

Effectiveness validation of LSTM for stock prices prediction on four stocks

Mingxuan Huang

School of Software, Shandong University, Jinan, Shandong, 250101, China

201900301189@mail.sdu.edu.cn

Abstract. Due to the significant volatility and complexity of financial data, stock price prediction is a difficult undertaking. Researchers have also begun using various models to predict stock prices. LSTM (Long short-term memory) are used in this study to predict the stock prices of four companies between 2013 and 2019. The author compares the performance of two loss functions, MSE (Mean squared error) and MAE (Mean absolute error), and evaluate the effectiveness of the proposed approach. The experiments show that LSTM is a promising model for stock price prediction, and the data's long-term dependencies can be captured by it. The predicted stock price curves closely match the real data curves, although there are some limitations in predicting the changes in stock prices. Furthermore, the study demonstrates that the choice of loss function has a significant impact on the prediction results. As a result, it's crucial to choose the loss function carefully depending on the dataset's features. Overall, the research shows that LSTM and other machine learning models have great potential for stock price prediction. Yet it's crucial to remember that the stock market is incredibly volatile and uncertain, and these models should be used as complementary tools to assist human experts in making investment decisions.

Keywords: deep learning, LSTM, stock price prediction.

1. Introduction

Stock prediction is a significant research field in the financial sector. Research into stock forecasting dates back to the early 20th century, when the stock market just emerged, and investors started to use fundamental and technical analysis to predict stock prices [1,2]. Fundamental analysis is a way to predict by looking at things like a company's financial health, the overall economy, and market movements. Technical analysis is another method, which predicted the future trend of stocks by analyzing technical indicators such as stock prices and trading volumes. However, the accuracy of this method was often limited by issues such as data reliability and timeliness of historical data [3,4].

In the 1960s and 1970s, scholars began to use mathematical models to predict stock prices. The earliest mathematical models were based on linear regression, which predicted future price trends by fitting historical data [5]. However, the accuracy of this method was difficult to achieve as stock price changes were influenced by multiple factors and involved nonlinear relationships.

With the development of computer technology, researchers began to use more complex and advanced models to predict stock prices, such as neural network models and SVM (support vector machine). Among them, neural network models were widely used in stock prediction, especially in the late 1990s,

with the introduction of the LSTM model, which further improved the prediction ability of neural network models. This article will also use LSTM for stock prediction [6-8].

2. Method

2.1. Dataset and preprocessing

This dataset is collected from Yahoo Finance [9,10]. It has historical price information for 3 high growth stocks and bitcoin from May 2013 to May 2019. An overview of the data is shown in Figure 1, which displays price line charts of four stocks from 2013 to 2019.



Figure 1. Prices of the four stocks.

As can be seen, the prices of various stocks fluctuate greatly, the range for AMZN (Amazon) is 250 to 2000, the range for DPZ (Domino's Pizza) is 50 to 300, the range for BTC (Bitcoin) is 50 to 17500, and the range for NFLX (Netflix) is 30 to 400, this may produce anomalous samples and cause adverse effects. The author here will skip the model building step for now and directly use raw data for training, taking BTC as an example, and its performance on the test set is shown in Figure 2. It is evident that if the raw data is not processed, the training loss will be difficult to reduce, resulting in extremely poor test results. Therefore, the first step of data preprocessing is normalization, which maps the data to the range of 0 to 1, eliminates the adverse effects caused by anomalous samples, and enables the model to converge faster.

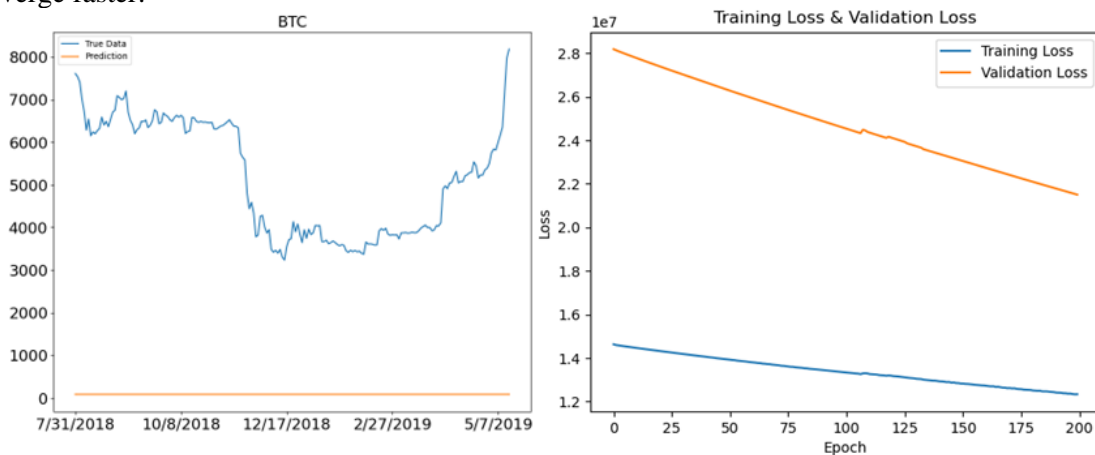


Figure 2. Prediction and convergence situation without data preprocessing.

