Two models for predicting stock prices in combination with LSTM

Yu Ma*

College of mathematics, Sichuan University, Chengdu, China

* Corresponding Author Email: 1455955616@qq.com

Abstract. In addition to its practical and theoretical significance, stock forecasting has long been a hot research topic for scholars domestically and abroad. Stock data are time-series in nature, and neural networks have achieved relatively good performance in dealing with time series problems, among which long-short-term memory neural networks are well suited to dealing with such time-series data with long-term dependence. However, the stock market is an environment that changes with the external environment, with high stochasticity and complex intrinsic nonlinear relationships between different phenomena. Relying on a single method to identify the series directly cannot fully extract the complex information of the series changes, so the combined forecasting method is proposed. One idea is to combine empirical mode decomposition (EMD) with long-short-term memory (LSTM), and another idea is to incorporate LSTM or its variant GRU into generative adversarial networks (GAN/CGAN/WGAN). After an empirical study of Guanhao Bio's stock price and Apple's stock price, both methods present better prediction results.

Keywords: Stock Price Forecast, EMD, LSTM, CGAN, WGAN.

1. Introduction

1.1. Research Background

With the continuous development of applied mathematics, mathematical models have been widely used in the financial industry, and the academic branch of financial mathematics has been derived. At present, the analysis methods of financial mathematics mainly use traditional time series models such as ARMA models, ARCH-type models. These traditional methods only analyze the series from the perspective of time-aligned data acquisition, but the calculation efficiency is low, the calculation error is large, and the research content is single.

Modern financial mathematics time series forecasting continues to accept the benefits of modern technology, and the application of artificial intelligence methods to time series forecasting, such as support vector machines, deep learning, etc. Deep learning is widely used in socially critical financial areas, where stock market prediction is one of the most popular and valuable areas in finance. Deep learning algorithms can extract features from large amounts of raw time series data without relying on a priori knowledge, making them well suited for financial time series forecasting, especially LSTM networks with long-term memory due to their recurrent structure. However, the stock market is an environment that changes with the external environment, with high stochasticity and complex intrinsic nonlinear relationships between different phenomena. Relying on a single method to directly identify the series cannot fully extract the complex information about the series changes, so combined forecasting methods are proposed.

One idea is to combine empirical mode decomposition (EMD) with long-short-term memory (LSTM), commonly used to analyze nonlinear, non-stationary high-frequency data in the time and frequency domains. Empirical mode decomposition (EMD) is an important component of the Hilbert-Huang transform, a new method for processing non-stationary signals, proposed by Dr. Norden e. Huang, a Chinese-American scientist at NASA, in 1998. The EMD-based time-frequency analysis method is suitable for analysing both nonlinear and non-stationary signals and linear and stationary signals. It also reflects the physical meaning of the signal better than other time-frequency analysis methods for linear and stationary signals.

Another idea is to incorporate LSTM or its variant GRU into generative adversarial networks (GAN), both of which have better prediction results. GAN is widely used for image generation but rarely applied to time series prediction [1]. Since few studies use GAN for time series prediction, their findings are inconsistent. This paper aims to use LSTM based on its variant GRU as a generator of GAN to predict stock prices and check whether adversarial neural networks can help improve time series prediction.

1.2. Research Methodology

1.2.1EMD-LSTM

Akita et al. [2] compared a long short-term memory network LSTM with a traditional neural network and found that LSTM has good fitting performance for stock series data. Peng Yan et al. [3] performed wavelet noise reduction in data preprocessing on the basis of LSTM network, and the prediction effect was improved. Jiannan Shi et al. [4] proposed an LSTM stock price prediction method based on dynamic modal decomposition (EMD), which has higher price prediction accuracy than traditional prediction methods in a specific market context. In traditional forward neural networks such as BPNN, the neuron signals can only flow in one direction, which leads to the processing of samples independently in time and loss of important time-series information for time-series data [5].

1.2.2 GAN, WGAN, CGAN based on LSTM and GRU

This paper proposes a stock prediction model using generative adversarial networks GAN[6] (including traditional GAN, CGAN, and WGAN), firstly comparing LSTM and its variant GRU, selecting GRU as a generator based on the resultant gated recursive units, inputting historical stock prices and generating future stock prices, and convolutional neural network CNN as a discriminator to discriminate between the real stock prices and the generated stock prices. In this paper, the closing price of Apple's stock is selected as the target price, along with some technical features. In addition, the news sentiment index of Apple is also generated using Bert as an additional predictive feature [7]. Finally, this paper compares the result of GAN, CGAN and WGAN models.

