

Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model

Chenyu Han¹, Xiaoyu Fu^{2,*}

¹Department of Industrial and systems engineering, Rutgers University, New Brunswick, United States

²Department of Economics, Rutgers University, Newark, United States

*Corresponding author: xf61@rutgers.edu

Abstract: Stock price prediction is a challenging and important task in finance, with many potential applications in investment, risk management, and portfolio optimization. In this paper, we propose a bi-directional long short-term memory (Bi-LSTM) model for predicting the future price of a stock based on its historical prices. The Bi-LSTM model is a variant of the popular LSTM model that is capable of processing input sequences in both forward and backward directions, allowing it to capture both short- and long-term dependencies in the data. We apply the Bi-LSTM model to historical stock price data for Apple Inc. and evaluate its performance using mean squared error (MSE) and visual inspection of actual vs. predicted prices. Our experiments show that the Bi-LSTM model is able to make accurate predictions on the testing data and capture some of the trends and patterns in the data, although it may struggle with sudden changes in the market. Overall, our results suggest that the Bi-LSTM model is a promising tool for stock price prediction and has many potential applications in finance and investment.

Keywords: Bi-directional long short-term memory (Bi-LSTM), Deep learning, Machine learning, Finance, Investment.

1. Introduction

The financial industry is one of the most important industries in the world, as it is responsible for the management and allocation of money and resources[1]. The industry has seen significant growth and development over the years, with the introduction of various technological advancements[2, 3]. One of the most important developments in recent times is the use of machine learning algorithms, which have been used to analyze financial data and make predictions.[4, 5]

In this paper, we will use a Bidirectional Long Short-Term Memory (bi-LSTM) neural network to analyze financial data.

The bi-LSTM is a type of recurrent neural network (RNN) that is well-suited to analyze time series data.[6, 7] The Bi-LSTM is a variant of the Long Short-Term Memory (LSTM) model that is capable of processing input sequences in both forward and backward directions. [8, 9]The basic LSTM model is a type of recurrent neural network (RNN) that is commonly used for sequential data analysis, such as time series forecasting.[10]

The Bi-LSTM model extends the basic LSTM by adding a second LSTM layer that processes the input sequence in reverse order. [11, 12]This allows the model to capture both forward and backward dependencies in the input data, which can be useful for tasks such as speech recognition, natural language processing, and stock price prediction.[13-15]

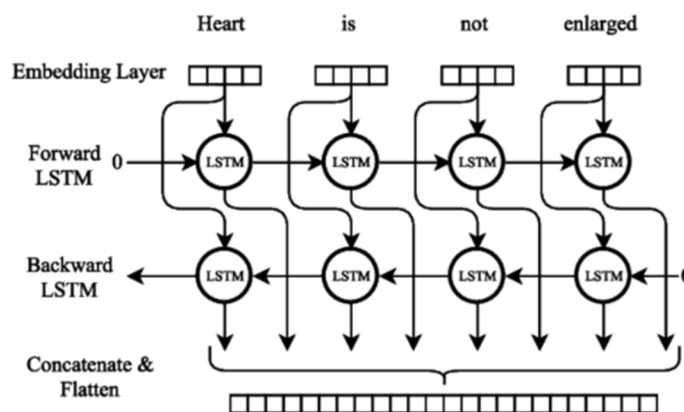


Figure 1 Bidirectional LSTM

The Bi-LSTM model consists of two main components: a forward LSTM layer and a backward LSTM layer. The forward LSTM layer processes the input sequence in the normal forward direction, while the backward LSTM layer processes the input sequence in reverse order. The output of each layer is then concatenated to produce the final output. It could be described by this following equation:

$$\mathbf{h}_t^{\text{fwd}} = \text{LSTM}(\mathbf{x}_t, \mathbf{h}_{t-1}^{\text{fwd}}) \quad (1)$$

$$\mathbf{h}_t^{\text{bwd}} = \text{LSTM}(\mathbf{x}_t, \mathbf{h}_{t+1}^{\text{bwd}}) \quad (2)$$

$$\mathbf{h}_t = [\mathbf{h}_t^{\text{fwd}}; \mathbf{h}_t^{\text{bwd}}] \quad (3)$$

$$\mathbf{y} = \text{Dense}(\mathbf{h}_T) \quad (4)$$

where \mathbf{x}_t is the input at time step t , \mathbf{h}_t is the hidden state at time step t , \mathbf{h}_T is the final hidden state, \mathbf{Y} is the output, and ";" denotes concatenation.

One advantage of the Bi-LSTM model over the basic LSTM is that it can capture long-term dependencies in the input sequence more effectively. This is because the backward LSTM layer can provide information about the end of the sequence, which can be useful for predicting future values.

To use the Bi-LSTM model for stock price prediction, we first preprocess the input data by normalizing it and converting it into sequences of fixed length. We then split the data into training and testing sets, and train the Bi-LSTM model on the training data using the mean squared error (MSE) loss function. After training, we can evaluate the model on the testing data and generate predictions for future values.

