long-term memory neural network is proposed, which has a good prediction effect on financial time series.

2. Gold Price Forecasting Problems Review

Jing R L et al. propose the LSTM model to predict the future price of gold and bitcoin within the acceptable range of error, which is beneficial for people to optimize their portfolio of the two assets [1]. Liang Y H forecasts gold prices by using a novel hybrid model based on ICEEMDAN and LSTM-CNN-CBAMMT. Firstly, the original gold prices are decomposed into different-frequency sublayers by ICEEMDAN. Then, the researcher combines three models, long short-term memory, convolutional neural network and convolutional block attention module, to forecast all the sublayers. It helps to improve the accuracy and generalization ability of predicting time series data. Eventually, the final result is to synthesize the predicted value of the sublayers by summation method [2]. Liu W J makes predictions by using the VMD-LSTM model based on the wavelet denoising method. Variational mode decomposition (VMD), suitable for nonlinear time series signals, applies the ideas of solving the variational problem to the signals, after the procession of which, an original signal is decomposed into several signals with different center frequencies. The researcher uses VMD to decompose the original time series, finds out the high-frequency noisy components, reconstructs the high-frequency noisy components, performs wavelet threshold denoising processing on the reconstructed sequences, and reconstructs all components into the LSTM model to build a VMD combination model based on wavelet denoising [3].

When the time series is too long, RNN is easy to forget the information in the previous relatively distant time period, the more recent the time point, the greater the impact [4]. In this paper, a prediction model based on LSTM for the time series of gold prices is proposed, which has a good prediction effect on financial time series. Moreover, the RMSE (root mean square error) acts as the loss function and is used to evaluate the accuracy of the prediction model.

3. LSTM Neural Networks

LSTM neural networks are a variant of recurrent neural networks (RNNS) and were proposed by Sepp et al. in 1997 [5]. Gate structure, which helps to solve the problem of information redundancy, is added in appropriate places. As it flows through neurons, the original information may well be enhanced since the gate structure allows information to be retained or abandoned selectively. The proportion of irrelevant information is weakened, and it makes sense to solve the problems of traditional gradient disappearance, gradient explosion and inability to deal with long-term dependence in recurrent neural networks [6].

The working mode of LSTM is very similar to the working mode of the human brain for memory. There are three main types of memory storage in the human brain: sensory stage, short-term memory, and long-term memory. The perception of the awareness stage is stored in the short-term memory, but the capacity of the short-term memory is very limited, and it is maintained for a very short time. The important short-term memory can be gradually transformed into long-term memory, and long-term memory can store a great deal of information for a relatively long period. The hidden state of

DOI: 10.54254/2754-1169/61/20231255

the LSTM is called the cells, and these cells are a repeating structure, each cell receives the previous cell state and the current input and decides what information to keep and what to forget, which can effectively preserve long-term information [7].

The key to constructing the LSTM model is the introduction of a set of memory units that enable the network to learn the timing of forgetting historical information and updating memory units with new information. At time t, all the historical information up to the end of the current moment are recorded by the memory unit c_t and the process is monitored and controlled by three "gates": the input gate i_t, the output gate o_t and the forgetting gate z_t. The value of elements of the three gates is between $[0,1]$. The forgetting gate f_t controls the amount of information required for each memory unit, the input gate i_t controls the amount of information newly added to each memory unit, and the output gate o_t controls the amount of information exported by each memory unit [8, 9].

$LSTM$ updates at time t in the following way:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + V_i c_{t-1}) \tag{1}$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + V_f c_{t-1}) \tag{2}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o c_{t-1}) \tag{3}$$

$$\widetilde{c_t} = \tanh\,(W_C x_t + U_c h_{t-1}) \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c_t} \tag{5}$$

$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

$i_t = \sigma(W_i x_t + U_i h_{t-1} + V_i c_{t-1}) f_t = \sigma(W_f x_t + U_f h_{t-1} + V_f c_{t-1}) o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o c_{t-1}) \widetilde{c_t} = \tanh\,(W_C x_t + U_c h_{t-1}) c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c_t} h_t = o_t \odot \tanh(c_t)$ Where, x_t is the input at time t; σ is a Logistic function; h_{t-1} is the external state at time $t-1$; $V_i, V_f\ and\ V_o$ are diagonal matrices; c_{t-1} is a $t-1$ moment memory unit. $\tilde{c}$ is the candidate function obtained by nonlinear function; W_f, W_i, W_c, W_o are the weight matrix of corresponding door multiplied by the input x_t and the output h_{t-1} of the middle-hidden layer.