2.2. LSTM

Before proceeding to the second step of data preprocessing, it is important to consider how LSTM works. LSTM, a type of recurrent neural network model, could solve the deficiencies lie in previous models, which is the long-term dependencies when processing sequence data. Its basic cell is illustrated in Figure 3. It contains four main operations, including input, forget, output, and cell state ones. The first gate is input gate, it chooses how much new information will be put into the cell state. Which outdated information should be ignored or maintained is decided by the forget gate. Output gate determines what information should be output. The cell state in LSTM can be thought of as a memory unit that can pass information and update when necessary.

Overall, the basic workflow of LSTM is as follows: it receives an input sequence, which is typically a sequence of input vectors across multiple time steps. Then, the input sequence goes through respective activation functions to calculate the aforementioned gates and cell state.

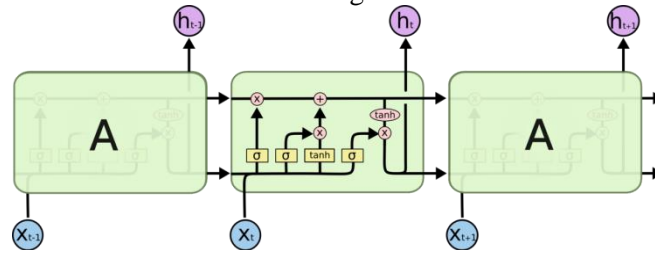


Figure 3. Structure of LSTM basic cell (image from: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>).

So according to the working principle of LSTM, it is necessary to properly split the data into sequences in order to make it a suitable input for the model. First, set the sequence length, for example 30, and then slice this set of data. The thirtieth data point is estimated leveraging its previous 29 data points. Next, simply shift by one and repeat the process with the next 30 numbers, until the entire training set is used.

2.3. Model constructing

LSTM is usually combined with other types of neural network layers to build more complex models. For example, adding a Dropout layer can randomly drop out some neurons during the training process to prevent overfitting. A Dense layer can also be added to map the output to a continuous numerical value as the prediction result for regression tasks.

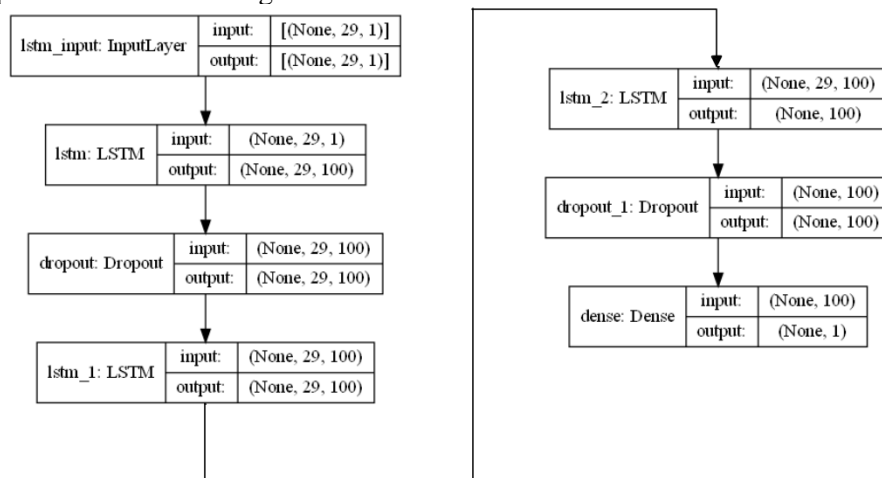


Figure 4. Architecture of the LSTM model leveraged in this research.

As depicted in Figure 4, the first layer is a LSTM layer with 100 neurons, time step of 29, and return_sequences set to true. A Dropout layer makes up the second layer, which has the rate of 0.3. The

third and the fourth layer are both LSTM layers but the fourth layer's return sequences is set to false. Then, it has a dropout layer and a dense layer, which make up the last two layers.

