Analysis and forecast of Amazon stock price based on LSTM algorithm

Yunxiang Liu*,
Guangzhou College of Commerce,
Guangzhou, China,
739047146@qq.com

Abstract—**In the 21st century, the stock is a high-risk, high-yield common financial products, in order to better analyze the stock investment, effectively and accurately predict the stock price trend has become the current research hotspot in recent years. In this paper, we put forward a LSTM prediction model to predict the closing stock prices of Apple. Experimental results indicate that this method displays better prediction performance than other machine learning methods. Specifically, the RMSE value is 24.43% and 16.69% lower than ridge algorithm and neural network algorithm respectively. In addition, we make a detailed data analysis of the data set, including the stock price changes over time, the average daily return and the moving average of the stock, and give how much risk to invest in a specific stock, which can provide some guidance for our future work.**

Index Terms—Stock Forecast, LSTM, Machine learning algorithms, Data Analysis,

I. INTRODUCTION

At present, with the continuous improvement and maturity of the financial market, the stock market has become an indispensable part of the world economic development. Because of its high rate of return, more and more investors choose to buy stocks as their way of investment [1]. However, due to the influence of economic environment, political policy, market behavior and other factors, the stock market always fluctuates frequently and the stock price is unstable, which would bring greater investment risk to investors [2]. Therefore, the prediction of stock price trend has become one of the hot issues of scholars all over the world. In this paper, we mainly forecast the stock prices of different companies which is provide by KAGGLE competition.

In recent years, as the quickly developing of computer technology, machine learning has been the current focus of research, its application fields and scope continue to expand and extend, including the financial information field [3]. Recently, many scholars use machine learning algorithms to predict the future price of stocks, including ridge regression, genetic algorithm, support vector machine and so on; with the rise of neural network technology, by building a deep neural network model to establish the stock price model and predict the trend of stock price, it has been deeply studied by many scholars

[4, 5]. Therefore, this paper mainly uses the stock prices dataset of several companies to predict the future stock trend of corresponding companies, including Apple, Google, Microsoft, Amazon and other companies. In this paper, we use the LSTM (long-short term memory) algorithm model based on neural network to predict the trend of Apple's stock closing price. The results show that this algorithm has more accurate prediction than the other algorithms. LSTM model is a type of gated recurrent neural network, and it has excellent prediction performance in time series problems and is a popular model in the field of quantitative investment [6].

We used the historical stock price time series data set of four international companies on the platform of KAGGLE to study the method of analyzing stock risk, and used LSTM algorithm to predict stock price. In this paper, Python was mainly used to process, model and analyze the above data. In the experimental stage, we first analyzed the relevant data of stock, including the change of stock price with time, the average daily return and moving average of stock, etc. Then, we predict the closing price of apple stock based on LSTM algorithm, and compare with other algorithms, including Ridge, Nerual Network and so on. The results indicate that the comprehensive form of our algorithm is better than other algorithms. Specifically, the RMSE of LSTM algorithm is 24.43% lower than Ridge algorithm and 16.69% lower than Nerual Network. And the correlation between different stocks is given, and some interesting conclusions are obtained, which can provide some guidance for future work.

In a word, the main innovation points and contributions are:

● We raise a prediction model using on LSTM algorithm to solve the task of stock trend prediction, which can effectively deal with the prediction problem with time series.

●In the process of the experiment, we first analyze the data related to the stock price, including the change of the stock price with time, the average daily return rate, the moving average value of the stock, the correlation between different stocks and the risk of investing in a specific stock, which can guide the related work in the future.

● In the prediction of stock price trend, the performance of LSTM algorithm model proposed in

this paper is better than other machine learning algorithms. More accurate trend prediction can provide certain guidance for financial investment.

A. Related Work

Data mining and analysis can find valuable and effective information from a large number of data, and the conclusions can be used to provide decision-making and support for investors. With the rapid development of science and technology, many scholars combine data analysis with the application of the stock market, with the full help and application of Python, SPSS, MATLAB and other software to achieve data processing in recent years [7]. And then complete the analysis and trend prediction of the stock price. It is necessary and feasible to use data analysis method to study the development trend of stock market.

In 2003, Fan *et al*. [8] proposed a linear prediction model ARMA based on statistical method. Although this mode can make linear approximation for complex stock system, it can't predict the evolution of nonlinear process and non-stationary process. With the continuous development of machine learning technology and neural network, many nonlinear algorithms have been developed to predict more complex problems. For example, Yao *et al*.[9] used Ridge Regression algorithm to predict stock market trend and obtained good results. Qiu *et al*. [10] used genetic algorithm to optimize the artificial neural network model and forecast the price trend of Japanese stock market index. Besides, Back propagation [11], radial basis function [12] and recurrent neural network [13] have been widely used to predict the development trend of financial time series. In reference [14], the LSTM algorithm is used to predict the stock price, and the time series learning is carried out through the historical stock price data. The obtained results present that the LSTM algorithm has better form in time series prediction. But this paper does not consider the impact of different types of data on the stock price, which makes the model too simple and lack of robustness. Therefore, this paper uses different dimensions of data and LSTM model to predict the stock price.

