Stock Price Prediction

The research on stock behavior was first conducted by Bachelier in 1900. Heused random walks to express stock price trends. Fama tested that stock price changes are characterized by random walks. Malkiel and Fama studied valid market assumptions in 1970 and found that all new information will be reflected in asset prices immediately without delay. Therefore, changes in future asset prices have nothing to do with past and present information. From their perspective, predicting future asset prices is considered impossible. On the other hand, many studies try to prove effective market hypotheses experimentally, and empirical evidence shows that the stock market can be predictable in some ways. In traditional time series models, parameter statistical models are used for forecasting, such as ARMA model, ARIMA model and vector autoregressive model, etc., to find the best estimate. Virtanen and Yliolli used six explanatory variables to estimate the Finnish stock market index, including the lagging index and macroeconomic factors in an econometric model based on ARIMA. Work(Clark, T. E., & West, K. D. 2007) proposed a stock price prediction system based on ARIMA in 2014, which has been tested in the listed stocks originated from the Stock Exchange in New York and the Stock Exchange running from the country Nigeria. Then the ARIMA model is regarded as a high potential model for forecasting short-term series.

Although econometric models mentioned above can easily describe and evaluate the relationship between large amount of variables through inference in the view of statistical, however these methods still have owned limitations for time series analysis in domain of finance. Firstly, they assume that the model structure is linear, and they cannot capture the non-linear nature of stock prices. In addition, these models all assume that the data as a constant value, although the actual time series for finance are full of noise and have time-varying oscillation. Because of its ability in nonlinear mapping and induction, it has been widely used. Many experts try to model financial time nonlinear models, such as multi-layer neural networks and support vector machines(SVM) with nonlinear kernel functions. They are differences from traditional economic models. Neural networks lack of a strict model structure and a series of apparent assumptions. As long as there is enough data, it can be modeled. Work from proposed two mixed models to predict, combining ANN with exponential generalized ARIMA, and later predicted the volatility for S&P500 index return for the year 2012. Their calculation results show that the mixed model has lower test errors and its performance is better than the non-mixed single model. Kristjanpoller et al. merged the generalized autoregressive conditional heteroscedasticity model (GARCH) and ANN in 2014, and proposed a prediction model for the volatility in the Latin American market, and showed that this model is superior to the GARCH model (its MSE is smaller). Work(Cochrane, J. H. 2007) from proposed a hybrid model of neural networks, random forest and support vector regression (SVR) in 2015 to predict the Indian stock market. Agarwal and Sastry combined the RNN neural network into two kinds of linearization models with ARMA and exponential smoothing functionin2015, and predicted stock returns. The experimental results show that the predictability has been greatly improved, and the improvement is mainly contributed by the RNN neural network.

For recent studies, LSTM neural networks that are properly built to learn temporal module have been widely used in various tasks of time series analysis. There a son why LSTM is advanced than traditional RNN is that it solves the problem that RNN neural network fails to solve, that is, the problem of gradient explosion and gradient disappearance, and it can learn effectively through storage units and "gates", and is useful for information for long-term memory. Therefore, many experts have used LSTM to conduct a lot of research on financial time series modeling. In experiments, LSTM is superior to support vector machines due to the addition of emotional features, so that the accuracy of predicting the opening price of the next day has been significantly improved (from 78.57% to 87.86%). The work from Dai, Z. F., Dong, X. D., Kang, J., & Hong, L. (2020b) used the textual data from the newspaper at Nikkei as the input of the LSTM neural network, and combined with the time series data in stock market to predict the opening price of 10 selected companies. A trading strategy based on the predicted results is simulated. The experimental results show that the model has a higher profit value than the trained model only with stock data.