2. Models and Methods

2.1. EMD-LSTM

Empirical Mode Decomposition (EMD) is a major breakthrough in Fourier transform-based linear and steady-state spectral analysis, which is based on the time-scale characteristics of the data itself to decompose the signal without any predefined basis functions. This feature makes the EMD method theoretically applicable to the decomposition of any type of signal, and thus has an undeniable advantage in dealing with non-stationary and non-linear data. For arbitrary time series data $x(t)$, the computational flow of the EMD decomposition is as follows.

Step 1 Find the magnitude and location of all local maxima and minima in $x(t)$.

Step 2 The very large and very small values are fitted with the upper and lower envelopes and set $U(t)$ and $L(t)$.

Step 3 Calculate the mean values of the upper and lower envelopes at each time point:

$$m(t) = [U(t) + L(t)]/2 \tag{1}$$

Step 4 let

$$h_1(t) = x(t) - m(t) \tag{2}$$

If $h_1(t)$ satisfies the judgment condition of IMF, then an IMF is obtained. Otherwise, let

$$x(t) = h_1(t) \tag{3}$$

Step 5 The operations of the above steps can be decomposed step by step to the 1,2,...,n $IMF_i(t)$, i=1,…,n and a residual $r(t)$. At this point, the original signal $x(t)$ can be expressed as:

$$x(t) = \sum_{i=1}^{n} IMF_i(t) + r(t) \tag{4}$$

The same modal state function contains multiple periodic features.

2.2. LSTM or GRU based as GAN/WGAN/CGAN generators

2.2.1 GAN principle

GAN is widely used for image generation, but rarely applied for time series prediction. Since there are few studies using GAN for time series prediction, their findings are inconsistent. The purpose of this paper is to use GAN to predict stock prices and to check whether adversarial neural networks can help improve time series prediction. A comparison between LSTM and GRU and the basic GAN, Wasserstein GAN (WGAN) and conditional Conditional GAN (CGAN) models is also included.

In GAN, the loss function is based on KL-JS divergence, and the GAN model will use cross-entropy loss to minimize the difference between the two distributions during training, which is equivalent to minimizing KL-JS divergence. In this paper, the goal of the discriminator is to maximize the probability of the correct label assigned to the sample. The objective mathematical function of the discriminator is defined as

$$\hat{v} = \frac{1}{m}\sum_{i=1}^{m} logD\left(y^i\right) + \sum_{i=1}^{m}\left(1 - logD\left(G\left(x^i\right)\right)\right) \tag{5}$$

Then this project trains the generator to minimize its objective function：

$$\hat{v} = \frac{1}{m}\sum_{i=1}^{m}\left(1 - logD\left(G\left(x^i\right)\right)\right) \tag{6}$$

Where x is the generated input data, y is the target in the real dataset, and G(xi) is the data generated by the generator (fake target). To be computed in GAN through the training process, the loss function of the discriminator is：

$$-\frac{1}{m}\sum_{i=1}^{m} logD\left(y^i\right) - \frac{1}{m}\sum_{i=1}^{m}\left(1 - logD\left(G\left(x^i\right)\right)\right) \tag{7}$$

Then this project trains the generator to minimize its objective function:

$$-\frac{1}{m}\sum_{i=1}^{m}\left(logD\left(G\left(x^i\right)\right)\right) \tag{8}$$

The loss function needs to be minimized during the training process to obtain better results. the GAN principle is shown in Figure 1.

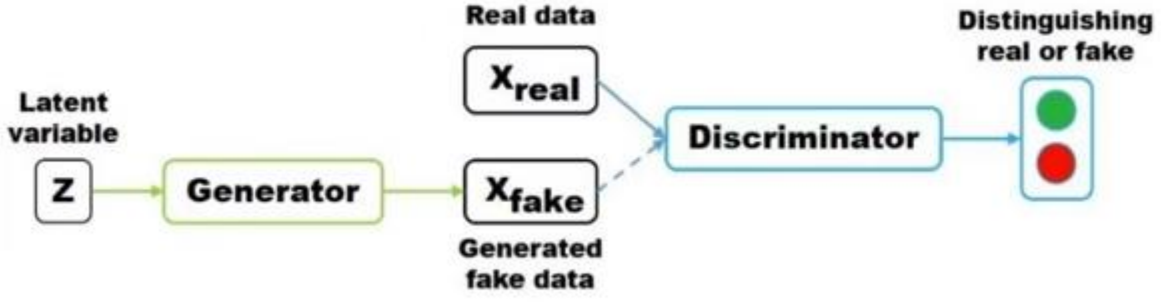

Figure 1. GAN principle structure diagram

2.2.2 WGAN Principle

Sometimes GAN is not powerful enough as the training process is slow and unstable. WGAN can improve the training of GAN by proposing the Wasserstein distance to solve this problem [8]. The Wasserstein distance is the minimum transport when converting a data distribution into a data quality cost. The Wasserstein distance between the true data distribution Pr and the generated data distribution Pg is defined as the maximum lower bound of any transportation plan.

