International Core Journal of Engineering
ISSN: 2414-1895
Volume 8 Issue 8, 2022
DOI: 10.6919/ICJE.202208_8(8).0125

Analysis of Trading Strategy based on LSTM

Shaokang Zhang, Zizhan Wei

School of Science, North China University of Science and Technology, Hebei 063210, China

Abstract

To maximize their total returns, market traders frequently buy and sell volatile assets. According to the daily price data in the given 5 years, a transaction decision model based on Long Short Term Memory Network Model (LSTM) is established to judge whether traders should buy, sell or hold gold and Bitcoin. Preprocess the missing data in the attachment, then use the given data to build an LSTM model that can predict the currency price in the next 10 days, and perform 10,000 iterative training on it to finally obtain a prediction model with a loss rate of 0.00040486435. Make an investment rule of buying, selling or holding every 5 weeks, and evenly allocate the initial $1,000 to bitcoin and gold for investment, and redistribute the funds according to the future income of the two, with the highest price during the trade cycle Sell at the lowest point, buy at the lowest point. The maximum return that ended up with $1,000 was $5,985.6348.

Keywords

LSTM Model; Price Forecast; Trading Decision.

1. Establishment of LSTM Model [4]

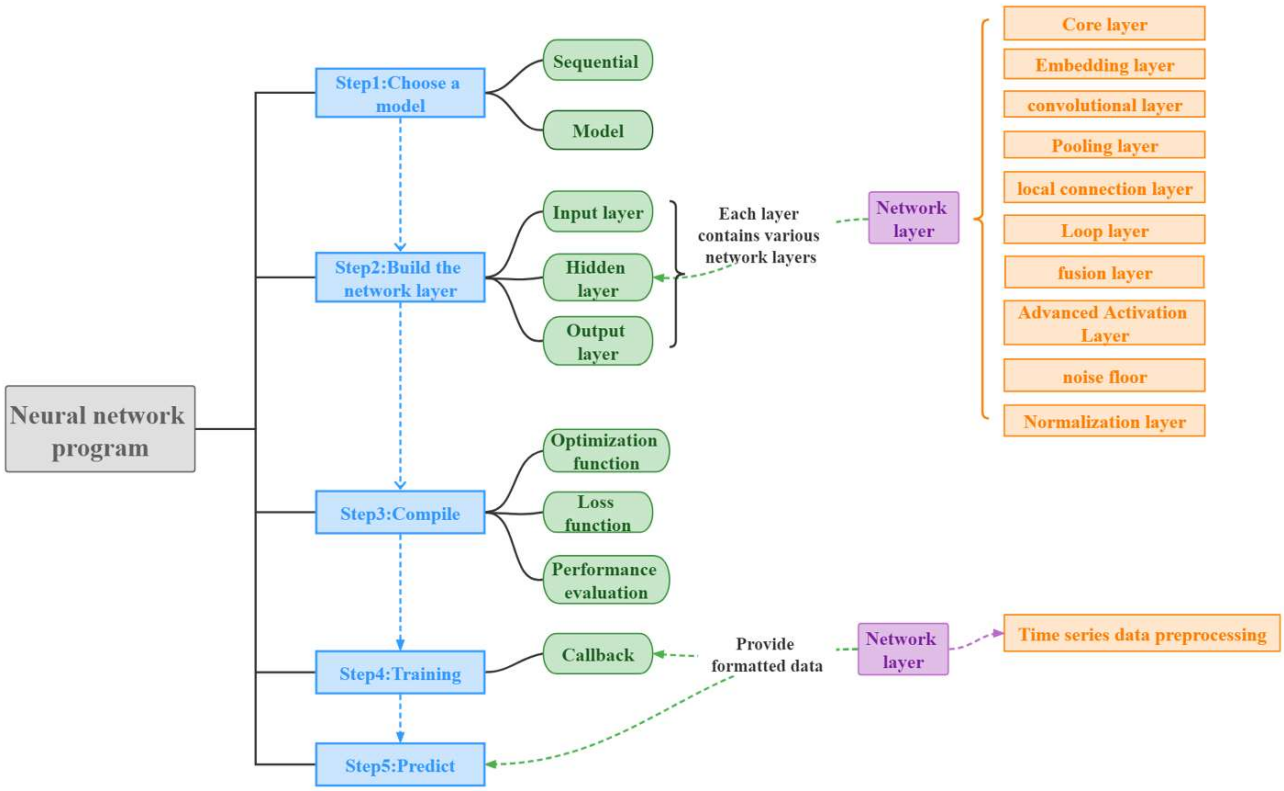

Figure 1. Neural network structure diagram

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0125

In recent years, machine learning and deep learning have achieved rapid development and have been widely used in many engineering fields [1]. The input of the model in this paper is multi-modal time series data, which has typical time characteristics. The results obtained by using the model are not only related to the current input, but also related to the past input. Therefore, the recurrent neural network can better solve the problem. However, when the time series is too long, the recurrent neural network has problems such as gradient explosion and gradient disappearance [2]. Therefore, the selection of the long short-term memory network (LSTM) model can make full use of the recurrent neural network, which is currently the most advanced for sequence learning. One of the deep learning architectures for the task [3], the basic process of the neural network is shown in Fig. 1.