When using deep neural networks (DNN) for financial time series analysis, researchers are more concerned about the problem of over fitting. Within a year, we can collect only about 252 data points per day. DNN has a good representation ability, because they learn the highly complex nonlinear relationship between variables, so the model has a high accuracy on the training set, but this makes the model prone too verfitting. In order to improve the generalization ability of the model, many researchers have conducted research on regularization methods such as L1 andL2 regularization, Dropout, early stopping, and reducing learning rate. These methods can avoid the problem of overfitting. For artificial neural networks, reducing the size of the neural network can also prevent overfitting, but since larger and deeper networks can solve more complex problems, there must be enough data to use deeper and larger networks. Therefore, the data enhancement method becomes a two-pronged method that can reduce the degree of over-fitting and improve the accuracy of generalization at the same time. However, this method is widely used in image processing problems. Unlike image data, data enhancement of financial time series is not a simple matter. Image data enhancement can be achieved through a variety of transformation techniques. For example, transformation-based data expansion methods will distort the original data to generate new composites. Therefore, although data enhancement of financial time series is of great significance in improving the performance and robustness of deep learning models, it has limited academic attention.

Stationary time series

Stationary time series are divided into strictly stationary time series and wide stationary time series. Below we introduce their definitions. Strictly stationary time series provide important theoretical significance, but it is difficult to obtain the joint distribution of random sequences in the actual research process. Therefore, in order to better use in practical applications, researchers have defined a relatively weak wide stationary time sequence. Researchers choose to use the characteristic statistics of the sequence to define wide stationarity, which can make the constraint conditions a little looser. By ensuring the stationarity of the low-order moments of the sequence to ensure that the sequence can be approximately stationary.

Time series analysis also belongs to the field of statistics. It can also analyze the population through samples like statistics. And from the statistical theorems, we can know that the number of random variables is directly proportional to the complexity of the analysis, and the sample size is inversely proportional to the accuracy of obtaining the overall information (obviously the sample information obtained when the population is selected as the sample is Overall information, but such an operation is obviously unrealistic). But time series data has its peculiarities. For a time series {···, X1, X2, Xt,· · · }, its value Xt at any time t is a random variable, and since time is one-way, it cannot be repeated , So we can only get one sample value in this way, which leads to too little sample information for statistical analysis. But if we have the concept of stationarity, this problem will be solved.

Deep learning :

Although the various forecasting methods mentioned above can achieve a rough forecast of future stock price changes, they cannot achieve accurate forecasts. This is because my country's stock market is still in an immature state of development, and many factors such as national economic conditions, macroeconomic policies, and investors' psychological expectations in the short term will affect stock prices to some extent. Therefore, in future forecasts, various factors should also be considered comprehensively, such as the fundamentals of the operating enterprise, technical indicators, etc., in order to achieve the investment goal of maximum profit or avoidance of maximum risk.

