Team Control Number

202010231013

Problem Chosen

B

2020

ShuWei Cup

Summary Sheet

(Your team's summary should be included as the first page of your electronic submission.)

Summary

The changes in the stock price of listed companies can directly reflect the operating conditions of the listing formula and the recognition of the market. However, since the stock market is a very typical nonlinear complex system (stock prices have both trend and random factors), stock price forecasts have always It is a difficulty. Starting from the chaos theory, this study intends to establish a mathematical model that can reflect the laws of the stock market to a certain extent, so as to provide answers to such problems or provide reliable suggestions.

For question one, this research first preprocesses the data given in the title to obtain daily, weekly and monthly data, and calculates a number of technical indicators based on basic indicators, which are used as the basis for judging stock trends;

For question two, this research comprehensively considers basic indicators and technical indicators, establishes a stock price and randomness prediction model based on LSTM, and uses the data of the past 20 days to evaluate the accuracy of the model, and the results show that the model is reliable;

For question three, this research introduces key turning point indicators on the basis of question two, establishes a stock price reversal judgment model based on LSTM, and uses the data of the past 20 days to evaluate the accuracy of the model. The results show that the model is reliable.

Key word: Stock market, Technical indicators, Key turning points, LSTM

Content

1. Introduction

1.1 Background

The change of stock price of listed companies can directly reflect the operation status of listed companies and the recognition degree of the market. The modeling and forecasting of stock price is always a difficult problem. The most important factor is that the stock price has both trend and random factors. Therefore, the stock market is a very typical nonlinear complex system. In the aspect of solving nonlinear complex system modeling, practice has proved that chaos theory is an effective theory, and has achieved certain theoretical and application effects in power, communication and other fields. As the stock market is also a nonlinear and complex system, can we use the idea and theory of chaos theory to establish the stock price model? Please use the mathematical modeling method to solve the following problems with the transaction data of three stocks provided:

2. Problem analysis

2.1 Analysis of question one

This study first preprocessed the data given by the subject to obtain weekly and monthly data, and then calculated based on the basic indicators given by the subject (opening price, highest price, lowest price, closing price, trading volume, turnover rate) Obtained a series of technical indicators, and analyzed the daily, weekly and monthly trends of the three stocks;

2.2 Analysis of question two

In this study, the original data of each indicator was preprocessed to filter invalid and non-compliant data. And through the construction of a multi-category feature system for stock trend prediction, basic indicators and technical indicators are calculated to obtain trend indicators and strength indicators based on price or trading volume, and finally a stock price trend model based on LSTM is established;

2.3 Analysis of question three

This study builds a stock price reversal judgment model based on the results of question two. By introducing important turning point indicators, a price reversal judgment model that can reflect the cycle of stock price changes is established. And use historical data to predict the cyclical changes of the selected stocks in the past 20 days to verify the reliability of the model.

3. Symbol and Assumptions

3.1Symbol Description

Symbol	Explanation
EMA n	Exponential moving average on day n
MACD	Moving Average Convergence and Divergence
DIF	Deviation
DEA	***DIF*** 's 9-day ***EMA***
KDJ	Stochastic indicator

3.2 Fundamental assumptions

Ⅰ: Assuming that when the time series is long enough, the stock price data presents a normal distribution;

Ⅱ: Assume that the stock market for the three types of stocks given in the question is stable and healthy;

Ⅲ: It is assumed that there are no major external factors in the time interval of the data given in the question.

4. Question one

4.1 Selection of technical indicators

The data given in the title includes indicator data such as opening price, highest price, lowest price, closing price, trading volume, and turnover rate. Such indicators are generally called basic indicators in the stock market. Since the stock information directly reflected by the basic indicators is relatively simple (most of the information that can reflect the development trend of the stock market is hidden), this study determined several technical indicators based on the basic indicators by consulting the literature, and gave three types of topics based on this Analyze the trend of stocks day by day, week by week, and month by month.

The selected technical indicators include: ***EMA5***, ***EMA10***, ***EMA20***, ***EMA30***, ***EMA60***, ***MACD*** fast line, ***MACD*** slow line, ***MACD*** column, ***Roc***, ***MAROC***, ***K***, ***D***, ***J*** stochastic indicators, a total of 13 technical indicators.

4.2 Calculation of technical indicators

4.2.1 Exponential Moving Average (EMA)

EMA is the index average index, which is a trend index. Its construction principle is: the arithmetic average of the closing price is used to judge the change trend of the price trend in the future. Compared with the ***MACD*** indicator and ***DMA*** indicator, the ***EMA*** indicator focuses on the weight of the price of the day (current) in its calculation formula, which determines that it is a type of trend analysis indicator, which overcomes the lag of the ***MACD*** indicator for price trends in use Sexual defects, and at the same time, to a certain extent, eliminate the signal advancement generated by the ***DMA*** indicator for price movements at certain times, and it is a very effective analysis indicator. Its definition is:

$$EMA_N(x_n) = \frac{2}{N+1}\sum_{k=0}^{\infty}\left(\frac{N-1}{N+1}\right)^k x_{n-k}$$

Among them, define $EMA_N(x_1) = x_1$, the recurrence formula is:

$$EMA_N(x_n) = \frac{2x_n + (N-1)EMA_N(x_{n-1})}{N+1}$$

4.2.2 Moving Average Convergence and Divergence (MACD)

MACD is called the moving average of similarity and difference, and the meaning of ***MACD*** is basically the same as that of double moving average, that is, the dispersion and aggregation of fast and slow moving averages characterize the current long-short status and the possible development trend of stock prices. The change of ***MACD*** represents the change of market trend, and the ***MACD*** of different K-line levels represents the buying and selling trend in the current level cycle. The calculation steps are as follows:

Ⅰ.Calculate the 12-day moving average (EMA12) and the 26-day moving average (EMA26)

$$EMA12 = EMA12' \times \frac{11}{13} + C_{12} \times \frac{2}{13}$$

$$EMA26 = EMA26' \times \frac{25}{27} + C_{26} \times \frac{2}{27}$$

Among them, $EMA12'$ and $EMA26'$ represent the $EMA12$ and $EMA26$ of the previous day, and C_{12} and C_{26} represent the closing prices on the 12th and 26th.

Ⅱ. Calculate the deviation value (***DIF***)

$$DIF = EMA12 - EMA26$$

Ⅲ. Calculate the 9-day ***EMA*** (***DEA***) of ***DIF***

$$DEA = DEA' \times \frac{8}{10} - DIF \times \frac{2}{10}$$

Among them, $\boldsymbol{DEA'}$ represents the $\boldsymbol{DEA}$ of the previous day.

In addition, use $(DIF - DEA) \times 2$ to get the $\boldsymbol{MACD}$ histogram.

4.2.3 *KDJ* Stochastic Index

The ***KDJ*** stochastic indicator is a fairly novel and practical technical analysis indicator. It was first used in the analysis of the futures market, and then widely used in the short-term trend analysis of the stock market. It is the most commonly used technical analysis tool in the futures and stock markets. The stochastic indicator ***KDJ*** is generally a statistical system used for stock analysis. According to statistical principles, through the highest price, lowest price that occurred in a specific period, the closing price of the last calculation period and the proportional relationship between these three, To calculate the immature random value RSV of the last calculation cycle, and then calculate the ***K*** value, ***D*** value and J value according to the smooth moving average method, and draw a graph to study the stock price trend. The calculation steps are as follows:

Ⅰ. Calculate immature stochastic index value (***RSV***)

$$RSV_n = \frac{C_n - L_n}{H_n - L_n} \times 100$$

Among them, C_n is the closing price on the n-th day, L_n is the lowest price in n days, and H_n is the highest price in n days.