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} E_{(x,y)\sim\gamma}[\| x - y \| \tag{9}$$

where Π (Pr, Pg) denotes the set of all joint distributions between Pr and Pg and Πcontains all possible transportation plans γ. Using the Kantorovich-Rubinstein duality can simplify the calculation as follows：

$$W(P_r, P_g) = \sup_{\|f\|_L \le 1} E_{x\sim P_r}[f(x)] - E_{x\sim P_g}[f(x)] \tag{10}$$

where sup is the smallest upper bound and f is a 1-Lipschitz function that satisfies the Lipschitz constraint

$$|f(x_1) - f(x_2)| \le |x_1 - x_2| \tag{11}$$

2.2.3 CGAN Principle

CGAN is similar to GAN in principle. The difference with the GAN is only a random conditional sequence z appended to the target $y_i$in the real dataset. Hensman et al. [9] proposed conditional adversarial networks, i.e., adding constraints to the GAN, thus effectively avoiding model training collapse. The mathematical objective function of the discriminator and generator is defined as：

$$\hat{v} = \frac{1}{m}\sum_{i=1}^{m} logD(y^i|z) + \sum_{i=1}^{m}\left(1 - logD\left(G(x^i|z)\right)\right) \tag{12}$$

$$\hat{v} = \frac{1}{m}\sum_{i=1}^{m}\left(1 - logD\left(G(x^i|z)\right)\right) \tag{13}$$

The loss function of the discriminator and the generator is：

$$-\frac{1}{m}\sum_{i=1}^{m} logD(y^i|z) - \frac{1}{m}\sum_{i=1}^{m}\left(1 - logD\left(G(x^i|z)\right)\right) \tag{14}$$

$$-\frac{1}{m}\sum_{i=1}^{m}\left(logD\left(G(x^i|z)\right)\right) \tag{15}$$

2.2.4 GAN Model Generator

In the GAN model, GRU is set as a generator to improve stability. The dataset includes historical stock prices and some features such as investor technical analysis indices. This model uses three layers of GRU and then connects two layers of Dense to form a generator.

The choice of LSTM or GRU as the generator can improve the training stability. Among them, GRU is a variant of LSTM. Unlike the traditional RNN, GRU solves the problem of gradient disappearance and explosion; unlike LSTM, GRU has fewer parameters than LSTM because there is no gate, which can save time when the results are similar to LSTM.

2.2.5 GAN model discriminator

The discriminator in the GAN model is a CNN designed to distinguish whether the input data to the discriminator is true or false. CNNs are typically used for image-related tasks. They are very powerful when it comes to extracting features from features, and the ability of CNNs to detect features can be used to extract information about stock price movement patterns. Another reason to use CNNs is that they work well with spatial data, meaning the data points closer to each other are more closely related to each other than to data points distributed everywhere. This holds true for time series data as well. In this example, each feature is continuous.

3. Empirical Analysis

3.1. Data sources and pre-processing

3.1.1 EMD-LSTM data source and pre-processing

Using the closing price of Guanhao Bio from January 1, 2020, to December 31, 2021, Guanhao Bio is a high-tech enterprise in life and health-related fields based on the regenerative medicine industry. By comparing the traditional LSTM model and EMD-LSTM, the results show that the proposed EMD-LSTM model has a better prediction effect. The stock price data were obtained from Netflix. The training data and test data are divided into 9:1.

A financial time series forecasting model based on the EMD-LSTM network is proposed for the characteristics of high noise as well as the nonlinearity of financial time series. The time series data are subjected to multi-step empirical modal decomposition through a certain size of time window, and the decomposed series are denoised and reconstructed, and the reconstructed series are used as the input of the LSTM network to obtain the final prediction results. The decomposition results of the reconstructed sequences are shown in Figure 2.