When $f_t = 0$ and $i_t = 1$, the memory unit empties the historical information and loads all candidate states $\tilde{c}$. At this time, the memory unit c_t is still related to the historical information at $t-1$ moment. When $f_t = 1$ and $i_t = 0$, the memory unit will copy the contents of time $t-1$, and no new information will be loaded.

LSTM neural network parameters are learned by gradient descent method [10]. Given a training sample (x, y), $x_{1:T} = (x_1, \dots, x_T)$ is an input sequence of length T, and $y_{1:T} = (y_1, \dots, y_T)$ is a label sequence of length T. Define the loss function at time t as:

$$\tau_t = \tau[y_t, g(h_t)] \tag{7}$$

$\tau_t = \tau[y_t, g(h_t)]$Where, $g(h_t)$ is the output at time t and τ is the differentiable loss function, and the expression is:

$$\tau = \sum_{t=1}^{T} \tau_t \tag{8}$$

$\tau = \sum_{t=1}^{T} \tau_t$The whole gradient of the loss function τ with parameter U is:

$$\frac{\partial \tau}{\partial U} = \sum_{t=1}^{T} \frac{\partial \tau_t}{\partial U} \tag{9}$$

DOI: 10.54254/2754-1169/61/20231255

$\frac{\partial \tau}{\partial U} = \sum_{t=1}^{T} \frac{\partial \tau_t}{\partial U}$In the LSTM neural network, the gradient calculation method is back-propagation over time. The formula of calculating the gradient of the loss function τ with parameters U, W and b is as follows:

$$Z_k = Uh_{k-1} + Wx_k + b \tag{10}$$

$$\frac{\partial \tau}{\partial U} = \sum_{t=1}^{T} \sum_{k=1}^{t} \delta_{t,k} h_{k-1}^{T} \tag{11}$$

$$\frac{\partial \tau}{\partial W} = \sum_{t=1}^{T} \sum_{k=1}^{t} \delta_{t,k} x_{k}^{T} \tag{12}$$

$$\frac{\partial \tau}{\partial b} = \sum_{t=1}^{T} \sum_{k=1}^{t} \delta_{t,k} \tag{13}$$

$Z_k = Uh_{k-1} + Wx_k + b\,\frac{\partial \tau}{\partial U} = \sum_{t=1}^{T} \sum_{k=1}^{t} \delta_{t,k} h_{k-1}^{T}\,\frac{\partial \tau}{\partial W} = \sum_{t=1}^{T} \sum_{k=1}^{t} \delta_{t,k} x_{k}^{T}\,\frac{\partial \tau}{\partial b} = \sum_{t=1}^{T} \sum_{k=1}^{t} \delta_{t,k}$Where, Z_k is the hidden layer's net input at each time k $(1 \leq k \leq t)$; Error term $\delta_{t,k}$ is the derivative of the loss at time t with respect to the hidden layer's net input Z_k at time k [11].

4. Empirical Research

4.1. Data Analysis

The dataset used in this paper is time series of large amounts of financial data published on Kaggle. It is provided by *Quandl*, a platform for economic, financial and alternative datasets. It contains 10787 pieces of information on the daily gold prices featuring univariate time-series from Timestamp ('1970-01-01 00:00:00') to Timestamp ('2020-03-13 00:00:00')) and is a typical time series.

Set date column as an index. First of all, draw a chart of the gold prices over time, which can be seen in the following Figure 1 is that since 1970, the gold prices have been on the rise in general, and there are fluctuations. The price of gold reached a peak in 2012.

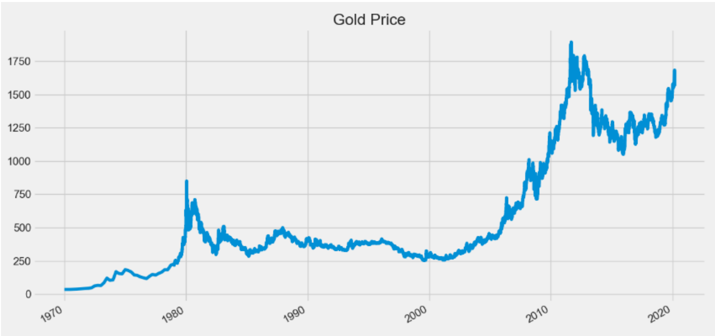

Figure 1: A chart of the gold prices over time.