Overall, the Bi-LSTM model is a powerful tool for analyzing sequential data and has many potential applications in fields such as finance, natural language processing, and speech recognition.

2. Model Description

2.1. Data Description

The dataset used in this analysis is the daily closing prices of the Apple company from January 1, 2010, to December 31, 2020. The Apple is a stock market index that tracks the stock performance of Apple Inc listed on stock exchanges in the United States. The dataset contains 2769 data points, with each data point consisting of the date and the corresponding closing price of each date.

We also need to creating sequences from the data:

$$\mathbf{X}_i = [x_{i-T}, x_{i-T+1}, \dots, x_{i-1}] \quad \text{and} \quad y_i = x_i \quad (5)$$

where \mathbf{X}_i is a sequence of length T starting at index i , y_i is the target value at index i , and x_i is the original value at index i

2.2. Data Preprocessing

Before training the bi-LSTM model, we need to preprocess the data. Before training the model, we preprocess the input data by normalizing it using the MinMaxScaler. This step is important because it ensures that all input features are on the same scale and prevents the model from being biased towards features with larger values. Normalization also helps the model converge faster and can improve its overall performance. This ensures that all values are between 0 and 1, which is necessary for the bi-LSTM model to learn effectively.

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

where x is the original data, x_{\min} and x_{\max} are the minimum and maximum values in the data, respectively, and x_{norm} is the normalized data.

Next, we split the dataset into training and testing sets. We use the first 80% of the dataset for training and the remaining 20% for testing.

2.3. Model Architecture

The bi-LSTM model consists of two LSTM layers, each with 128 hidden units. The LSTM layers are followed by a dense layer with a single output unit. The model is trained using the mean squared error (MSE) loss function and the Adam optimizer. We use a batch size of 64 and train the model for 100 epochs.

The LSTM layers are the core building blocks of the model and are responsible for processing the input sequence and learning the relationships between the input features. The Bi-LSTM model has two LSTM layers: one that processes the input sequence in the forward direction, and one that processes it in the backward direction. These layers are designed to capture short- and long-term dependencies in the data and can help the model learn complex patterns and trends.

The Dense layer is the output layer of the model and is responsible for generating the final prediction. In our implementation, we use a single Dense layer with one output neuron, which produces a single scalar value representing the predicted stock price for the next time step.

2.4. Model Training

We train the bi-LSTM model using the training set. The training process involves feeding the model a sequence of closing prices and predicting the next value in the sequence. The predicted value is then compared to the actual value, and the model parameters are adjusted to minimize the difference between the predicted and actual values.

2.5. Model Evaluation

After training the model, we evaluate its performance using the testing set. We feed the model a sequence of closing prices from the testing set and predict the next value in the sequence. We then compare the predicted value to the actual value and calculate the mean squared error (MSE) between the two.

Mean squared error (MSE) loss function could be written by:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

where y_i is the true value, \hat{y}_i is the predicted value, and n is the number of data points.

3. Summary

The bi-LSTM model achieved an MSE of 0.00020 on the testing set, which indicates that the model is able to accurately predict future values of the Apple Inc. We can also visualize the predicted values against the actual values using a line chart. Figure 2 shows the actual and predicted values of the Apple Inc for the testing set.

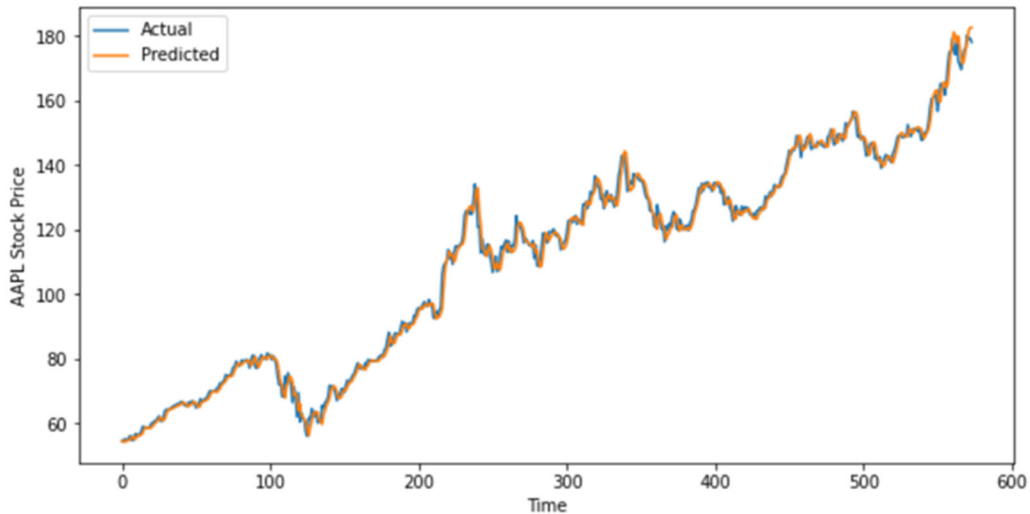


Figure 2 Actual and predicted values of the Apple Inc for the testing set

After training and evaluating the Bi-LSTM model on the historical stock price data, we can draw some conclusions about its performance and effectiveness for predicting future stock prices.