Ⅱ. Calculate ***K*** value and ***D*** value

$$K = K' \times \frac{2}{3} + RSV \times \frac{1}{3}$$

$$D = D' \times \frac{2}{3} + K \times \frac{1}{3}$$

Among them, K' and D' are the ***K*** value and ***D*** value of the previous day, respectively. If there is no ***K*** value and ***D*** value of the previous day, 50 can be used instead.

Ⅱ. Calculate ***J*** value

$$J = K \times 3 - 2 \times D$$

4.3 Data processing

This research uses Excel tools to process the relevant data of the three stocks given in the title, and obtains the daily, weekly and monthly data needed for the analysis.

4.4 Stock analysis (taking 000400 as an example)

In this study, Matlab software was used to visualize the technical indicators of the stock numbered 000400. The results of daily changes are shown in Figure 1, Figure 3 and Figure 5.

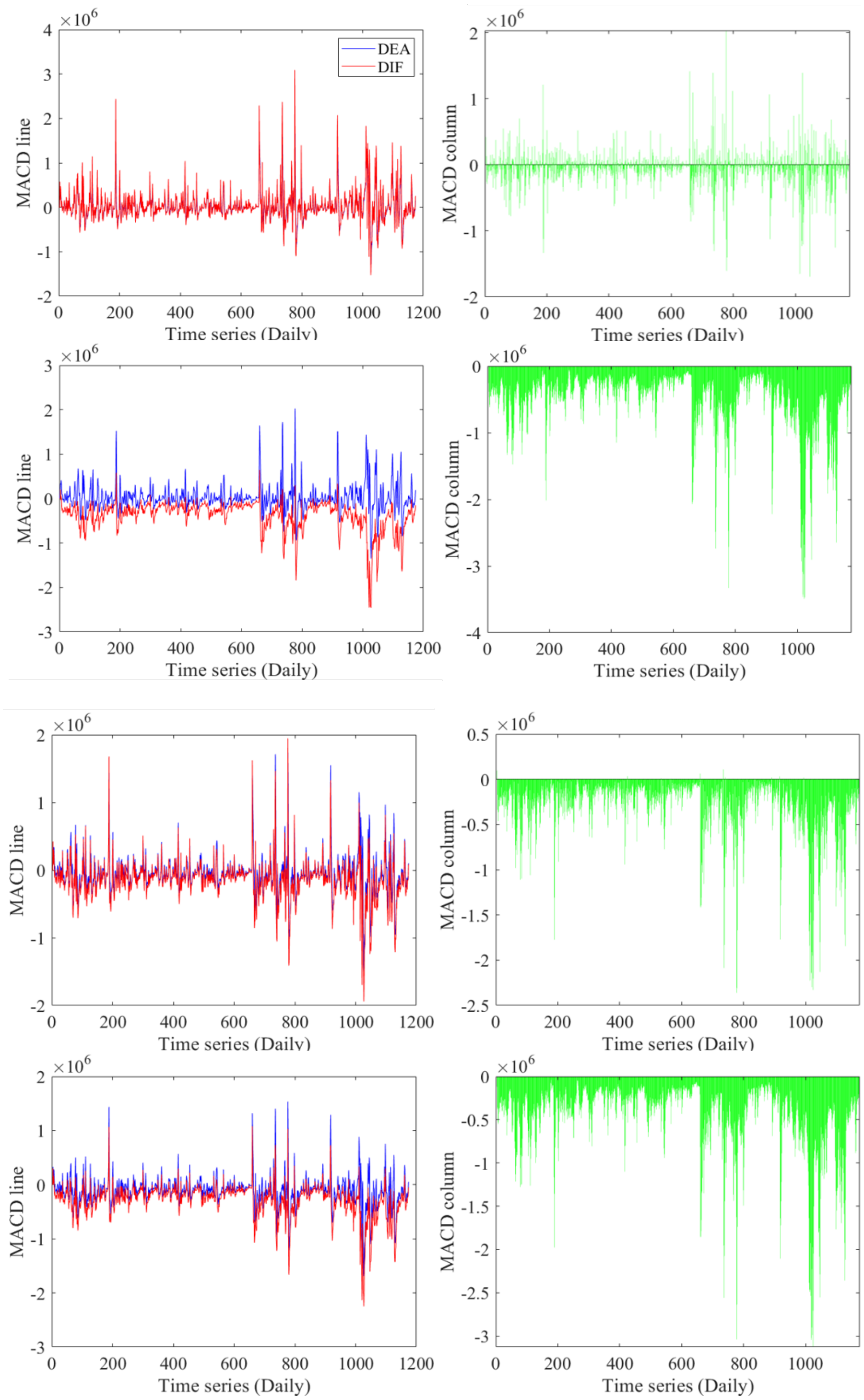

Figure 1. Analysis of ***MACD*** index

Note: 1. From top to bottom are the opening price, the highest price, the lowest price and the closing price;

2. The left column is ***MACD*** fast line (***DIF)*** and slow line (***DEA***), and the right column is ***MACD*** histogram.

This study combines the ***MACD*** indicator usage rules (Table 1-2, Figure 2) to analyze the daily trend of the stock.

Table 1. ***MACD*** line usage rules

DIF	***DEA***	***Signal***	***Conclusion***
>0	>0	***DIF*** breaks through ***DEA*** upward	Buy signal
<0	<0	***DIF*** fell to break ***DEA***	Sell signal
--	--	***DIF*** line deviated from ***K*** line	The market may reverse
The values of ***DIF*** and ***DEA*** from positive to negative or from negative to positive			Not a trading signal (behind the market)

Table 2. Rules for using ***MACD*** column

Signal	***Conclusion***
MACD golden fork: ***DIFF*** breaks through ***DEA*** from bottom to top	Buy signal
MACD dead fork: ***DIFF*** breaks through ***DEA*** from top to bottom	Sell signal
MACD value changes from negative to positive	The market turns from short to long
MACD value changes from positive to negative	The market turns from long to short
Both ***DIFF*** and ***DEA*** are positive and are likely to belong to a long market. ***DIFF*** breaks through ***DEA*** upwards	Buy signal
Both ***DIFF*** and ***DEA*** are negative, which is likely to be a short market, ***DIFF*** fell below ***DEA***	Sell signal
When the ***DEA*** line deviates from the ***K*** line trend	The market may reverse
DEA has a higher turnover rate during consolidation	Cooperate with ***RSI*** and ***KDJ*** indicators to make up for shortcomings appropriately

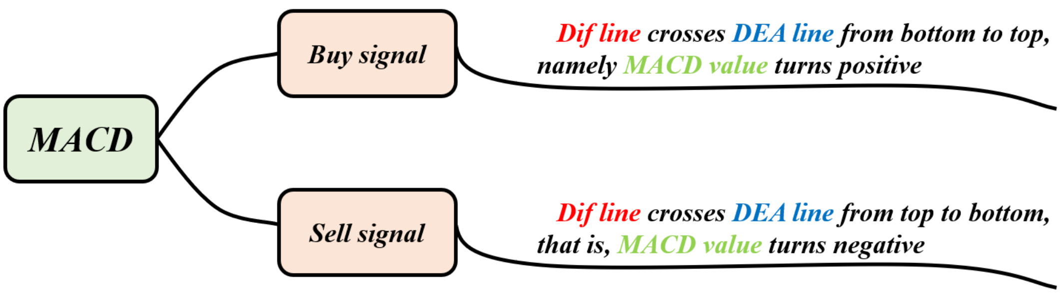

Figure 2. Schematic diagram of ***MACD*** index using method

In the first row and left column of Figure 1, when the market opens every day, the line basically completely crosses the line from bottom to top or top to bottom, that is, buy and sell at the opening. In the right column of the first row, when the market opens every day, the bar trend basically changes from negative to positive or from positive to negative, which corresponds to one-to-one correspondence between the bottom-up or top-down crossing of the center line in the left column of the first row

The left column of the second row of Figure 1 changes synchronously and reversely with the line. Most of the time, the line crosses the line from top to bottom, and rarely the line crosses the line from bottom to top. It is suitable for selling but not suitable for holding positions. The bar on the right column of the row is basically below zero, which often means the beginning of a new wave of decline, with extremely high reliability and a large decline. And the bigger the line breaks down, the bigger the bar line breaks down. Line to line floats between -1 and 1 most of the time, a small part of the time floats between -2 and 2, and even less time floats between -3 and 3.