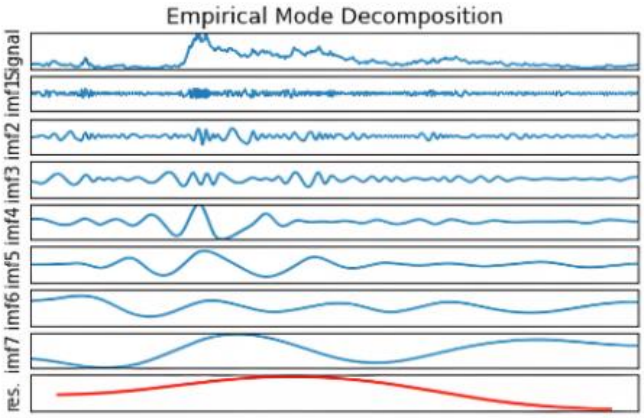

Figure 2. EMD decomposition result graph

3.1.2 GAN data sources and pre-processing

The closing price of Apple's stock was selected as the target price for the empirical analysis of the GAN model, along with the opening, low, high, closing price and volume, NASDAQ, FTSE 100, BSE SENSEX, NYSE, S&P 500, SSE, RUSSELL 2000 HENG SENG Index, Crude Oil, Gold, VIX, US Dollar Index, and other features. The scraped news is then analyzed using the bert model that analyzes the sentiment of financial texts, and a score between -1 and 1 is given to evaluate the news sentiment, and finally, the news sentiment index of Apple is generated as the NEWS feature. Around -1,0.1 means negative, neutral and positive. The target stock price in the model is the closing price of Apple's stock. The training and test data were divided by 7:3. The calculations include the most common technical indicators used by investors (7- and 21-day moving averages, exponential moving averages, momentum, Bollinger bands, MACD, etc.) The technical indicators are shown in Figure 3.

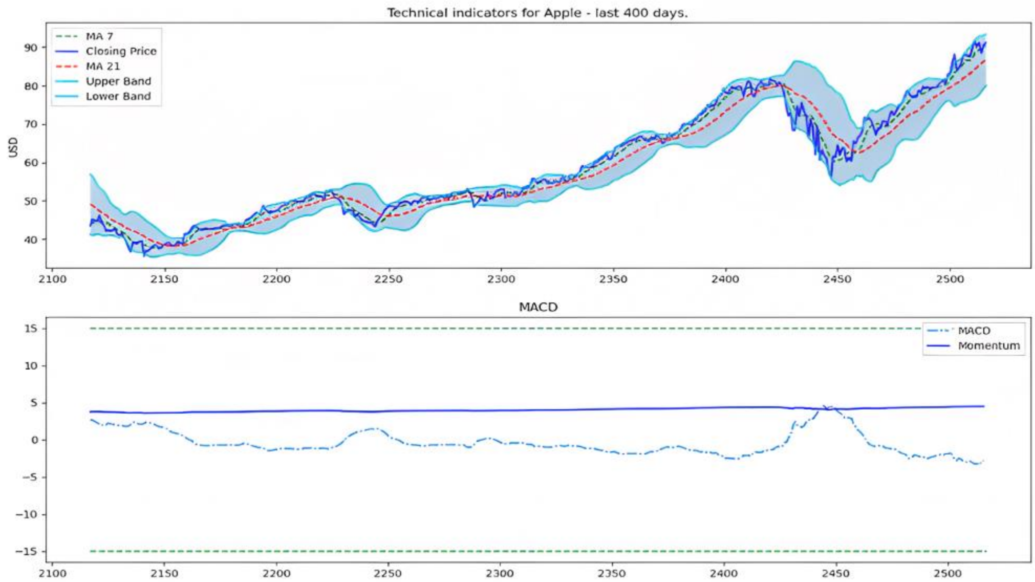

Figure 3. Chart of common technical indicators of stock prices

Fourier transforms with 3, 6 and 9 components are used to extract short and long-term trends based on daily closing prices. The Fourier principle reveals that any continuous time series or signal can be represented as an infinite superposition of sinusoidal signals of different frequencies. After extracting the local and overall trends by Fourier transform, an approximation of the real stock movement is created and used as input to the model with noise reduction to make the model prediction more accurate.