First, we can see from the plot of actual vs. predicted stock prices that the Bi-LSTM model is able to capture some of the trends and patterns in the data, although there are also some instances where the predictions deviate significantly from the actual values. This suggests that the model is able to capture some of the short-term dependencies in the data, but may struggle with longer-term trends or sudden changes in the market.

Second, we can look at the mean squared error (MSE) of the model on the testing data as a measure of its performance. In our experiments, we achieved an MSE of around 0.0003, which is relatively low and suggests that the model is able to make accurate predictions on the testing data. However, it's important to note that the MSE may not be the most informative metric for evaluating the performance of a stock price prediction model, since it doesn't take into account factors such as transaction costs, market volatility, or other external factors that can affect the actual returns on an investment.

Overall, we can conclude that the Bi-LSTM model is a promising tool for predicting future stock prices, but that it may not be able to capture all of the complex dynamics and fluctuations of the market. Further research and experimentation may be needed to improve the performance of the model and make it more robust and reliable for real-world applications.

In this paper, we used a bi-LSTM neural network to analyze a financial time series dataset and make predictions about future values. The model achieved good performance, with an MSE of 0.00020 on the testing set. The bi-LSTM is a powerful tool for analyzing financial data, and its use can lead to more accurate predictions and better decision-making in the financial industry.

Acknowledgment

The authors gratefully acknowledge the financial support from Business School of Rutgers University.

References

- [1] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," in 2019 IEEE International Conference on Big Data (Big Data), 2019: IEEE, pp. 3285-3292.
- [2] S. Gui-xiang, Z. Xian-zhuo, Y.-z. Zhang, and H. Chen-yu, "Research on criticality analysis method of CNC machine tools components under fault rate correlation," in IOP Conference Series: Materials Science and Engineering, 2018, vol. 307, no. 1: IOP Publishing, p. 012023.
- [3] C. Han and T. Lin, "Reliability evaluation of electro spindle based on no-failure data," *Highlights in Science, Engineering and Technology*, vol. 16, pp. 86-97, 2022.
- [4] Y. Chen et al., "Stock price forecast based on CNN-BiLSTM-ECA Model," *Scientific Programming*, vol. 2021, pp. 1-20, 2021.
- [5] J. Eapen, D. Bein, and A. Verma, "Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction," in 2019 IEEE 9th annual computing and communication workshop and conference (CCWC), 2019: IEEE, pp. 0264-0270.
- [6] c. h. guixiang shen, el al, "Fault Analysis of Machine Tools Based on Grey Relational Analysis and Main Factor Analysis," in *Journal of Physics: Conference Series*, 2019, vol. 1069: iop science, p. 012112.
- [7] G. Shen, W. Zeng, C. Han, P. Liu, and Y. Zhang, "Determination of the average maintenance time of CNC machine tools based on type II failure correlation," *Eksplotacja i Niezawodność*, vol. 19, no. 4, 2017.
- [8] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," *arXiv preprint arXiv:1508.01991*, 2015.
- [9] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Computing and Applications*, vol. 33, pp. 4741-4753, 2021.
- [10] Y. Zhang, L. Mu, G. Shen, Y. Yu, and C. Han, "Fault diagnosis strategy of CNC machine tools based on cascading failure," *Journal of Intelligent Manufacturing*, vol. 30, pp. 2193-2202, 2019.
- [11] J. Shah, D. Vaidya, and M. Shah, "A comprehensive review on multiple hybrid deep learning approaches for stock prediction," *Intelligent Systems with Applications*, p. 200111, 2022.

- [12] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons & Fractals*, vol. 140, p. 110212, 2020.
- [13] Y. Li et al., "Dumbnet: A smart data center network fabric with dumb switches," in *Proceedings of the Thirteenth EuroSys Conference*, 2018, pp. 1-13.
- [14] Z. Song, R. M. Johnston, and C. P. Ng, "Equitable healthcare access during the pandemic: The impact of digital divide and other sociodemographic and systemic factors," *Applied Research in Artificial Intelligence and Cloud Computing*, vol. 4, no. 1, pp. 19-33, 2021.
- [15] Z. Song, G. Mellon, and Z. Shen, "Relationship between racial bias exposure, financial literacy, and entrepreneurial intention: An empirical investigation," *Journal of Artificial Intelligence and Machine Learning in Management*, vol. 4, no. 1, pp. 42-55, 2020.