2. LSTM Neural Network Structure

The LSTM network is composed of the input layer, the hidden layer and the output layer like the RNN. At the same time, a state unit c is added on the basis, and the information selectively affects the LSTM network at each moment through the forget gate, the input gate, and the output gate. state, the internal structure of LSTM is shown in Figure 1. The LSTM unit has 3 inputs at time t: the input value xt of the network at the current time, the output value ht-1 of the LSTM hidden layer at the previous time, and the unit state ct-1 at the previous time. The LSTM unit has 2 outputs at time t: the output value ht of the hidden layer at the current time and the unit state ct.

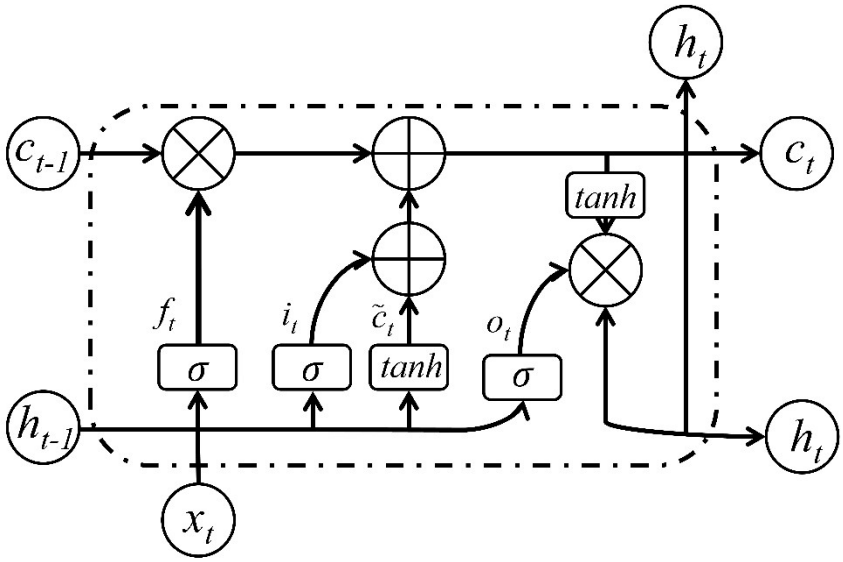

Figure 2. LSTM network internal structure diagram

The first is the forget gate:

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

Among them, f_t is the proportional coefficient that controls the forgetting information at the current moment; σ is the sigmoid activation function, which is used to smoothly map the data obtained after the operation to the (0,1) interval, which just corresponds to the degree of switching of the "control gate"; ω_f is the weight matrix of the forget gate; b_f is the bias term of the forget gate.

$$[\omega_f]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} = [\omega_{fh} \quad \omega_{fx}]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \tag{2}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0125

Among them, the weight matrix ω_f consists of two weight matrices, one is ω_{fh} and the other is ω_{fx}, corresponding to the input items h_{t-1} and x_t respectively. The function of the forget gate is to let LSTM "forget" the useless information before, assuming that the input time series of LSTM at time $x_t = {x_t}^1, {x_t}^2, {x_t}^3, \cdots, {x_t}^n$, hidden layer output at t-1 time $h_{t-1} = {h_{t-1}}^1, {h_{t-1}}^2, {h_{t-1}}^3, \cdots, {h_{t-1}}^n$, take the two as inputs and obtain the coefficient f_t that determines the proportion of forgotten information through formula (1), the information in the dimension where f_t is close to 0 will be completely forgotten, and f_t Information on dimensions close to 1 is fully preserved.