If f(x) is a one-dimensional continuous real function, the Fourier transform can be defined as

$$F(u) = \int_{-\infty}^{+\infty} f(x)\, e^{-j2\pi ux} dx \tag{16}$$

And the discrete Fourier change can be defined as

$$F(u) = \sum_{x=0}^{N-1} f(x) e^{-j2\pi ux}, u = 0,1,2, \dots, N-1 \tag{17}$$

The Fourier transform helps the GRU neural network to predict the trend more accurately. This paper will use these transforms to eliminate much noise and create an approximation of the real stock movement. As can be seen from figure 4, the more Fourier transform components are used, the closer the function value is to the actual value. This indicates that the Fourier transform can effectively extract the stock price trend and thus help the model to make predictions.

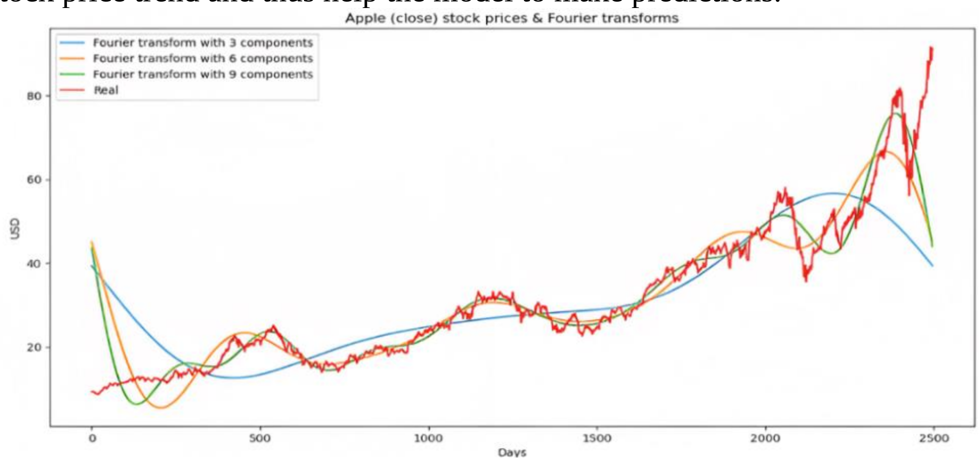

Figure 4. Fourier extracted stock price trend chart

3.2. EMD-LSTM based stock price prediction

The MSE of the conventional LSTM model is 0.10, the MSE of the one-layer LSTM model after EMD decomposition is 0.47, and the MSE of the two-layer LSTM model after EMD decomposition is 0.05.

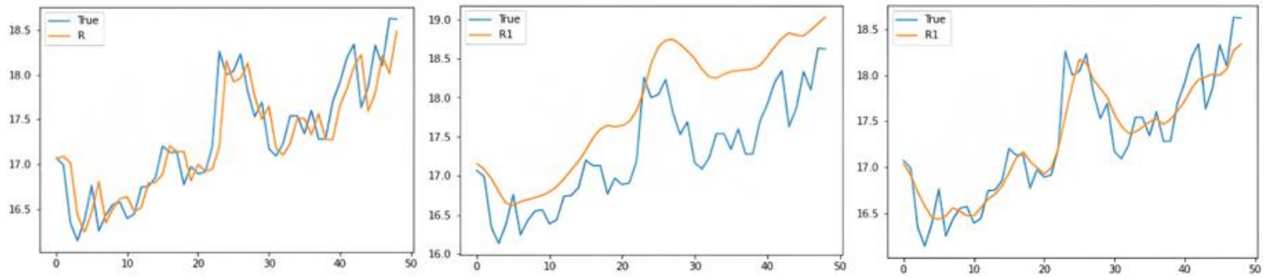

Figure 5. Traditional LSTM, one-layer LSTM after EMD decomposition, two-layer LSTM after EMD decomposition stock price prediction graph

The EMD decomposition was used for modal feature extraction to obtain potential correlation factors and trend information of Guanhao Bio. Then the LSTM model was applied to stock price prediction, which fully utilized the time series information and modal features, and the features extracted by EMD could effectively improve the model's prediction performance. In addition, it is known from the literature [10] that appropriately increasing the number of LSTM network layers can enhance the feature extraction of the input sequence, grasp the data trend, and improve the accuracy of the model prediction, and this paper increases to two-layer LSTM network based on the first two experiments, and the prediction accuracy is greatly improved.

3.3. Comparison of LSTM and GRU prediction results

3.3.1 LSTM and GRU

GRU and LSTM are both improved structures of RNN. The LSTM is a simplification of the GRU, as it appears earlier than the GRU, and has fewer parameters than the LSTM, making it easier to converge. Structurally, GRU has only two gates (update and reset), while LSTM has three gates (forget, input, output). GRU directly passes the hidden state to the next cell, while LSTM wraps the hidden state with a memory cell.