All models were trained for 200 epochs, using the loss functions MSE, and MAE during training. MSE is more sensitive to significant errors because it calculates the squared difference between actual and predicted values. MAE calculates the absolute difference between real and predicted values, and is more suitable for handling outliers because it treats all errors equally.

2.4. Evaluation indexes

The performances are measured by MSE and MAE. They could be calculated as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (2)$$

where n is the quantity of samples for evaluating and Y_i and \hat{Y}_i denote the true label and corresponding output results.

3. Result

3.1. Price prediction

In the test set, the time range is from 7/31/2018 to 5/14/2019. According to the method above, the sequence length is 30, with Adam optimizer with 200 epochs.

It can be seen from Figure 5 that the results of the four stocks fluctuated significantly during this period in the test set. The predicted data curve reflects most of the details of the changes and fits closely with the real data.

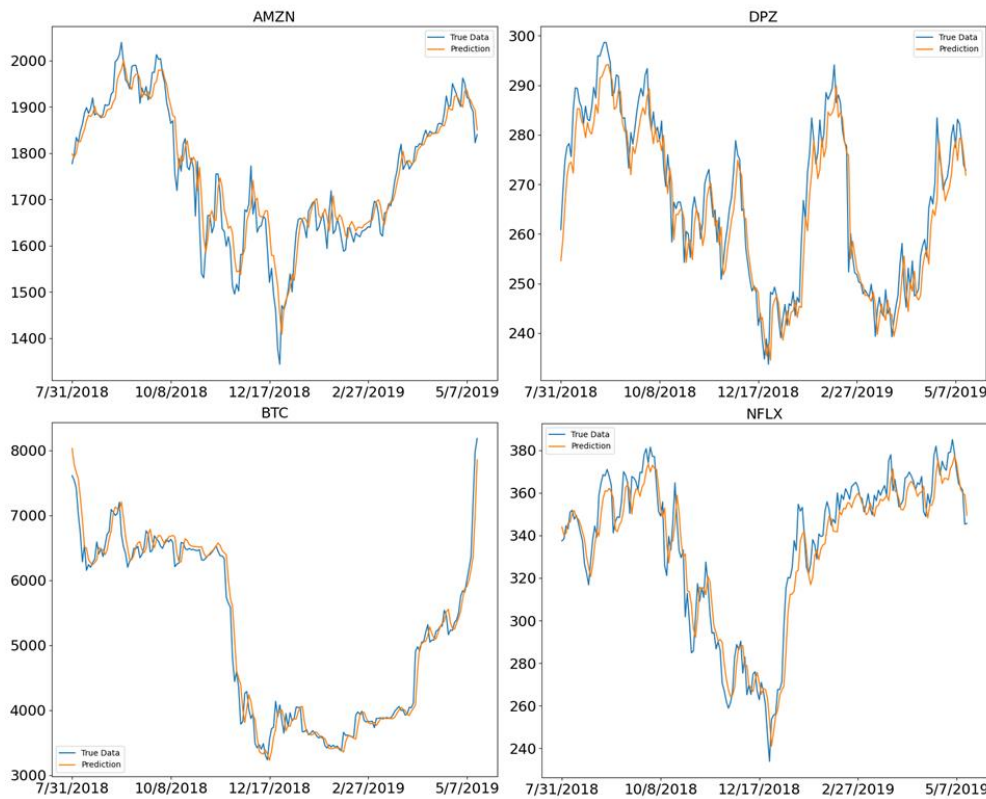


Figure 5. Prediction of the four stocks.

During the training of the model, MSE and MAE were used as loss functions separately, and the model was trained and tested 10 times on each stock. The performances are demonstrated in Table 1. It can be seen that using MSE and MAE as loss functions does not make much difference on the datasets of

AMZN and BTC. However, using MSE as the loss function has better performance on DPZ, while using MAE as the loss function has better performance on NFLX.

Table 1. Quantitative results of the four stock predictions.

	MSE Loss		MAE Loss	
	MSE	MAE	MSE	MAE
AMZN	0.000822209	0.022476487	0.000911061	0.023815813
BTC	0.000234275	0.011441033	0.000242462	0.011839177
DPZ	0.000553763	0.018286963	0.000948514	0.024826606
NFLX	0.001077164	0.02532778	0.000669694	0.020063605

3.2. Effectiveness of early stopping

In addition, EarlyStopping and ModelCheckpoint were used. EarlyStopping is a regularization technique used to prevent overfitting of the model. During the model training process, the main strategy is to keep an eye on the performance on the validation set and cease training when the performance stops getting better, thus preventing the model from overfitting on the training set. Therefore, setting the monitor as val_loss, training can be stopped when there is no improvement in the test set loss within the corresponding patience. Meanwhile, by using ModelCheckpoint, the model with the lowest val_loss can be saved.