Most of the time in the third row and left column of Figure 1 is the line crossing the line from top to bottom, that is, the time suitable for selling is much longer than the time suitable for buying. Corresponding to the right column of the third row, ***DIF*** and ***DEA*** are both below the zero axis, which is a dead cross at zero, which often means the beginning of a new wave of decline, and the reliability is extremely high, and the decline is also large. In a very small amount of time, the ***DIF*** line crosses the ***DEA*** from bottom to top, and the ***MACD*** turns from below zero to above zero, which means that the stock price has turned from falling to rising, which means the beginning of a long market.

The left column of the fourth row of Figure 1 clearly crosses the line from top to bottom, which is suitable for selling, not suitable for buying and holding positions. Corresponding to the right column of the fourth row, ***DIF*** and ***DEA*** are both below the zero axis, which means that the stock price has fallen, and the reliability is extremely high, and the decline is also large.

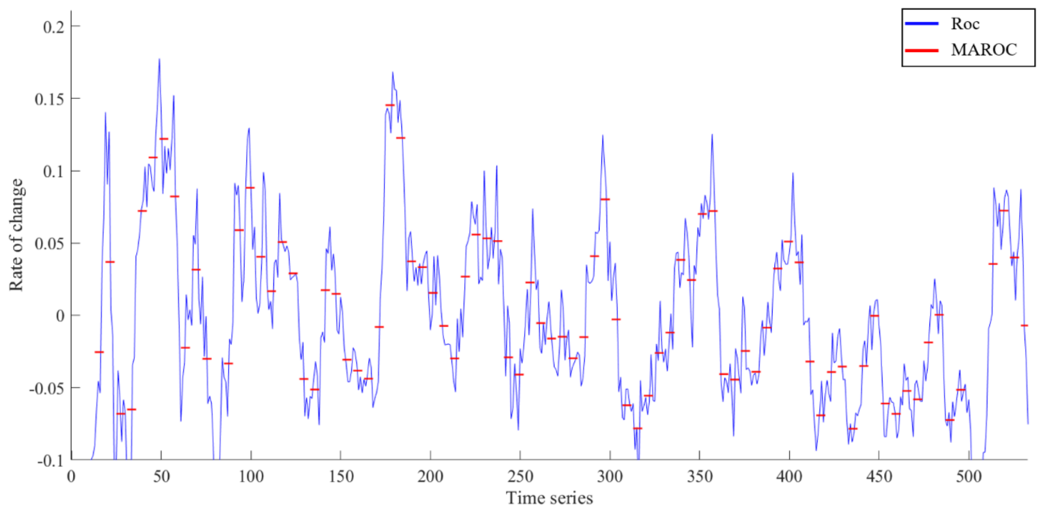

Figure 3. Roc and ***MAROC*** indicators（Intercepted part）

This study combines the use rules of ***Roc*** and ***MAROC*** indicators (Table 3, Figure 4) to analyze the daily trend of the stock.

Table 3. ***Roc*** and ***MAROC*** indicators usage rules

Signal	***Conclusion***
ROC wears through from bottom to top 0	Buy signal
ROC drops below 0 from the top 0	Sell signal
ROC penetrates ***MAROC*** from bottom to top	Buy signal
ROC fell from top to below ***MAROC***	Sell signal
When the stock price hits a new low and the ***ROC*** does not cooperate with the decline, it means that the downward momentum is weakened.	Deviation phenomenon should be undertaken on dips
When the stock price hits a new high and the ***ROC*** fails to rise, it means that the upward momentum is weakened.	Deviation phenomenon should be cautious to prevent the stock price from turning down.
Stock price and ***ROC*** are rising simultaneously at low levels	The short-term trend is normal or the stock price rebounds in the short-term
Stock price and ***ROC*** are falling simultaneously at a high level	The short-term trend is normal or the stock price will fall in the short-term
ROC fluctuates within the "normal range" and rises to the first overbought line,	Sell signal
ROC fluctuates within the "normal range" and drops to the first oversold line,	Buy signal
When ***ROC*** crosses the third overbought line upwards,	Try not to sell shares easily.
When ***ROC*** crosses the third oversold line downwards,	Should refrain from buying stocks easily.
When the ***ROC*** indicator crosses the third overbought and oversold line,	Hand over stock to ***SAR*** indicator

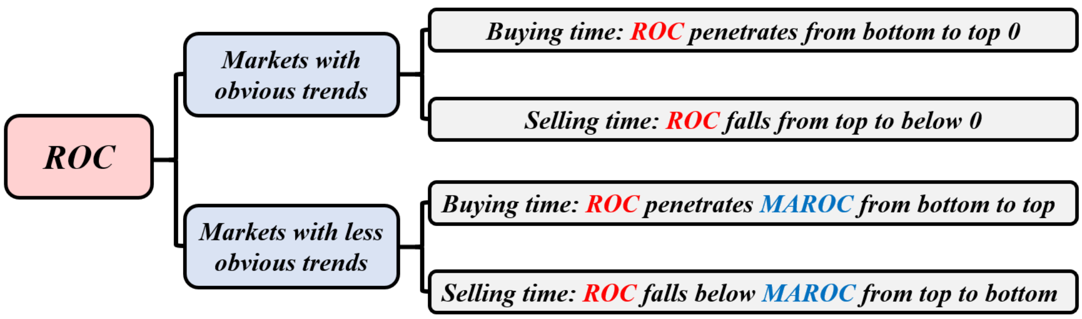

Figure 4. Schematic diagram of ***Roc*** and ***MAROC*** indicators

It can be clearly seen from Figure 3 that the ***ROC*** indicator line fluctuates between -0.1-0.1 most of the time. In a market with a clear trend, the time for ***ROC*** to break through 0 from bottom to top is minimal, that is, the time for buying is very good. It takes more time for ***ROC*** to break through 0 from top to bottom than it takes for ***ROC*** to break through 0 from bottom to top, that is, there are more opportunities for selling than for buying; in the trend of balanced and volatile markets, ***ROC*** from bottom to bottom The time to break ***MAROC*** upward is much longer than the time to break 0 from bottom to top in a market with obvious trends. That is, in an unobvious market, the time for buying has increased slightly, and ROC as a whole fell from top to bottom. The time to break ***MAROC*** is still dominant, that is, the timing of selling is dominant.