3.3.2 Empirical results

The prediction results of the model with GRU or LSTM as the generator on the training and test sets are shown in Figures 6 and 7.

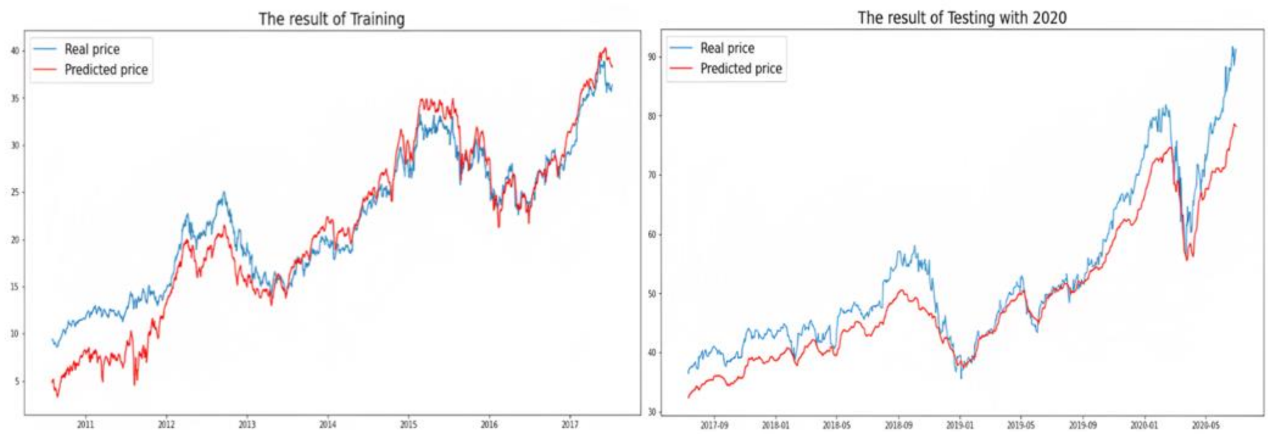

Figure 6. Prediction results of LSTM training set and prediction set

The RMSE is 2.73 for the first 70% training set and 4.67 for the second 30% test set under the LSTM model.

Figure 7. Prediction results of GRU training set and prediction set

The RMSE is 1.71 for the first 70% of the training set and 4.91 for the last 30% of the test set under the GRU model.

Choosing LSTM or GRU as the generator can improve the training stability. Where GRU is an RNN, which is a variant of LSTM. Unlike the traditional RNN, GRU solves the problem of gradient disappearance and explosion; unlike LSTM, GRU has fewer parameters than LSTM due to the absence of a gate, and the training time is shorter in the case of comparable results with LSTM. Therefore, GRU is chosen as the generator of GAN model.

3.4. Analysis of prediction results based on GRU as a GAN/WGAN/CGAN generator

In the model, GRU is used for the generator, which consists of 1024, 512 and 256 neurons in layers 1, 2 and 3, and then two additional dense layers of Dense. In addition, there are two dense layers. The discriminator is composed of one-dimensional convolutional layers with 32, 64 and 128 neurons on layers 1, 2 and 3, and three other dense layers with 220, 220 and 1 neurons on layers 4, 5, and 6, with epochs adjusted accordingly. The parameters of the model are shown in Form 1.

Table 1. The parameters of GAN/WGAN/CGAN

	GAN	WGAN	CGAN
Epochs	100	160	100
Learning Rate	0.00015	0.0001	0.00015
Batch size	128	128	128
Optimizer	Adam's Algorithm	Adam's Algorithm	Adam's Algorithm

Choosing epochs of 100 can get better prediction results in a short time.

3.4.1 Empirical results of choosing various generators to make predictions

In the GAN and CGAN model, the optimizer for our models is the Adam algorithm with a learning rate of 0.00015. The batch size is 128, and then the model on this dataset has been trained for 100 epochs. In the WGAN model, the optimizer for our models is the Adam algorithm, with a learning rate of 0.0001. The batch size is 128, and then the model on this dataset has been trained for 160 epochs. The prediction results of GAN, WGAN and CGAN on the training and test sets are shown in Figures 8, 9 and 10.