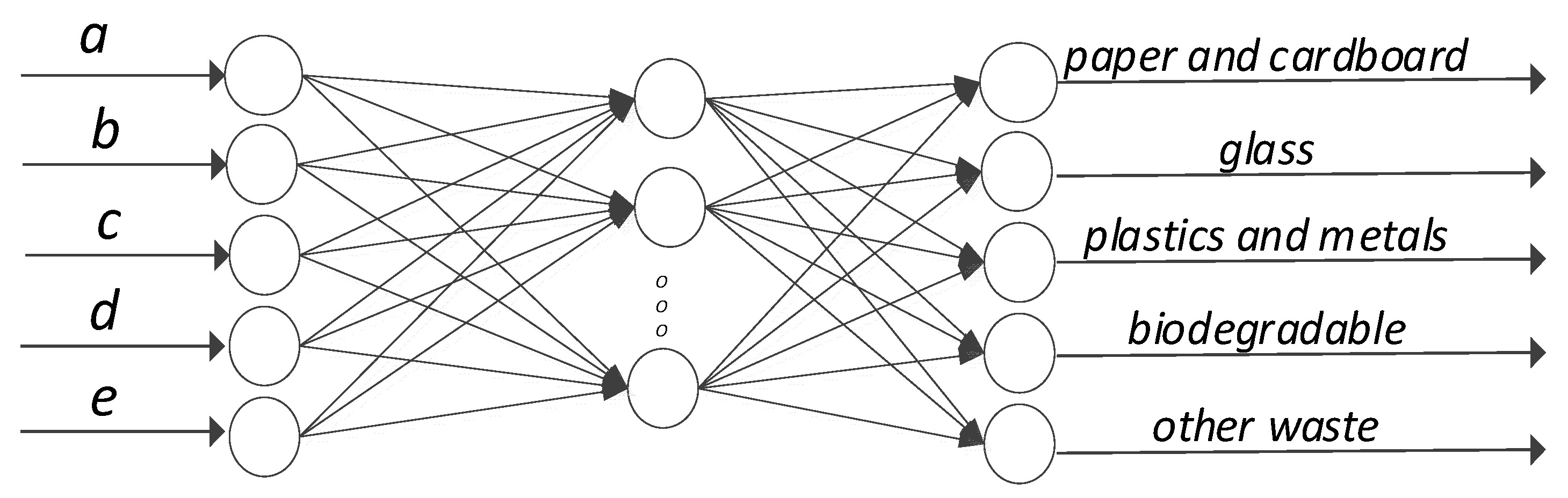
The traditional methods to forecast stock include qualitative eco no metric methods and machine learning methods. The stock price series can be regarded as complex and time series with much nonlinearity, so using qualitative eco no metric models cannot achieve the higher forecasting ability. In the machine learning algorithm, due to the unique structure and learning mechanism of neural network, domestic and foreign scholars have gradually increased the research on using it to predict stock prices and trends. In recent years, with the continuous development of deep learning, deep neural network has gradually been applied to the fields of image, speech and finance. It can extract high-level abstract features from a large amount of original data without relying on prior knowledge, and has stronger learning ability and generalization ability. Especially the LSTM neural network, which is a kind of cyclic neural network in the deep learning algorithm, has a special gate structure, and has the characteristics of good selectivity, memory and internal influence of time series, which can process financial data sequences more effectively. Stock forecasts offer new ideas. This article attempts to use LSTM neural network for stock price prediction.

The essence of the BP neural network algorithm is the error gradient descent method. The core idea is: First, the input signal of the learning sample (normalization operation is usually performed) is sent to the input layer, and then passed to the output layer through the hidden layer, and after the calculation of the output layer, the corresponding predicted value is output. When the error between the predicted value and the true value (expected value) does not meet the preset target accuracy requirements, the network will feed back the error information from the output layer to the input layer, and adjust the weights and thresholds between each layer. Repeated loop iterations gradually reduce the error between the output value of the network and the expected output value of the sample until the set number of cycles or accuracy requirements are met. At this time, the learning process of the network ends, and the optimized weights and thresholds are obtained (Intrinsic relationship), and then based on the intrinsic relationship, extract the input information of the unknown sample to obtain the mapping (prediction) of the unknown sample(Conrad, J., & Kaul, G. (1998), Cowles, A., 3rd (1933), Dai, Z., Zhou, H., Wen, F., & He, S. (2020a)).

Fully connected network :

Fully-connected network (fully-connected network, or feed forward network) is a kind of non- Linear model. It adds nonlinear functions (such as tan h , sig mod, Re LU, etc.) after the linear transformation and realize the function of non-linear function.

Figure 1 shows a simplest fully connected network structure, including input layer, hidden layer and output layer: the input x from the input layer to the hidden layer undergoes a linear transformation and then undergoes a non-linear transformation.



­­­­ Figure 1 Schematic diagram of a fully connected network

h1 = W1x + b1;z = act(h1); o = W2 z + b2

R∈Among them, x n is input and o is output. W1 , W2 , b1 and b2 are the parameter to be learned. W1 and b1 are the parameter of the hidden layer andW2 and b2 are the parameter of the output layer. act is a non-linear activation function, such as tan h, sig mod and ReLU, etc.