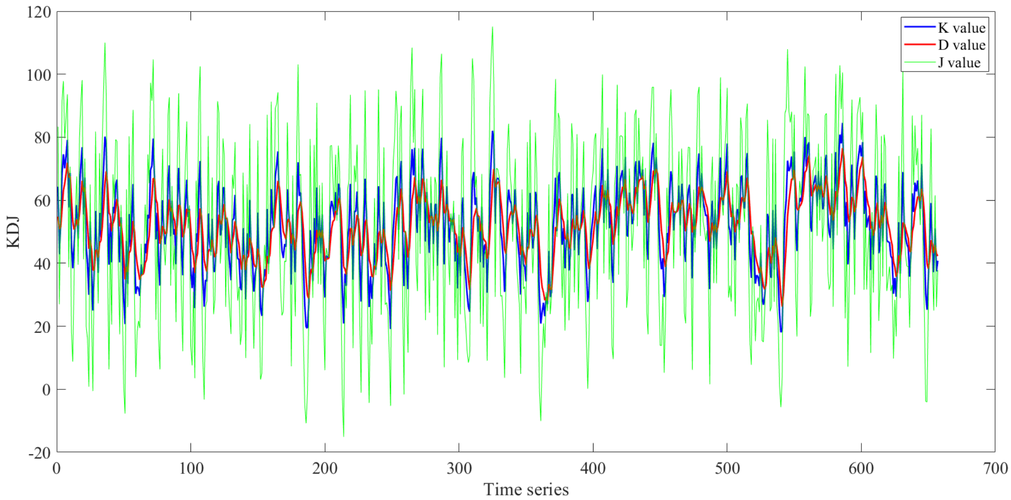

Figure 5. ***KDJ*** random index

This study combines the ***KDJ*** stochastic indicator usage rules (Table 4) to analyze the daily trend of the stock.

Table 4. ***KDJ*** random index usage rules

Signal	***Conclusion***
D> 80,	The market is overbought.
D<20	The market is oversold.
In an uptrend, ***K<D*** value, when ***K*** line breaks ***D*** line upward	Buy signal
In a downtrend, ***K>D*** value, when ***K*** line falls and breaks ***D*** line	Sell signal
When the stochastic index deviates from the stock price	Form change
The rate of increase or decrease of ***K*** and ***D*** values decreases, and the slope tends to be flat	Short-term change signal

It can be seen from Figure 5 that the values of ***K*** and ***D*** are between 20-80 most of the time, and the phenomenon of overbought and oversold is almost zero. The ***K*** value basically breaks upward completely or falls below the ***D*** value completely, that is, the buy and sell signals are obvious. It can be seen from Figure 5 that there are fewer overbought and oversold phenomena, but oversold phenomena are slightly more frequent than overbought.

Similarly, use the same method to analyze weekly and monthly data (Because the ***MACD*** indicator has a greater reference value at the close, the ***MACD*** indicator in the following text only considers the situation at the close).

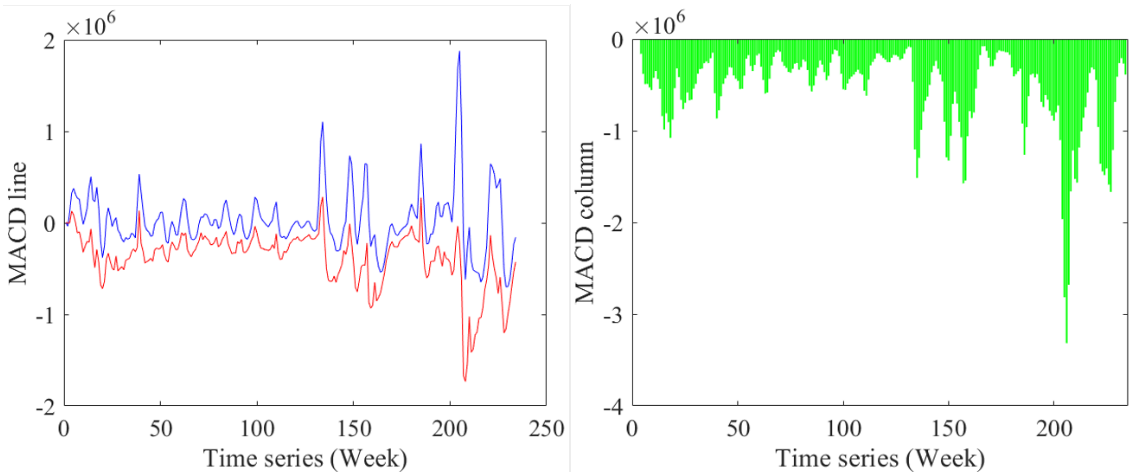

Figure 6. Analysis of ***MACD*** index (Week)

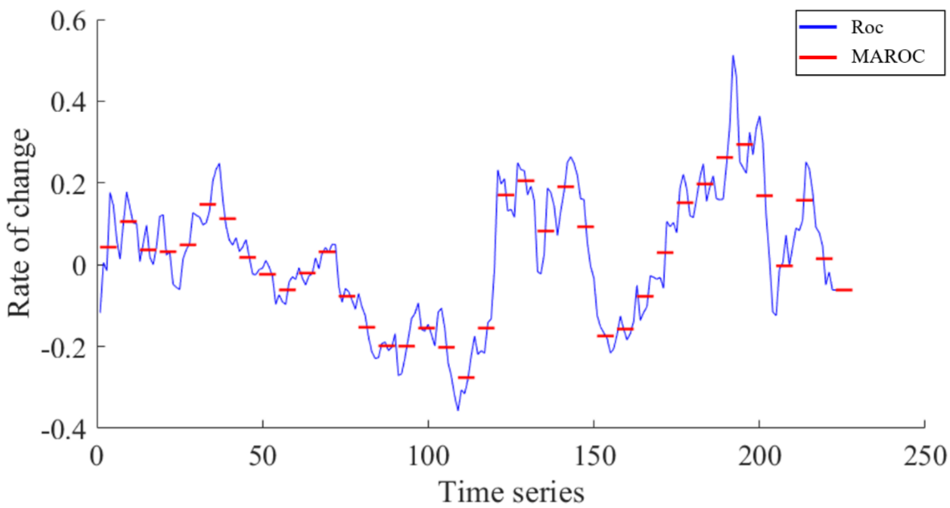

Figure 7. ***Roc*** and ***MAROC*** indicators (week)

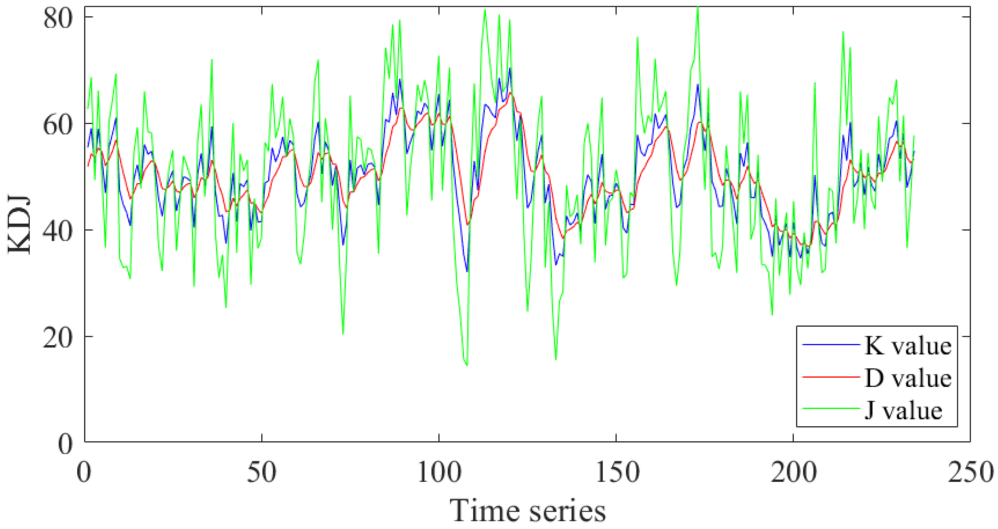

Figure 8. ***KDJ*** random index (Week)

It can be seen from Figure 6-8 that at the close of each day, the ***DIF*** line basically changes in synchronization with the ***DEA*** line. It is obvious from the figure that the ***DIF*** line and the ***DEA*** line do not overlap or cross, and there is no upward breakthrough or downward break. Both of them fluctuate between -1 and 1 most of the time. The ***MACD*** histogram fluctuates completely below zero, which basically corresponds to the downward deviation of the ***DIF*** line from the zero line. The ***ROC*** indicator line is mostly in the trend of balance and shock, and the ***ROC*** is from bottom to bottom. The time to break through ***MAROC*** is almost the same as the time to break from top to bottom, that is, in an unobvious market, the timing for buying is equivalent to the timing for selling; ***K*** and ***D*** values are mostly between 40-60 At that time, the time for the ***K*** line to break above the ***D*** line is equivalent to the time for the ***K*** line to break below the ***D*** line, that is, the timing for buying and selling is equivalent.