（1）GAN empirical results

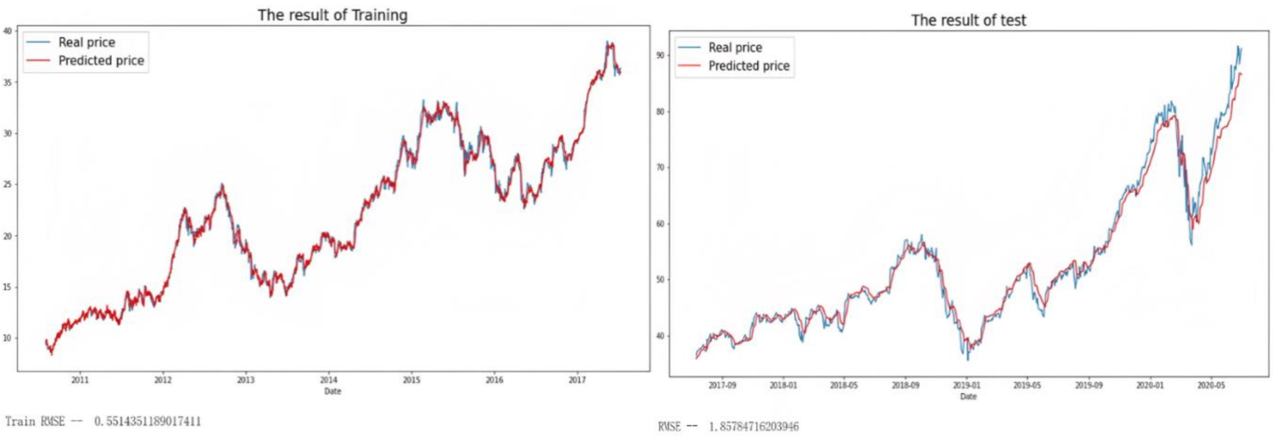

Figure 8. Prediction results of GAN training set and prediction set

RMSE of 0.55 for the first 70% of the training set and 1.86 for the last 30% of the test set under the GAN model.

（2）WGAN empirical results

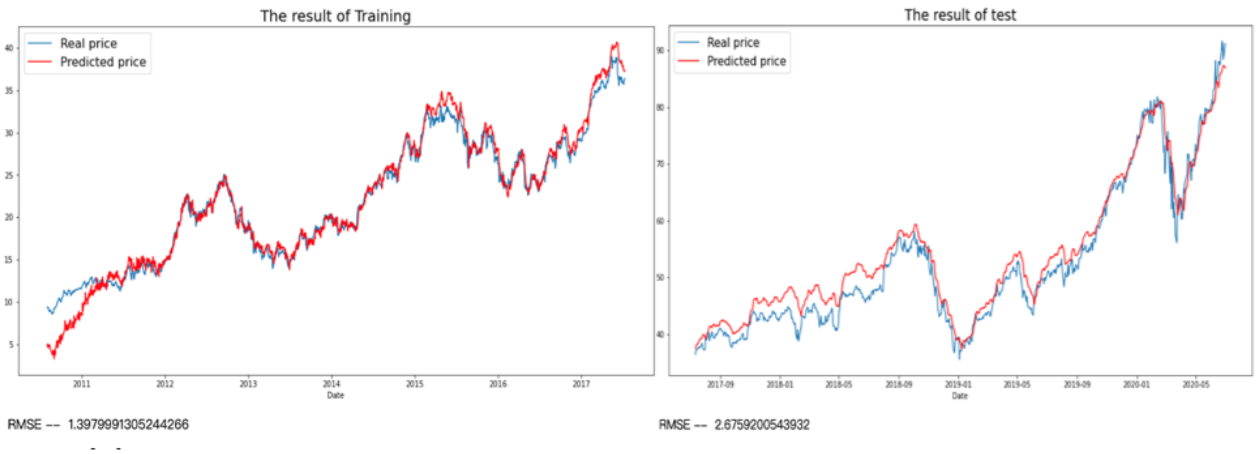

Figure 9. Prediction results of WGAN training set and prediction set

The RMSE is 1.40 for the first 70% training set and 2.68 for the second 30% test set under the WGAN model.