The structure of this paper is as follows. Section II introduces the data set and data analysis process. Next, the models, including Ridge Regression, Neural Network and LSTM, is introduced in Section III. Then, we analyzed the performance of different algorithms using RMSE value. Finally, the main conclusions in this paper and future work are given in Section V.

II. DATA ANALYSIS

In this section, we introduce the datasets used in this paper and the relevant processes for data analysis.

The stock price data set mainly contains the historical stock price data of four technology companies from 2012 to 2021, including apple, Amazon, Microsoft and Google on the platform of KAGGLE competition. In the dataset, there are seven fields, namely company name, opening price, highest price and lowest price of the day, closing price, adjusted closing price and daily volume, so that we can study the trend of stock prices from different aspects.

Then, we use Python to analyze the stock related data, mainly including the stock price changes over time, the average daily return and the moving average of the stock, and give how much risk to invest in a specific stock. Taking Microsoft as an example, we give moving averages in different time periods, as shown in Figure.1. It can be seen that the change trend of the moving average in different time periods is the same, which is an important basis for judging the stock market [7]. However, when the time is too long (50 days), the change trend of the moving average becomes more gentle.

Figure.1 Moving averages of Microsoft
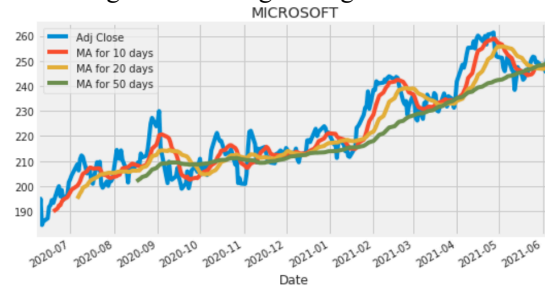
III. ALGORITHM MODELS

In this section, we introduce several prediction algorithm models used in this paper, including ridge regression, neural network and LSTM.

Ridge regression is an improved least square estimation method, which can be used to estimate collinear data. Compared with the traditional linear regression model, Ridge Regression algorithm adds a penalty term of L2 norm to the target function [15]. The objective function of Ridge regression can be expressed as:

$$J(\beta) = \sum(y - X\beta)^2 + \lambda\|\beta\|_2^2,$$

where λ is always positive. when λ is larger, in order to make $J(\beta)$ minimum, regression coefficient β is even smaller. Because of the penalty term, the accuracy of the algorithm would decrease, but the ability to solve ill conditioned matrix problems is greatly improved, and the robustness of the algorithm is also improved.

Artificial neural network model is a highly complex and nonlinear model, and can be applied to solve complex problems. The algorithm is based on the mechanism of information transmission of neurons in human brain, and aims to simulate the operation mode of brain. Building a complete neural network mainly includes two steps: building a model and determining the weight. Generally speaking, there are two kinds of network structure: forward type and feedback type. A very important ability of neural network is to constantly adjust the weights and thresholds of neurons, so that the network can constantly learn from the environment. When the output error is zero or less than the set error threshold, the model is established [5].

LSTM model is a method of time series prediction analysis, which can effectively use the characteristics of an event in the past to predict the characteristics of the event in the future. Compared with the traditional neural network algorithm, LSTM algorithm can calculate the weight of the memory part and the new information part, and use the weighted average to replace the matrix multiplication, so as to realize the long-term memory. A standard unit of LSTM model is given below:

Figure. 2. The basic unit of LSTM algorithm
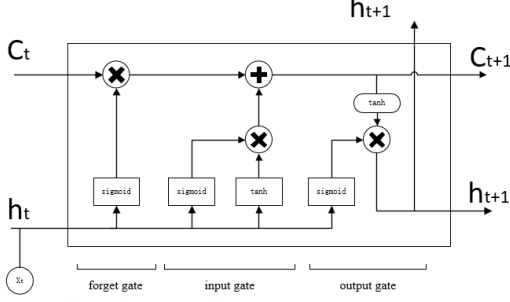
where $X^{(t)}$, $h^{(t)}$, and $c^{(t)}$ are the input, output and cell state at time t respectively. The function of cell state is to provide some memory to remember the previous data. In an LSTM model unit, it mainly includes three kinds of "gates" functions, including forget gate, input gate and output gate, which are controlled by sigmoid function.

IV. EXPERIMENTS

In this part, the prediction results and experimental analysis of different algorithms are introduced separately.