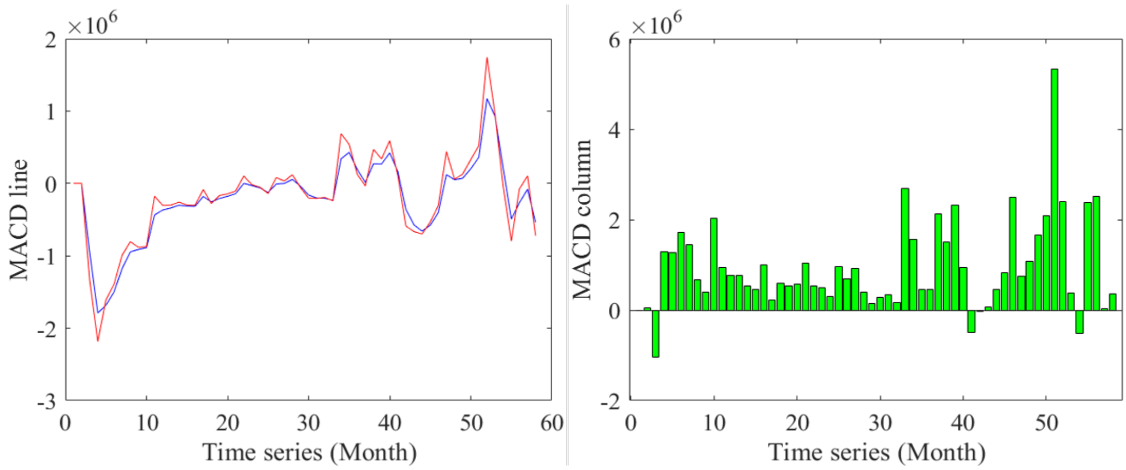

Figure 9. Analysis of ***MACD*** index (Month)

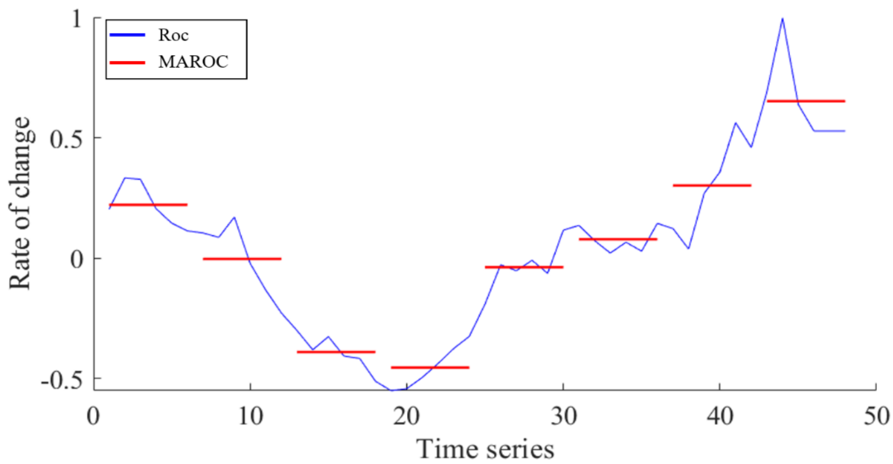

Figure 10. ***Roc*** and ***MAROC*** indicators (Month)

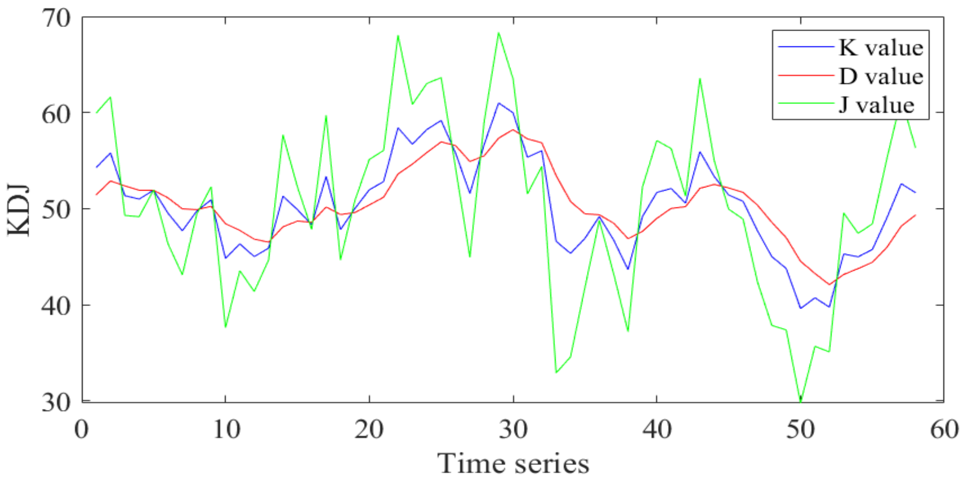

Figure 11. ***KDJ*** random index(Month)

From Figure 9-11, it can be seen that the ***DIF*** line breaks through the ***DEA*** line from bottom to top to occupy the dominant position, that is, the time to buy is more than the time to sell. The stock begins to have ***DIF*** and ***DEA*** lines below the zero line and down Moving means that you are in a short market, you can sell to open a position or wait and see, and then move upward, it means that the market is rising, you can buy to open a position or a long position. The ***MACD*** histogram fluctuates above zero most of the time, and the trend is the same as the ***DIF*** line and ***DEA*** The degree of deviation of the line from the zero line basically corresponds. When the ***DIF*** line falls from the top to the ***DEA*** line, the corresponding ***MACD*** histogram is below zero, that is, the time to sell; in the trend of the balance and shock market, the ***ROC*** falls from the top to the time when it breaks the ***MAROC*** In the early stage, the time for ***ROC*** to break through ***MAROC*** from bottom to top is in the later stage, that is, in an unobvious market, it is suitable for selling in the early stage and suitable for buying in the later stage; ***K*** and ***D*** values are mostly between 40-60, ***K*** line The time to break the ***D*** line upward is the same as the time to break the ***D*** line downward, that is, the time for buying and selling is the same.

The analysis methods of the remaining two types of stocks are the same as those of the 000400 stock. Due to the present value of space, they are not shown in the main text.

5. Question 2

5.1 Model preparation

This study constructed a multi-category feature system for stock trend prediction by consulting relevant literature, and divided stock data features into two categories, namely, basic indicators and technical indicators. As shown in Table 5.