4.1.1. Data Preprocessing

In the experiment, choose column of prediction and normalize data. The rules of normalization are as follows:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{14}$$

$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$In the paper, the MinMaxScaler class is used to normalize data. MinMaxScaler is a common way to scale data by linearly transforming the raw data into a range of minimum and

DOI: 10.54254/2754-1169/61/20231255

maximum values so that the data values fall within a predetermined range. The fit_transform method then performs two steps: first, it fits the given data (finding the minimum and maximum values), and then it transforms the data to scale the data to the specified range. This paper divides the preprocessed data set into two sections: training set and test set, with the training set accounting for 75% of the original and the test set for 25%. At this time, in the training set, the number of samples is 8090; the number of test samples is 2679, as shown in Table 1.

Table 1: Training set and test set after dividing the data set.

Data set	sample	proportion
Raw data set	10787	1
Training set	8090	75.00%
Test set	2697	25.00%

Then, create a training set with 100 time-steps and reshape the training set.

Time-steps is the number of time steps supposed to be used, the scope of each iteration will take continuous time-steps as input, and will take the next-time step as the target output. In particular, for each iteration, x_train is a two-dimensional array, said feature vector of continuous time step. And y_train is a one-dimensional array representing the feature vectors for the next time step.

The purpose of reshaping is to make the current data meet the input requirements of subsequent model training.

4.1.2. Model Training

Model parameters are as follows:

Table 2: Parameters of LSTM Model.

Layer(type)	Output Shape	Param
lstm (LSTM)	(None, 100, 50)	10400
lstm_1 (LSTM)	(None, 64)	29440
dense (Dense)	(None, 32)	2080
dense_1 (Dense)	(None, 16)	528
dense_2 (Dense)	(None, 1)	17

Fit the LSTM set to the training set, evaluate the model through losses. Figure 2 shows changes in losses of the LSTM neural network model during the process of model training.

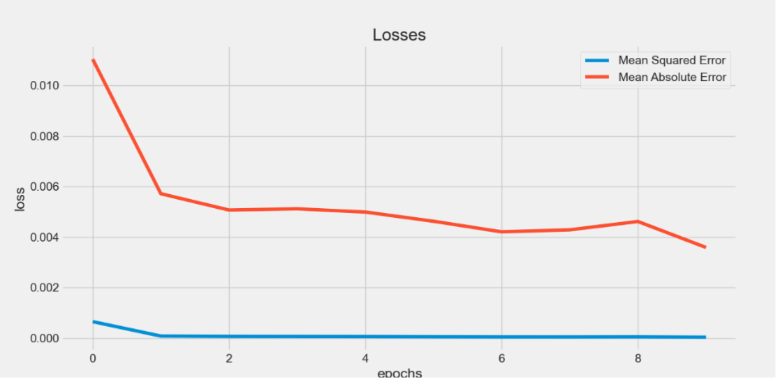

Figure 2: Changes in losses of the LSTM model.

DOI: 10.54254/2754-1169/61/20231255

Make model predictions and perform a restore operation on the predictions to bring them back to the scale of the original data. The predictions are stored and this step maps the input dataset to the corresponding predicted output based on the results of the model training. Next, restore predictions. Because the data has been normalized before, the normalized prediction results need to be converted back to the scale of the original data. Restoring it to the range of the original data allows a better comparison between the predicted results and the actual data.

4.2. Result Analysis

4.2.1. Evaluation Index

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{test}^{(i)} - \hat{y}_{test}^{(i)})^2} RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{test}^{(i)} - \hat{y}_{test}^{(i)})^2} \tag{15}$$

$RMSE$ (root mean square error) is used to measure the deviation from the observed value to the true value and it is the square root of the ratio, which is the square of the deviation to the number of observations N. $RMSE$ is also denoted by σ, which reflects the degree of deviation, that is, smaller σ indicates higher measurement accuracy.

4.2.2. Model Evaluation

Table 3: The predictive effect of the test set.