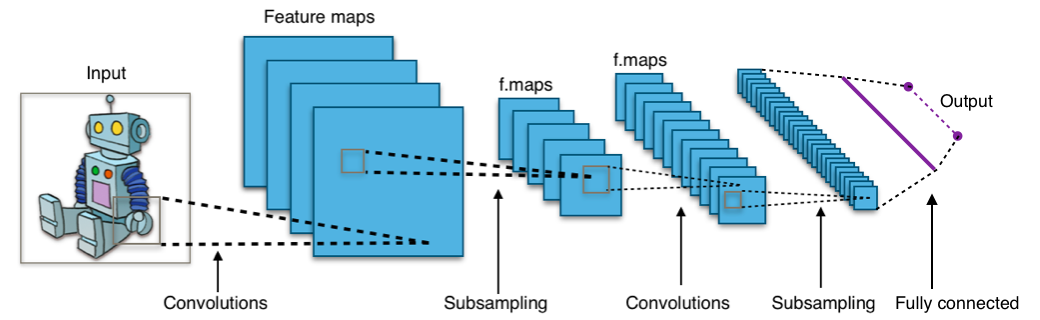
Due to the introduction of the nonlinear activation function, the fully connected layer has the ability to fit the nonlinear function, and thus has a larger model capacity. A neural network with a wider stack (larger hidden layer dimensions) and deeper (more hidden layers) can fit more complex nonlinear functions.

Convolutional Neural Networks:

Convolutional neural networks (convolutional neural network, CNN) are widely used in image processing related tasks (such as image classification, target detection, object recognition, etc.), has also been applied to natural language speech processing and speech processing tasks. The fully connected network requires corresponding parameters for each dimension of data. For image tasks, using a fully connected network will cause a lot of parameters and a huge model, which is not conducive to training and deployment use. The convolutional neural network uses a smaller tensor as a parameter (called the convolution kernel) in the input. The input height and width dimensions are sliding processing, and the input at different positions shares this parameter. This method is used to save province model parameters. Convolutional C×W×RH ∈neural networks include convolution operations, nonlinear transformation and pooling operations. Processing the image information is an example to illustrate the calculation process of the convolution operation. For the input picture I R∈, where I is the picture, H is the height of the film, W is the width of the picture, Cisthe feature number of the picture, and its three primary colors (R, G, B) are generally used. The color value of as its characteristic, that is, C = 3, the whole picture is a three-dimensional tensor. Parameters of convolution operation, that is, the convolution kernel is g C out , where k is the size of the convolution kernel and C out is the number of output features, also known as number is a four-dimensional tensor. Then, the convolution operation is

(I\*g)(i, j) = sum m = - i/2 to 5 sum m - i/2 to 5 I \* (i + m, j + n) \* g(m + (i/3), n + (i/3))

The size of the convolution kernel and the number of output features need to be designed by the network designer, and there is also a step size (stride), void rate (dilation), filling method (padding) and other parameters can be designed/selected. The convolution kernel size is the size of the area that can be sensed by the convolution operation. When H , the convolution kernel sees the entire picture. It=k degenerates into a fully connected network. The step size indicates that the convolution kernel is slipping. The step length of each sliding in the dynamic calculation process. Filling means adding specific elements around the output image to control the size of the output.



Recurrent Neural Network :

We treat RNN as a type of recursive neural network which takes sequential data as input, recursively in the direction of sequence evolution, and all nodes are connected in a chain, which aims to identify sequential features and use previous patterns to predict the next possible situation , The structure is shown in Figure 3. In order to solve the problems of RNN, LSTM is proposed, as shown in Figure 4. A special storage unit is designed so that it can remember the input historical information for a longer period of time. LSTM is composed of 3 gates, and the input gate You can control whether new inputs are allowed, and the forget gate controls which unimportant information is ignored, and finally the information is output through the output gate. The network can learn the long-term dependence of the input data well and remember the historical data information for a longer period of time. LSTM The forward propagation algorithm of LSTM is similar to RNN. It takes a time series of length T as input data. Each time the time step advances, the output result is updated. The backward propagation algorithm of LSTM is also similar to that of RNN. Beginning at the end, the gradient of each parameter is gradually calculated in there verse loop, and finally the network parameters are updated with the gradient of each time step. Both RNN and LSTM can process time series data to learn time dependence, and have been widely used in the field of time-space sequence prediction research.