Next is the input gate:

$$i_t = \sigma(\omega_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$[\omega_i]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} = [\omega_{ih} \quad \omega_{ix}]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \tag{4}$$

Among them, i_t is the proportional coefficient of the control input information at the current moment; ω_i is the weight matrix of the input gate, which is composed of ω_{ih} and ω_{ix} weight matrices corresponding to h_{t-1} and x_t; b_i is the bias term of the input gate.

$$\widetilde{c_t} = tanh(\omega_c \cdot [h_{t-1}, x_t] + b_c) \tag{5}$$

$$[\omega_c]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} = [\omega_{ch} \quad \omega_{cx}]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \tag{6}$$

Among them, c_t is the candidate unit state at the current moment; $tanh$ is the activation function. Different from the sigmoid activation function, the tanh activation function smoothly maps the data to the (-1,1) interval; ω_c is the weight matrix of the candidate unit state, which is determined by the corresponding It is composed of ω_{ch} and ω_{cx} weight matrices of h_{t-1} and x_t; b_c is the bias term of the candidate unit state.

$$c_t = f_t \circ c_{t-1} + i_t \circ c_t \tag{7}$$

The LSTM network needs to supplement the latest information from the current input, and this process is done through the "input gate". First, the "input gate" obtains the proportional coefficient i_t of the control input information through equation (3) to determine which information in the candidate unit state is added to the previous unit state c_{t-1} to generate a new state c_t. Secondly, the candidate unit state c_t is obtained by taking h_{t-1} and x_t as input through equation (5). Finally, through formula (7), f_t is multiplied element-wise by the previous unit state c_{t-1}, and then the current candidate unit state c_t is multiplied by i_t element-wise, and the two are added to obtain the current unit state c_t.

And finally the output gate:

$$o_t = \sigma(\omega_o \cdot [h_{t-1}, x_t] + b_o) \tag{8}$$

$$[\omega_o]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} = [\omega_{oh} \quad \omega_{ox}]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \tag{9}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0125

Among them, o_t is the proportional coefficient of the control output information at the current moment; ω_o is the weight matrix of the output gate, which is composed of the ω_{oh} and ω_{ox} weight matrices corresponding to h_{t-1} and x_t; b_o is the bias term of the output gate.

$$h_t = o_t \circ tanh(c_t) \tag{10}$$

The output process of LSTM generating the current moment is completed through the "output gate". First, the parameter o_t that controls the switching degree of the "output gate" is obtained by formula (8), and then the latest unit state c_t is multiplied by the corresponding parameter of o_t through the activation function $tanh$ operation by formula (10) to obtain the output h_t at the current moment.

Specifically in this paper, the input x of the LSTM network model is the filtered power consumption characteristic data and power quality historical monitoring data, and the final model output is the predicted value of the power quality index data.

3. LSTM Network Training

The current mainstream training methods for neural network models such as LSTM [6]are BPTT (Back Propagation Through Time) and RTRL (Real-Time Recurrent Learning). Both are algorithms based on gradient descent, in which the iterative direction of RTRL is consistent with the direction of forward propagation, and the amount of calculation is large, while the BPTT algorithm has a clear concept and is computationally efficient, and has more advantages than RTRL in terms of computational time. Therefore, BPTT is used in this paper. Algorithms to train LSTM networks.

The training of LSTM network is mainly divided into four steps:

1) Input the data of the divided training set into LSTM, and calculate the output value of each neuron forward. For the LSTM network, it is the value of the five vectors of f_t,i_t,o_t,h_t and c_t. The relevant expressions have been described in 5.1.1.

2) The error term δ is calculated in reverse for each neuron. The error term of the LSTM propagates in two directions: back-propagation in time; propagating the error term to the upper layer.

3) Recalculate the gradient of each weight matrix according to the error term, and add the gradients of the bias terms b_f,b_i,b_o, and b_c at each moment to obtain the final gradient [7].

4) The corresponding weights are updated according to the gradient and the learning rate, and the adaptive momentum estimation (Adam) gradient optimization algorithm is used.

4. Solution of the Model

According to the principle of the LSTM model in 5.1.1, the time series prediction models of Bitcoin and gold are established respectively based on the data given in the appendix. First, preprocess the data given in the attachment, fill in the vacant value of the gold value table in the attachment with the gold price of the previous day, and then use the data given in the attachment for 10,000 trainings, and finally obtain a prediction model with a loss rate of 0.00040486435 [9].

The specific results are as follows:

Table 1. Model training loss rate

Frequency	1	2	3	4
Loss Rate /%	3.6475	3.5431	4.6291	3.0636
Frequency	5	6	7	8
Loss Rate /%	1.9064	1.3747	1.0742	0.8735
	……			
Frequency	9997	9998	9999	10000
Loss Rate /%	0.0418	0.0405	0.0418	0.0405

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0125

According to the daily currency price, use the forecast model to predict the currency price in the next fifteen days, and get the currency value trend. For better investment, make the following rules:

1) Take 5 weeks as the limit for one buy and sell, and only one buy and sell or no trade can be made within 5 weeks.