Table 5. List of input characteristics

Feature type	***Feature name***
Basic index	1: Opening price; 2: Highest price; 3: Lowest price; 4: Closing price; 5: Volume; 6: Turnover rate
Technical index	7-11: ***EMA5***, ***EMA10***, EMA20, ***EMA30***, ***EMA60*** 12-14: ***MACD*** fast line, ***MACD*** slow line, ***MACD*** column 15-16: ***Roc***, ***MAROC*** 17-19: ***K***, ***D***, ***J*** stochastic indicators

The basic indicators include: opening price, highest price, lowest price, closing price, daily price change, trading volume, turnover rate, etc. Technical indicators are based on basic indicators and are obtained through further calculations, mainly including trend indicators and strength indicators based on price or transaction volume (see 5.2 for details of each technical indicator).

In addition, further preprocessing of the raw data of each indicator, including checking data consistency, processing invalid values, filling in missing values, and filtering data that does not meet the requirements, etc.

5.2 LSTM principle

Long short-term memory network (***LSTM***) is a kind of time cyclic neural network, which is specially designed to solve the long-term dependence problem of general RNN (circular neural network). All RNNs have a chain of repeated neural network modules. Style form.

Recurrent neural networks are good at solving problems related to time series, and stock trend forecasting is a very typical time series forecasting problem. Ordinary ***RNN*** will cause gradient disappearance and gradient explosion when calculating the connection between remote nodes, which cannot solve the problem of "long-term dependence". However, LSTM can effectively save the long-term through the unique "gate" structure and cell state update. memory.

5.3 Model establishment

This research defines the problem as a regression problem. It uses 20 consecutive days of stock data to predict the short-term stock trend in the next few days, and uses the next three days' return rate to express the stock's rise and fall trend, which is calculated by the following formul:

$$3daysPred = 0.4 \times \frac{closeprice(n+1)}{closeprice(n)} + 0.32 \times \frac{closeprice(n+2)}{closeprice(n+1)} + 0.28 \times \frac{closeprice(n+3)}{closeprice(n+2)}$$

Among them, $\boldsymbol{3daysPred}$ represents the three-day rate of return, and $\boldsymbol{closeprice(n)}$ represents the closing price on the nth day. If the three-day yield is greater than 1, it means that the closing price of the stock price in the next three days must be greater than the closing price of the previous day for at least one day, indicating that it has short-term investment value. In addition, a sliding window is used to input stock data into the model for training, and the sliding window size is set to 20, as shown in Figure 12.

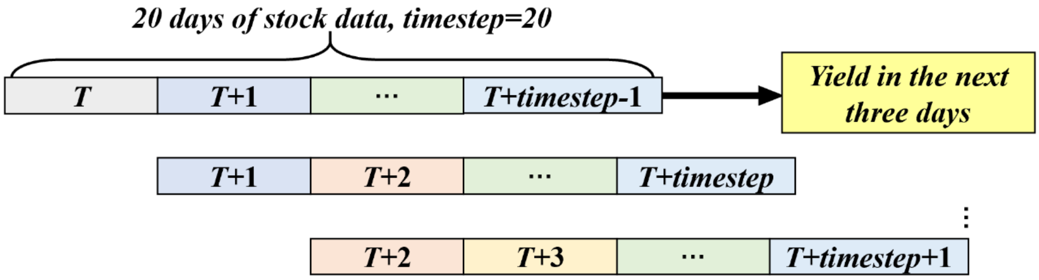

Figure 12. Sliding window settings for input variables

T represents the stock data on ***T*** day, including the 38 characteristics mentioned above. All the features of the day are expressed as $\boldsymbol{X_T} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\}$, where ***n***=38, and the input vector is represented by $\boldsymbol{X_T} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_{T+timestep}}\}$ $(\boldsymbol{timestep}) = \boldsymbol{20}$, The target vector ***Y*** is the three-day rate of return. The characteristic values of the basic indicators and technical indicators are standardized and quantified values.

This paper uses 20-day historical stock data to predict the stock rise and fall in the next 3 days, and it is suitable to use ***LSTM*** to capture the relationship between long-term series. The ***LSTM*** network structure is shown in Figure 13, using three special gate structures, "forgotten gate", "input gate" and "output gate" to form a loop body. The gate structure uses sig-moid neurons to control how much input information is retained. There is also a hidden memory structure C_t in ***LSTM***. This variable participates in the operation each time, but only outputs h_t, not C_t, but saves it and passes it to the next operation.

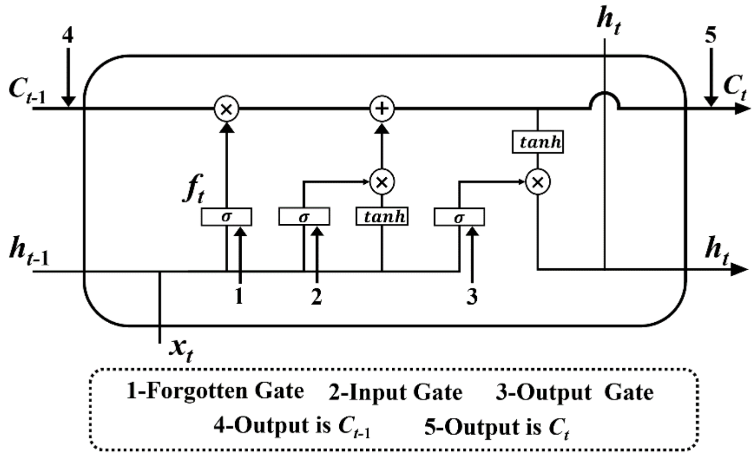

Figure 13. ***LSTM*** network structure diagram

According to the structure of the ***LSTM*** network, the calculation of each ***LSTM*** unit is as follows:

$$X_T = \{x_1, x_2, \dots, x_{T+timestep}\}timestep$$

$$f_t = \sigma\left(W_f \cdot \left[\hbar_{t-1}, x_t\right] + b_f\right)$$

$$i_t = \sigma\left(W_i \cdot \left[\hbar_{t-1}, x_t\right] + b_i\right)$$

$$\widetilde{C_t} = tan\hbar(W_C \cdot \left[\hbar_{t-1}, x_t\right] + b_C)$$

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C_t}$$

$$o_t = \sigma\left(W_o \cdot \left[\hbar_{t-1}, x_t\right] + b_o\right)$$

$$\hbar_t = o_t \times tan\hbar(C_t)$$

Where $\boldsymbol{f_t}$ represents the forget gate, $\boldsymbol{i_t}$represents the input gate, $\boldsymbol{O_t}$ represents the output gate, $\boldsymbol{x_t}$ represents the input at this moment, ***t*** represents the information to be input, $\boldsymbol{C_{t-1}}$ represents the cell state at the previous moment, and $\boldsymbol{C_t}$ represents the cell state at this moment. $\boldsymbol{h_t}$ represents the output of the current unit, and $\boldsymbol{h_{t-1}}$ represents the output of the previous unit.

So far, this article has completed the establishment of a stock trend prediction model based on ***LSTM***.

5.4 Model solving

This research constructs features based on the preprocessed data set, divides the processed data set into training set and test set, selects relevant input features, and establishes model training. Finally, the model prediction results are compared with the real measurement results, and the prediction accuracy and ***MSE*** are used to evaluate.

Continuous adjustments are made according to the evaluation results to improve the generalization ability of the model. The processing flow is shown in Figure 14.