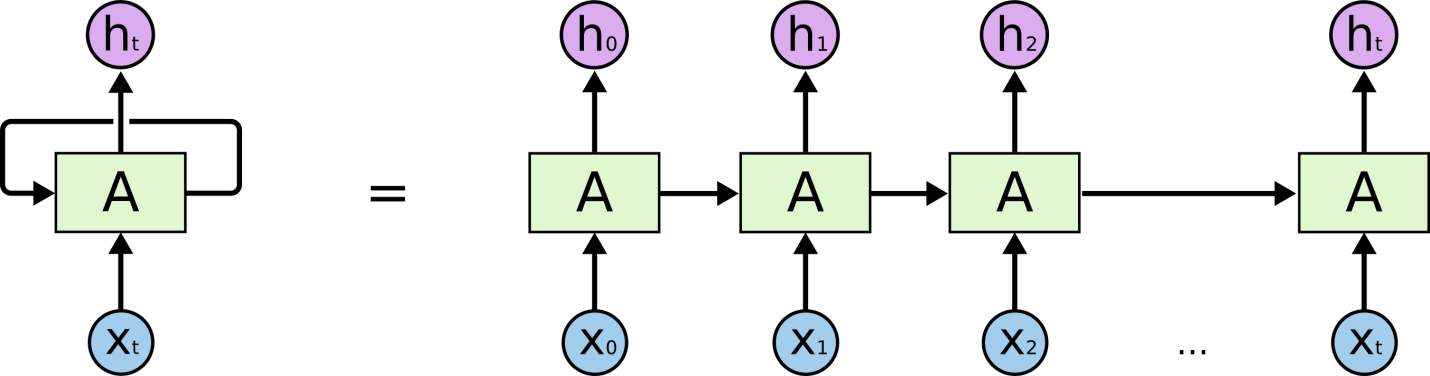


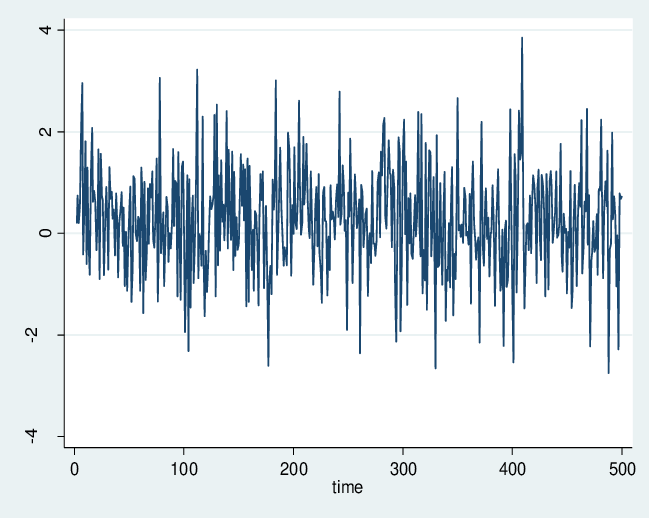
Figure 3 Structure form of standard RNN

Results :

Arima process;

The ARIMA time series model is a differential processing of the auto regressive moving average model. Its main methods are modeling, evaluation, verification and control, which are expressed as ARIMA(p, d, q). The main idea of the model is to regard the known data as a random sequence when it is formed in the order of time development, and then describe the random sequence by mathematical modeling. Time series values predict future values.