Table 2 shows the average size of the loss generated by using AMZN as the dataset, MSE as the loss function, and testing 10 times with and without EarlyStopping. It is evident that using EarlyStopping generates a smaller loss compared to not using it, which directly proves the superiority of EarlyStopping.

Table 2. Effectiveness of early stopping.

	Loss
With Early Stopping	0.000822209
Without Early Stopping	0.001307001

4. Discussion

Of course, LSTM models also have disadvantages, which will be shown below. (1) Dataset: LSTM needs to learn long-term dependencies, which often requires a big amount of training data. If the dataset is insufficient or of low quality, it can significantly affect the prediction results. (2) Inability to handle sudden events: In real life, there are many unpredictable events, such as natural disasters, wars, and other uncontrollable factors that can greatly affect the market. However, LSTM and similar models learn and predict based on past data, making it difficult to react to sudden events. (3) Difficulty in handling other information: The input data of LSTM models usually only include historical stock price data. However, there are many other factors that can affect stock prices, such as recent policy information and international situations. If these factors cannot be analyzed and integrated into the model, it will be difficult to make accurate predictions. (4) Lagging: As with other time series models, LSTM requires historical data as input. Due to the internal structure of the model, changes in data in the previous time step will be calculated and reflected in the next time step, leading to a certain degree of lagging and less accurate predictions.

Although using LSTM for stock prediction seems to have some shortcomings, there are also methods for optimization and improvement to address these issues. The most direct approach is to combine other models for prediction, such as using text data from news events, social media, policy information, etc., and analyzing them with NLP (Natural language processing) -related models. This not only solves the singularity of input data but also improves the reliability of predictions.

5. Conclusion

This article uses LSTM model to predict stock prices, with historical data of four stocks between 2013-2019, using both MSE and MAE as loss functions. LSTM, as a variant of RNN (Recurrent neural network), has shown certain ability in solving time-series problems such as stock price prediction and

can uncover long-term dependencies. From the results, the predicted curve and the real data curve fit closely, but there is some room for improvement in reflecting the ups and downs of stock price changes. Meanwhile, choosing either MSE or MAE as the loss function can also have a certain impact on the prediction results, and this requires comparison and selection based on the specific situation of the dataset. Currently, using LSTM or other models for stock price prediction has a promising future, and researchers are constantly exploring and trying other methods. However, it should be noted that even if more data information can be combined and other models can be used for more reliable prediction, it is still important to understand that there is no perfect model that can predict stock prices with certainty, and there will always be some degree of uncertainty and unpredictability in the market. Overall, all models are just auxiliary tools, and people still need to have professional knowledge to make better judgments.

References

- [1] Avdalovic, S. M., & Milenković, I. (2017). Impact of company performances on the stock price: An empirical analysis on select companies in Serbia. *Economics of Agriculture*, 64(2), 561-570.
- [2] Masulis, R. W. (1980). Stock repurchase by tender offer: An analysis of the causes of common stock price changes. *The Journal of Finance*, 35(2), 305-319.
- [3] Lu, W., Li, J., Wang, J., & Qin, L. (2021). A CNN-BiLSTM-AM method for stock price prediction. *Neural Computing and Applications*, 33, 4741-4753.
- [4] Hu, Z., Zhao, Y., & Khushi, M. (2021). A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4(1), 9.
- [5] Soni, P., Tewari, Y., & Krishnan, D. (2022). Machine Learning approaches in stock price prediction: A systematic review. In *Journal of Physics: Conference Series*, 2161(1), 012065.
- [6] Li, H., Shen, Y., & Zhu, Y. (2018). Stock price prediction using attention-based multi-input LSTM. In *Asian conference on machine learning*, 454-469.
- [7] Ding, G., & Qin, L. (2020). Study on the prediction of stock price based on the associated network model of LSTM. *International Journal of Machine Learning and Cybernetics*, 11, 1307-1317.
- [8] Pramod, B. S., & Pm, M. S. (2021). Stock price prediction using LSTM. *Test Engineering and Management*, 83, 5246-5251.
- [9] Lawrence, A., Ryans, J., Sun, E., & Laptev, N. (2018). Earnings announcement promotions: A Yahoo Finance field experiment. *Journal of Accounting and Economics*, 66(2-3), 399-414.
- [10] Ferdiansyah, F., Othman, S. H., Radzi, R. Z. R. M., Stiawan, D., Sazaki, Y., & Ependi, U. (2019, October). A lstm-method for bitcoin price prediction: A case study yahoo finance stock market. In *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*, 206-210.