	Predictions	Actuals	MAPE
0	1112.566895	1114.75	2.183
1	1110.668945	1104.00	6.669
2	1101.577759	1130.00	28.422
3	1127.871216	1134.75	6.879
4	1130.406494	1149.00	18.594
2593	1604.819092	1655.70	50.881
2694	1589.775635	1653.75	63.974
2695	1586.996582	1570.70	16.297
2696	1508.904297	1562.80	53.896

Table 3 records the predictive effect of the test set and the predicted results are visualized in Figure 3. The orange line is the real price of gold, and the blue line is the predicted value of the model. It shows that the predicted curve generated by LSTM neural network model fits the actual curve well.

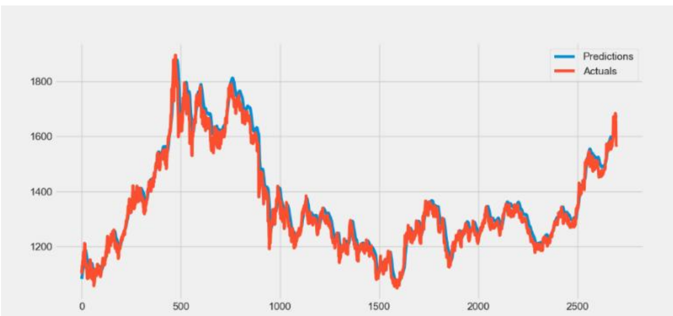

Figure 3: The results predicted by the model.

DOI: 10.54254/2754-1169/61/20231255

Table 4: Model prediction result.

Model	RMSE
LSTM	24.29

Use RMSE as an evaluation criterion for prediction accuracy and visualize it, as shown in the Table 4 above. The RMSE value of the final evaluation is 24.29 and the fitting accuracy is relatively high.

5. Conclusion

There is a lot of research gap in the prediction of gold future prices and thus, in this paper, a prediction model based on LSTM neural network is proposed. Compared with other time series prediction models, LSTM model can get prediction results of high precision by calculating the error evaluation index. The prediction results shows that the RMSE of the model is 24.29, indicating that the LSTM neural network can accurately predict the future price trend according to the long-term data changing with the time series, improve the prediction accuracy, and verify the high accuracy and applicability of the LSTM network, which will meet the demand of the actual future market.

References

[1] Jing R L, Qian H L. Analysis of gold and bitcoin price prediction based on LSTM model [J]. Academic Journal of Computing & Information Science. Volume 5.0, Issue 6.0. 2022

[2] Liang Y H, Lin Y, Lu Q. Forecasting gold prices using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAMMT [J]. Expert Systems With Applications. Volume 206, Issue. 2022

[3] Liu W J. Gold Futures price prediction and comparison based on VMD-wavelet denoising [D]. Jiangxi university of finance and economics, 2023.DOI: 10.27175/d.cnki.gjxcu.2022.001262.

[4] Qin X W, Wang Q J, Wang X M, et al. Short-term prediction of PM2.5 in Beijing based on VMD and LSTM methods [J]. Journal of Jilin University (Earth Science Edition), 202, 52(1): 214 - 221.

[5] Sepp H, Ju, Schmidhuber R. Long Short-Term Memory [J]. Neural Computation, 1997, 9(8): 1735 – 1780.

[6] Qiu X P. Neural Networks and Deep Learning [M]. Beijing: China Machine Press, 2020.

[7] Yuan D F, Gold Futures Price Prediction Based on CEEMDAN-PCA-LSTM Model [D], Shandong University, 2022.DOI: 10.27272/d.cnki.gshdu.2021.005040.

[8] Qian Z, PEI Y, CAO L, et al. Review of wind power prediction methods [J]. High Voltage Technology, 2016, 42(4): 1047-1060.

[9] Liu L L. Wind speed error correction model based on gated recurrent unit neural network for short-term wind power prediction [J]. Journal of Shenyang Institute of Technology (Natural Science Edition), 2021, 17(2): 11-17.

[10] LIU Y, GUAN L, HOU C, et al. Wind power short-term prediction based on LSTM and discrete wavelet transform [J]. Applied Sciences, 2019, 9(6): 1108.

[11] Wang H, Wang Z P, Liu X Y, Commutation failure boundary detection based on deep recurrent neural networks [J], Modern Electric Power, 2019, 36(6): 88 - 93.