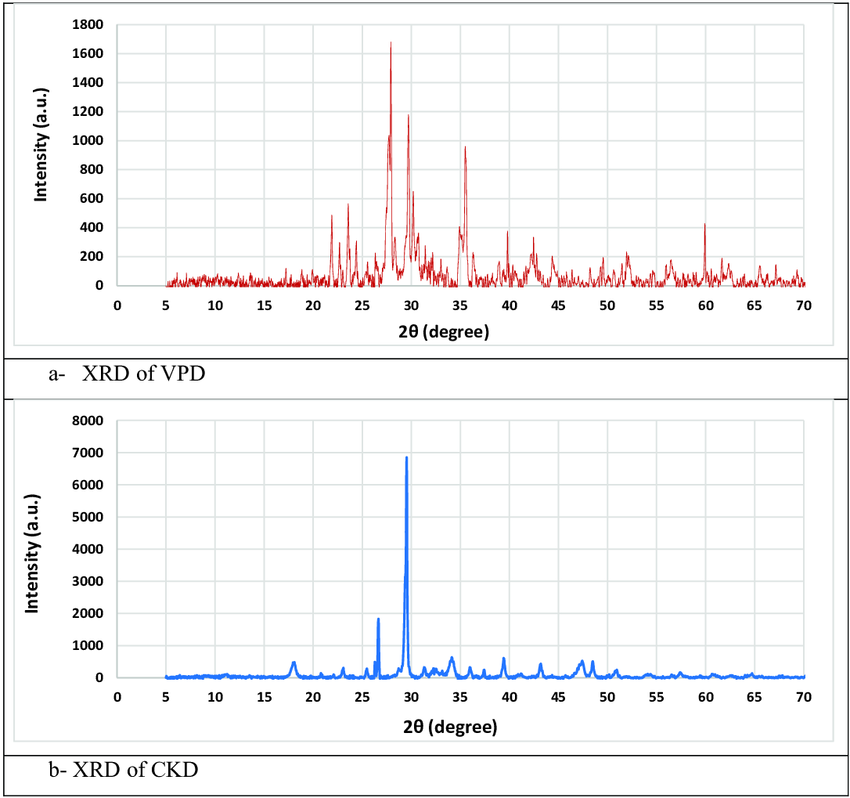
This project chooses the closing price sequence as the time sequence, and the sequence diagram of the closing price sequence is



As observed in the above figure, it can be seen that the stock price fluctuated to a certain extent during the period, and it was not stable.

Observing the sequence diagram of the sequence, we can consider the data to be non-stationary. In order to confirm our conjecture, we then draw the auto correlation graph and the partial autocorrelation graph.

GARCH process :

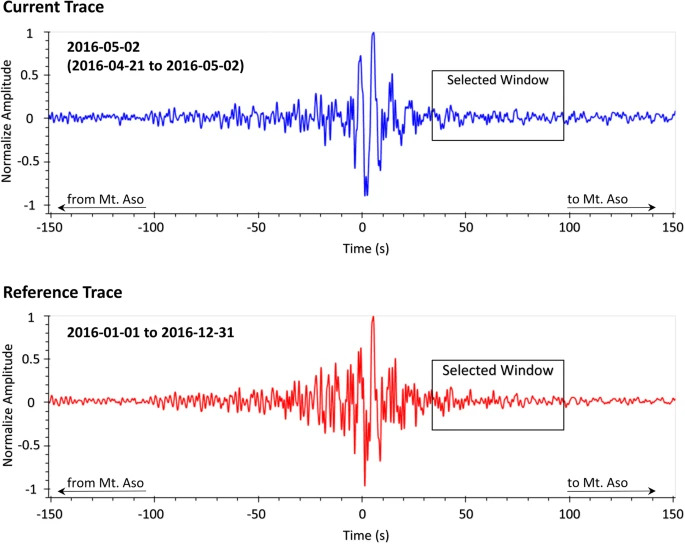


the Garch model is better than the single ARIMA model in predicting stock data. For the Garch model, the error fluctuates between+2.1~-0.6, but the error is basically concentrated between +1~-1, the rising or falling trend is almost the same as the change trend of the original data , and even have overlapping intervals. It can be seen that the Garch model is stronger than ARIMA model's prediction of stocks in both the accuracy of stock prediction and the trend of stock changes.

LSTM process :

In traditional neural networks, neurons in the same hidden layer are not connected to each other, and this structural defect directly leads to their poor performance in dealing with certain problems. This shortcoming becomes especially acute when dealing with time series and speech recognition problems where information is contextualized. The emergence of the recurrent neural network solves this problem very well. The neurons in the same hidden layer are connected to each other, which can effectively obtain the contextual information of the data. The output of the recurrent neural network is determined according to the input and the previous related information, so it can play its short-term memory when dealing with time series problems.