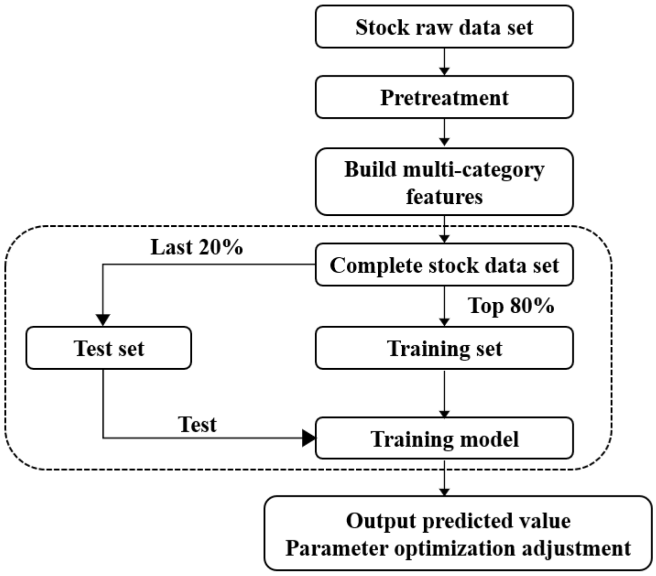

Figure 14. Training process flow chart

In the training process, in order to obtain better model performance, the parameters of the ***LSTM*** model are set and optimized as follows:

(1) Perform Z-score standardization on the attribute values in the original data to make the processed data conform to the standard normal distribution to accelerate the model convergence and ensure the validity of the training data;

(2) In order to reduce the memory burden and speed up the training speed, this study will set the value to 60, calculate the loss function of one batch at a time, and combine the back propagation algorithm to optimize the parameters;

(3) According to the size of the input and output vectors, set the number of hidden nodes to 20, the length of the ***LSTM*** loop body to 20 according to the length of the time series, and the number of hidden layers to 2;

(4) Use dropout between the loop bodies of different layers of ***LSTM***, and randomly let some hidden layer nodes not work during the model training phase to improve the generalization ability of the neural network;

(5) The activation function adopts the Relu function, which has a smaller amount of calculation, faster calculation speed, and avoids the phenomenon of gradient disappearance.

This study uses Matlab software to implement the above ***LSTM*** model, and the iteration results are shown in Figure 15.

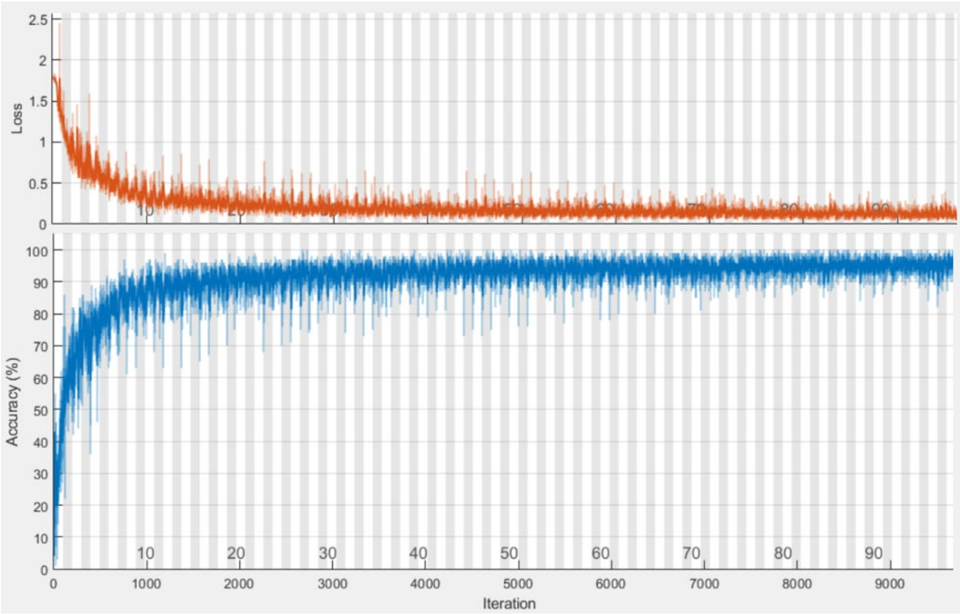

Figure 15. ***LSTM*** iteration results

As can be seen from the above figure, as the number of iterations continues to increase, the accuracy of the model continues to improve, and finally stabilizes between 90-100%; the model loss continues to decrease, and finally stabilizes below 0.5. This shows that the ***LSTM*** model training and prediction results are excellent.

In addition, the prediction comparison results of different output features on the LSTM model are obtained, as shown in Table 6.

Table 6. Comparison of prediction results of different input features on ***LSTM*** model

Different category characteristics	***Forecast accuracy***
Basic index	66.78%
Basic indicators + technical indicators	67.13%

5.5 Model checking

This study uses the ***LSTM*** model established in 6.3 to predict the price trend of the 000400 stock in the past 20 days, and uses the historical data of the past 20 days for evaluation. The result is shown in Figure 16.

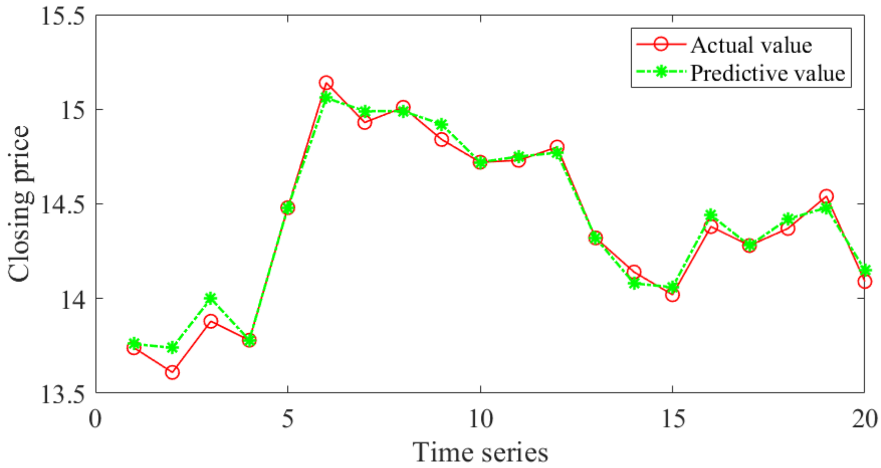

Figure 16. Test results

It can be seen from the figure above that in most cases, the prediction result of the stock price trend of the ***LSTM*** model is close to the actual situation, which shows that the model established in this research is effective and reliable.

6. Question 3

6.1 Selection and introduction of key turning points

This question requires the establishment of a stock price reversal judgment model based on the research results of Question 2. In order to better reflect the cyclical trend of stock price reversals, this research introduces key turning point indicators that measure the important turning points of stock prices. The results are shown in Table 7.

Table 7. List of key turning point indicators(Continued table 5)

Feature type	Feature name
Key turning point	20: The golden fork of the 5-day ***EMA*** line and the 10-day ***EMA*** line 21: The golden cross between the 10-day ***EMA*** line and the 30-day ***EMA*** line 22: ***K***-line and ***D***-line in ***KDJ'***s golden cross

6.2 Establishment of the judgment model of stock price reversal

Enter the key turning point indicators in 7.1 into the ***LSTM*** model in Question 2, and define as follows:

Ⅰ. If a golden cross appears on the ***EMA*** line, it is marked as 1, and a dead cross is marked as -1;

Ⅱ. If there is a golden cross on the ***KDJ*** line, it is marked as 1, and if a dead cross occurs, it is marked as -1;

Ⅲ. If either the ***EMA*** line or the ***KDJ*** line has a golden cross or a dead cross, it is marked as 0.

Similarly, use Matlab to implement the model, and get the prediction comparison results of different output features on the ***LSTM*** model. The results are shown in Table 8.