First, we give the prediction results of closing price of apple stock based on LSTM algorithm, as shown in Figure.3, where the blue, red and yellow lines represent training data, testing data and prediction data. It can be seen that the predicted values and the real values (testing data) have the same change trend, and the curves almost coincides.

Figure. 3. The closing price of Apple's stock
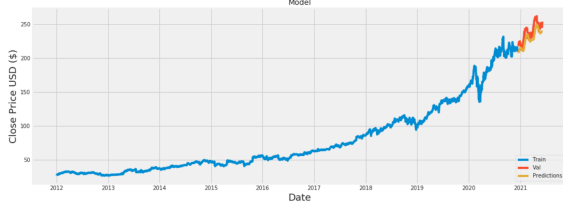
Then, in order to comprehensively evaluate the prediction performance of different algorithms, we use RMSE (Roo t Mean Square Error) index as the evaluation index. Generally speaking, the lower the RMSE value, the better the performance of the stock forecasting model. The general formula of RMSE is given as:

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(y_i - y_i')^2},$$

where M is the number of days, indicating the amount of data in the testing set. y_i and y_i' are the real and predicted closing prices of Apple's stock.

The RMSE values of different algorithms are shown in Table 1.

Obviously, our LSTM model has the lowest RMSE value. Specifically, the RMSE value of our proposed model is 2.13 lower than the Ridge algorithm and 1.32 lower than the Nerual Network. Therefore, results show our method using LSTM algorithm is effective for stock price forecast task.

Models	RMSE
LSTM	6.59
Ridge	8.72
Nerual Network	7.91

Table 1. The RMSE values on stock dataset for price forecast task.

Besides, through the analysis of stock related data, we give the correlation between the closing prices of different stocks, as shown in Figure. 4. It can be seen that the daily closing prices of Apple and Amazon stock have the greatest correlation. In addition, interestingly, the stock prices of all technology companies are positively correlated, which can guide our future work to a certain extent.

Figure.4 Correlation of closing prices among different companies

V. CONCLUSIONS

Based on the stock price data set provided by KAGGLE competition platform, this paper combines LSTM algorithm with data analysis. Experiments show that this method can effectively predict the closing price of Apple's stock, and its performance is better than ridge algorithm and traditional neural network algorithm. In addition, we analyzed the data set in detail and get some interesting and constructive guidance. In the future, we will improve our model to improve the forecasting effect, and apply our model to other stock trend problems.

REFERENCES

[1] Zhang Qianqian, Lin Tianhua, Qi Xuyang, et al. A Review of Stock Forecasting Based on Machine Learning[J]. Journal of the Hebei Academy of Sciences, 2020, 37(4):15-21.

[2] Chen Weibin, Lin Yizzhen, Wang Zongyue. Stock Information Mining and LSTM

Prediction[J]. Journal of Jimei University(Nature Science), 2020, 25(5):385-391.

[3] Gurav U , Sidnal N . Predict Stock Market Behavior: Role of Machine Learning Algorithms[M]. 2018.

[4] Thawornwong S , Dagli C H , Enke D L . Using Neural Networks and Technical Analysis Indicators for Predicting Stock Trends[J]. 2001.

[5] Liang X . Neural Network Method to Predict Stock Price Movement Based on Stock Information Entropy[C]// International Conference on Advances in Neural Networks. Springer-Verlag, 2006.

[6] Gf A , Schmidhuber J , F Cummins. Learning to Forget: Continual Prediction with LSTM[M]. Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale, 1999.

[7] Xu Xiaoxiao, Wang Lizhu, Cheng Congdian. A Method of Finding Stock Buying Point Based on Moving Average Price and Big Data Analysis[J]. 2020, 9(1):1-6.

[8] Fan Jianqing, Yao Qiwei. Nonlinear time series: Nonparametric and parametric methods[M]. Science Press, 2006.

[9] Yao Hongliang, Ma Xiaoqin, Wang Hao, et al. Stock Market Trend Prediction Algorithm Based on Morphological Characteristics and Causal Ridge Regression[J]. Computer Engineering, 2016, 042(002):175-183.

[10] Qiu M , Yu S . Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model[J]. Plos One, 2016, 11(5):e0155133.

[11] Kara Y , Boyacioglu M A , mer Kaan Baykan. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange[J]. Expert Systems with Applications, 2011, 38(5):5311-5319.

[12] Wei S A , Xg A , Chao W B , et al. Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm[J]. Knowledge-Based Systems, 2011, 24(3):378-385.

[13] Lalu J , Binu J A . A Survey on Time-Series Data Prediction Models Using Recurrent Neural Networks[M]. 2021.

[14] Bao W , Yue J , Rao Y , et al. A deep learning framework for financial time series using stacked autoencoders and long-short term memory[J]. PLoS ONE, 2017, 12(7):e0180944.

[15] Hoerl A E , Kennard R W . Ridge Regression[M]. American Cancer Society, 2006.