(3）CGAN empirical results

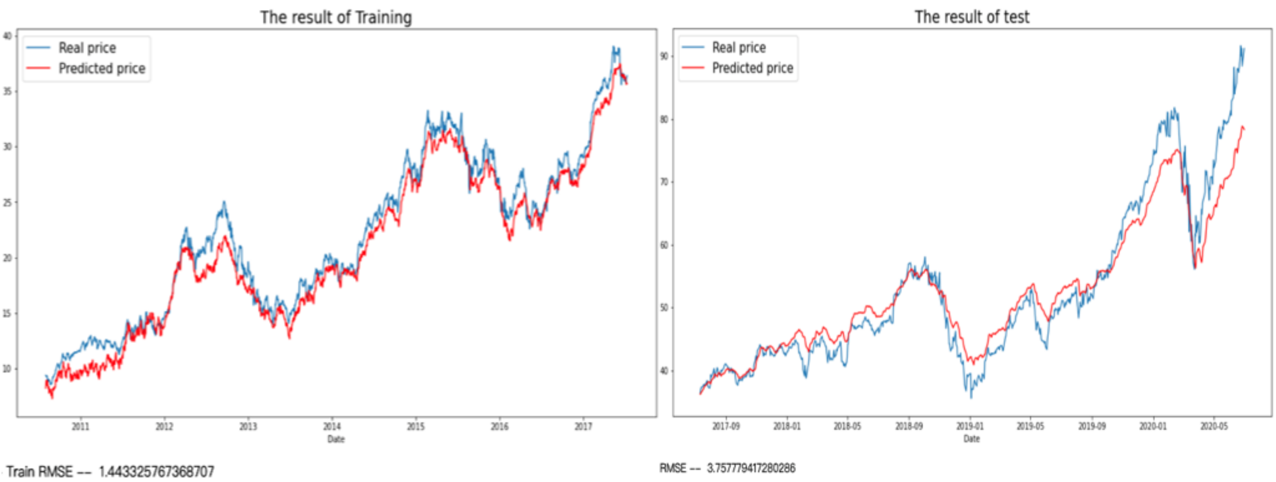

Figure 10. Prediction results of CGAN training set and prediction set

The RMSE of the first 70% training set under the CGAN model is 1.44, and the RMSE of the second 30% test set is 3.76

In summary, the results of GAN are slightly better than WGAN slightly better than CGAN, but the lag of GAN prediction sequence is significantly better than WGAN and CGAN, so WGAN is considered to have a better prediction effect

4. Conclusion

This paper proposes some prediction methods combined with LSTM. In the first half of the paper, we tried EMD-LSTM and compared the prediction results of conventional LSTM, one-layer LSTM after EMD and two-layer LSTM after EMD. In the second half of the article, we propose a stock prediction model using GAN (including GAN, CGAN, and WGAN), which sets GRU as a generator and CNN as a discriminator after comparing the advantages and disadvantages of LSTM and its variant GRU.

According to the experimental result, some conclusions have been summed. EMD can effectively handle financial data, and the best prediction results are obtained from the two-layer LSTM after EMD. The prediction results of LSTM and GRU are similar, but GRU is more time-saving. GAN, CGAN and WGAN can all perform better than traditional models.

References

[1] WANG J,XIA L M. Abnormal behavior detection based on deep-learned features[J].Journal of Hunan University(Natural Sciences),2017,44(10):130-138

[2] Akita R, Yoshihara A, Matsubara T, et al. Deep learning for stock prediction using numerical and textual information[EB/OL].(2016-07-26)[2021-0323].https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7550882.

[3] Peng Yan, Liu Yuhong, Zhang Rongfen. Modeling and analysis of stock price forecast based on LSTM. Computer Engineering and Applications, 2019, 55(11):209-212.

[4] Shi Jiannan, Zou Junzhong, Zhang Jian, et al. Research of stock price prediction based on DMD-LSTM model[J]. Application Research of Computers, 2020,37(3):662.

[5] Long Aoming, Bi Xiuchun, Zhang Shuguang. An arbitrage strategy model for ferrous metal futures based on LSTM neural network[J].Journal of University of Science and Technology of China, 2018,48(2):125-132.

[6] GOODFELLOW I,POUGET-ABADIE J,MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems.New York,NY,USA:ACM,2014:2672-2680.

[7] Araci D T. FinBERT:financial sentiment analysis with pre-trained language models[EB/OL].(2019-06-25)[2021-03-23].https://arxiv.org/pdf/1908.10063.pdf.

[8] Xu Feifei, Hu Yue, Wang Zhaobing. Research on stock price forecasting based on FWGAN model[J].Journal of Zhejiang University of Science and Technology,2022,(03):207-215.

[9] Hensman P, Aizawa K. cGAN-based manga colorization using a single training image[EB/OL].(2017-05-30)[2021-03-23].https://arxiv.org/pdf/1706.06918.pdf.

[10] Chen K, Zhou Y, Dai F.A LSTM- based method for stock returns prediction:a case study of China stock market[C]//IEEE International Conference on Big Data,2015:2823-2824.