2) In the initial stage, $1,000 will be equally divided between gold and bitcoin as initial investment funds, that is, the weight of gold and bitcoin in the initial stage are both 0.5. As the investment progresses, the weights are reconfigured based on the returns of the two currencies.

The specific principles are as follows:

Initial stage: $\omega_{BTC} = \omega_G = 0.5$, that is, Bitcoin and Gold each have USD 500 as the initial capital.

When a trading cycle ends:

$$Invest_{BTC} = \frac{Income_{BTC}}{Income_G + Inco_{BTC}}(Weight_G + Weight_{BTC}) \tag{11}$$

$$Invest_G = \frac{Income_G}{Income_G + Inco_{BTC}}(Weight_G + Weight_{BTC}) \tag{12}$$

$\mathrm{Invest_{BTC}}$, $\mathrm{Invest_G}$ represent the investment principal of Bitcoin and gold at the begin -ning of the next transaction respectively; $\mathrm{Income_{BTC}}$, $\mathrm{Income_G}$ respectively represent the income of Bitcoin and gold after the end of a trading cycle, $\mathrm{Weight_{BTC}}$, $\mathrm{Weight_G}$ represent the selling price of bitcoin and gold, respectively.

The buying and selling points of one of the trade cycles are determined as follows:

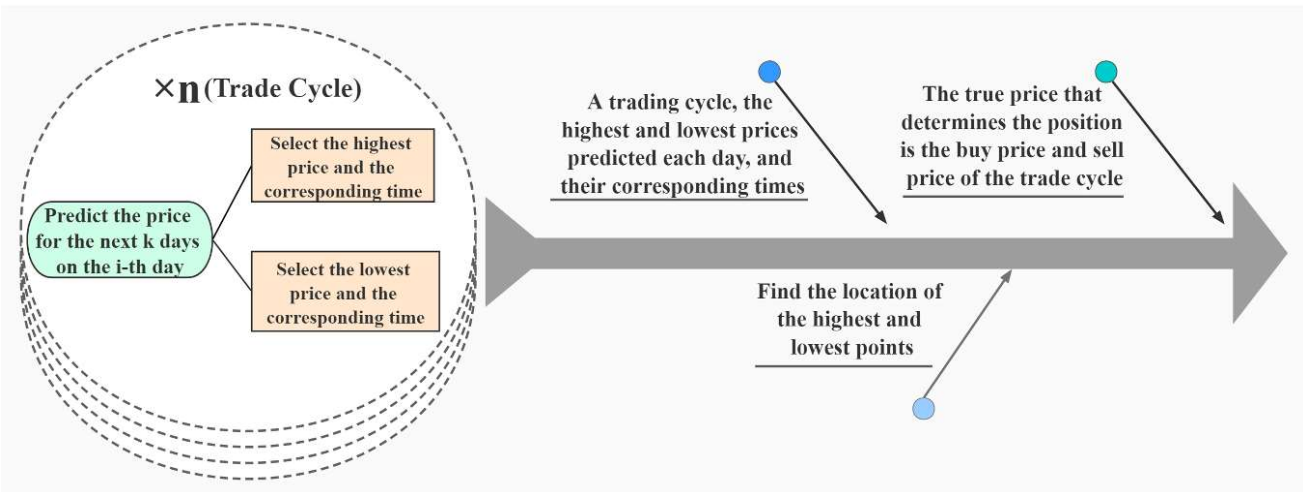

Figure 3. Determination of buy and sell points in a trade cycle

In order to get the maximum profit, it is necessary to find the maximum and minimum prices of Bitcoin and gold in each cycle to buy and sell, buy at the lowest price point, and sell at the highest point. In a trading cycle, there are the following situations:

Volume 8 Issue 8, 2022
DOI: 10.6919/ICJE.202208_8(8).0125

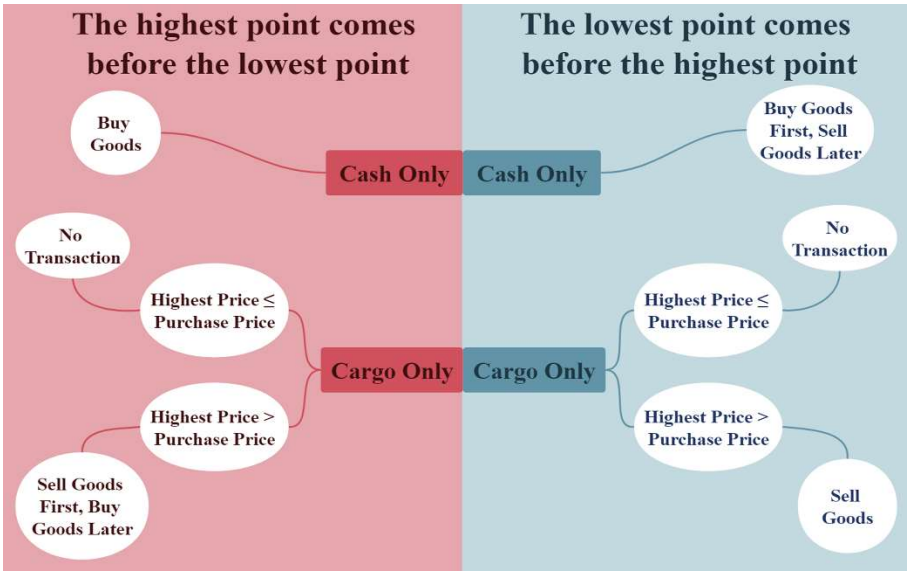

Figure 4. One-cycle buying and selling strategy

Earnings are calculated using the following formula:
Cash Only: Need to buy Bitcoin or Gold.

$$C_{BTC/G} = \frac{cash(1-a)}{price_1} \tag{13}$$

Cargo Only: Need to sell Bitcoin or Gold.

$$cash = C_{BTC/G} \cdot price_2(1-a) \tag{14}$$

According to the model built above, using python to calculate, a total of 48 transactions were performed, and the final result was 5985.6348 on September 10, 2021, and the income amount was 4985.6348 (see appendix for specific data). Calculating the profitability of each transaction, it can be seen that the use of this strategy will lead to a decrease in revenue. This is because the prediction accuracy of the LSTM model cannot reach 100%. In addition, the prices of Bitcoin and gold are often affected by external factors, and only rely on price for investment. Earnings are not guaranteed to be increasing all the time.

References

[1] Ma Shilong, Unni Riqiqige, Li Xiaoping (Ma Shilong, Unni Riqiqige, Li Xiaoping). Deep Learning with Big Data: state of the art and development [J]. Intelligent System Journal (CAAI Transactions on Intelligent Systems), 2016, 11(6): 728-742.

[2] Sun Yingying, Jia Zhentang, Zhu Haoyu (Sun Yingying, Jia Zhentang, Zhu Haoyu). Survey of multimodal deep learning [J]. Computer Engineering and Application, 2020, 56 (21): 1-10.

[3] Liu Jianwei, Liu Yuan, Luo Xionglin. Research progress of deep learning [J]. Application Research of Computers, 2014, 31(7): 1921-1930, 1942.

[4] Zhao Changwei, Pian Ruijun, Du Tianshuo, Ge Leijiao. LSTM-based power quality trend prediction and analysis model for important users [J/OL]. Journal of Electric Power Systems and Automation: 1-8 [2022-02-20]. DOI: 10.19635/j.cnki.csu-epsa.000949.

[5] Gao Yang, Yang Fuliang. "The Influence of Transaction Costs in Futures Market on Transaction Volume and Strategies for Reducing Transaction Costs." Chinese Journal of Nonferrous Metals. 01(1998): doi:10. 19476/j.ysxb.1004.0609.1998.01.040.

[6] Wang Xin, Wu Ji, Liu Chao, Yang Haiyan, Du Yanli, Niu Wensheng. Fault Time Series Prediction Based on LSTM Recurrent Neural Network [J]. Journal of Beihang University, 2018, 44(04):772-784 . DOI: 10.13700/j.bh.1001-5965.2017.0285.

[7] Zhu Qiaomu, Li Hongyi, Wang Ziqi, Chen Jinfu, Wang Bo. Ultra-short-term prediction of wind farm power generation based on long-short-term memory network [J]. Power Grid Technology, 2017,41(12): 3797-3802. DOI : 10.13335/j. 1000-3673. pst. 2017.1657.

[8] Xiong Wei, Chen Langnan. Currency idiosyncratic volatility, currency return and investor sentiment [J]. Management Science, 2015, 28 (05): 106-115.

[9] Zhang Lei. Research on Currency Market Trend Analysis and Forecast Based on Support Vector Machines [D]. Nanjing University of Posts and Telecommunications, 2020. DOI: 10.27251/d.cnki.gnjdc. 2020. 000099.