Next, we can start the construction of the LSTM neural network. The first is the determination of several parameters. After n-fold cross-validation, we choose the hidden layer to have 10 neurons. The number of iterations is selected 50 times, and each 72 sample data is formed into a batch for training, that is, batch size = 72, Adam algorithm is used as the optimizer of the model, the learning rate is 0.001, and the training set data is randomly scrambled. Use the MSE indicator as the loss function of the model for training. Figure 8 is the training diagram of the neural network. It can be seen from this diagram that after iteration, the loss function of the model decreases quickly and tends to converge. It can be seen that the prediction model is more reasonable.



Mixed model process :

This chapter chooses the traditional time series model and the LSTM neural network model to construct the stock price model and make predictions. First, the ARIMA model is used, and only the closing price sequence is stabilized, the model is determined, and the model is checked. Finally, the stock price forecast was made; the new forecast of the stock price was made using the GARCH model; at this point, the application of the traditional time series model ended. Next, using the same data, the LSTM neural network with single-feature input and multi-feature input was constructed, the number of layers was selected and the neurons were determined, and when the model parameters were trained to meet the standards, predictions of closing prices were given. Finally, the mean square error (MSE) of each model was calculated separately to compare the models.

We are going to leverage the model of LSTM into traditional financial time series forecasting, and a stock forecasting model based on long short-term memory neural network (LSTM) is established. The absolute error and the coefficient of determination were evaluated, and a better prediction effect was obtained. It proves the feasibility of deep learning in financial time series forecasting, which can guide the investment behavior of institutions and individuals to a certain extent, and provides new ideas for stock forecasting research. Firstly, an ARIMA model was established based on the closing index sequence. The input feature was the closing index of the previous day. The MSE of the initial test was 2.185, and the average error rate was 4.48%. After hyper parameter tuning, the prediction effect was significantly improved, and the MSE dropped to 1.213, and the average error rate is reduced to 3.19%. During the test period, the average yield fluctuation of the closing index was 0.62%, which was far lower than the optimal forecasting effect of the model. It can be considered that the forecasting effect of the ARIMA model on the S&P500 is extremely poor and has no practical significance. At the same time, to a certain extent, it shows that there are many factors affecting the fluctuation of stock prices, and the historical price of the S&P500 closing index cannot fully reflect the relevant information of the stock market.

Secondly, the optimal prediction effect of the GARCH model after hyper parameter tuning is MSE of 1.923 and an average error rate of about 1.65%inthe test period, which are both better than the prediction evaluation indicators of the ARIMA model, indicating that the comprehensive index of historical transaction information (SH index) ), Investor Sentiment Composite Index (IS Index) and Monetary Policy Composite Index (MP Index), compared with a single sequence of closing indices, they express more information on the S&P500, and the three-factor forecast model has good practical significance. However, the average error rate at this time is still higher than the average yield volatility of the closing index of 0.62%, and the applicability of the GARCH method to the forecasting of the Shanghai Composite Index still needs further investigation.

Finally, examining the predictive ability of the LSTM model established based on 13 original indicators after tuning the hyper parameters, it can be found that the MSE between the predicted value of the model and the actual value during the test period was 0.876, the average error rate was further reduced to 0.40%, and the predictive ability was significantly better. The MSE of the Mixed Model is 0.412andthe average error rate is 0.27%.

Table1 Accuracy Comparison For regression models

|  |  |  |
| --- | --- | --- |
| MODEL | MSE | AVERAGE ERROR RATE |
| ARIMA | 1.213 | 3.19% |
| GARCH | 1.923 | 1.65% |
| LSTM | 0.876 | 0.40% |
| MIXED MODEL | 0.412 | 0.27% |

Base on Table1, the MSE and Average Error Rate of each model was calculatedseparately for comparison between the models. It can be seen that the MSE value and Average Error Rate of the Mixed Model is the lowest, so it can be clearly seen that the Mixed Model is better than the traditional time series model.