Table 8. Comparison of prediction results of different input features on ***LSTM*** model (Continued table 6)

Different category characteristics	Forecast accuracy
Basic indicators + Ttechnical indicators + Key turning points	68.26%

It can be seen from the above table that when the key turning point is introduced, the accuracy of model prediction has been further improved, indicating that the methods and ideas adopted in this study are feasible.

So far, this research is based on the problem 2 stock price trend prediction model, and establishes a price reversal judgment model that can reflect the periodicity of stock price changes.

6.3 Result analysis

Similarly, predict the price reversal position of the 000400 stock in the past 20 days, and use the historical data of the past 20 days for evaluation. The result is shown in Figure 17.

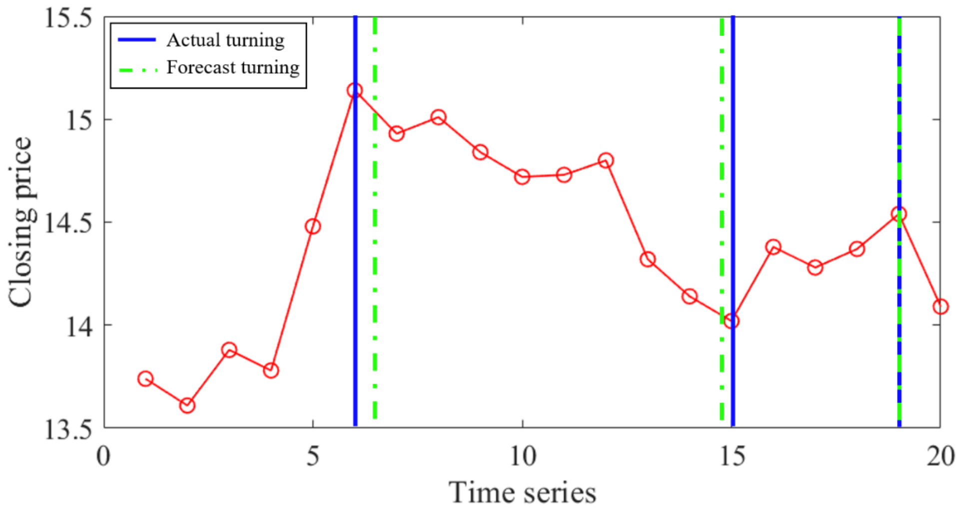

Figure 17. Forecast results of stock price reversal

It can be seen from the above figure that in most cases, the stock price reversal prediction model based on ***LSTM***, the prediction result of the stock price reversal position is close to the actual situation, which shows that the model established in this research is effective and reliable.

7.Strengths and Weakness

Strengths:

LSTM model has certain advantages in sequence modeling and has long-term memory function. Simple to implement. Solve the problem of gradient disappearance and gradient explosion in the process of long sequence training.

Weakness:

The LSTM model has disadvantages in parallel processing.

References

[1]Zhuoxi Yu et al. "Stock price forecasting based on LLE-BP neural network model". 553(2020)

[2]Sanjiban Sekhar Roy et al. "Random forest, gradient boosted machines and deep neural network for stock price forecasting: a comparative analysis on South Korean companies". 33.1(2020)

[3]Kun Zhou et al. "Comparison of Time Series Forecasting Based on Statistical ARIMA Model and LSTM with Attention Mechanism". 1631.1(2020):012141-.

[4]Yanwen Xue and Jun Jiang and Ling Hong. "A LSTM based prediction model for nonlinear dynamical systems with chaotic itinerancy". 8.prepublish(2020):1-12.

[5]Hongli Niu and Kunliang Xu and Weiqing Wang. "A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network". (2020):1-14.

Appendix

1. Matlab code for solving technical indicatorsclear;clc

```
%%
d_000400=xlsread('D:\000400.SZ.xlsx','Sheet1');
%%
%EMA5
Ema5=[];EMA5=[];
EMA_0=d_000400(1,:);
for i=2:length(d_000400)
    for j=1:length(EMA_0)
        if i == 2
            ema5=(2*d_000400(i,j)+4*EMA_0(1,j))/6;
        else
            ema5=(2*d_000400(i,j)+4*ema5)/6;
```

```
        end
        Ema5=[Ema5,ema5];
    end
    EMA5=[EMA5;Ema5];
    Ema5=[];
end
EMA5=[EMA_0;EMA5];
%EMA10
Ema10=[];EMA10=[];
EMA_0=d_000400(1,:);
for i=2:length(d_000400)
    for j=1:length(EMA_0)
        if i == 2
            ema10=(2*d_000400(i,j)+9*EMA_0(1,j))/11;
        else
            ema10=(2*d_000400(i,j)+9*ema10)/11;
        end
        Ema10=[Ema10,ema10];
    end
    EMA10=[EMA10;Ema10];
    Ema10=[];
end
EMA10=[EMA_0;EMA10];
%EMA20
Ema20=[];EMA20=[];
EMA_0=d_000400(1,:);
for i=2:length(d_000400)
    for j=1:length(EMA_0)
        if i == 2
            ema20=(2*d_000400(i,j)+19*EMA_0(1,j))/21;
        else
            ema20=(2*d_000400(i,j)+19*ema10)/21;
        end
        Ema20=[Ema20,ema20];
    end
    EMA20=[EMA20;Ema20];
    Ema20=[];
end
EMA20=[EMA_0;EMA20];
```

```
%EMA30
Ema30=[];EMA30=[];
EMA_0=d_000400(1,:);
for i=2:length(d_000400)
    for j=1:length(EMA_0)
        if i == 2
            ema30=(2*d_000400(i,j)+29*EMA_0(1,j))/31;
        else
            ema30=(2*d_000400(i,j)+29*ema10)/31;
        end
        Ema30=[Ema30,ema30];
    end
    EMA30=[EMA30;Ema30];
    Ema30=[];
end
EMA30=[EMA_0;EMA30];
%EMA60
Ema60=[];EMA60=[];
EMA_0=d_000400(1,:);
for i=2:length(d_000400)
    for j=1:length(EMA_0)
        if i == 2
            ema60=(2*d_000400(i,j)+59*EMA_0(1,j))/61;
        else
            ema60=(2*d_000400(i,j)+59*ema10)/61;
        end
        Ema60=[Ema60,ema60];
    end
    EMA60=[EMA60;Ema60];
    Ema60=[];
end
EMA60=[EMA_0;EMA60];
%%
%EMA12
Ema12=[];EMA12=[];
EMA_0=d_000400(1,:);
for i=2:length(d_000400)
    for j=1:length(EMA_0)
        if i == 2
```

```
            ema12=(2*d_000400(i,j)+11*EMA_0(1,j))/13;
        else
            ema12=(2*d_000400(i,j)+11*ema12)/13;
        end
        Ema12=[Ema12,ema12];
    end
    EMA12=[EMA12;Ema12];
    Ema12=[];
end
EMA12=[EMA_0;EMA12];
%EMA26
Ema26=[];EMA26=[];
EMA_0=d_000400(1,:);
for i=2:length(d_000400)
    for j=1:length(EMA_0)
        if i == 2
            ema26=(2*d_000400(i,j)+25*EMA_0(1,j))/27;
        else
            ema26=(2*d_000400(i,j)+25*ema26)/27;
        end
        Ema26=[Ema26,ema26];
    end
    EMA26=[EMA26;Ema26];
    Ema26=[];
end
EMA26=[EMA_0;EMA26];
%DIF
DIF=EMA12-EMA26;
%DEA
Dea=[];DEA=[];
DEA_0=DIF(1,:);
for i=2:length(DIF)
    for j=1:length(DEA_0)
        if i==2
            dea=DEA_0(1,j)*0.8+DIF(i,j)*0.2;
        else
            dea=dea*0.8+DIF(i,j)*0.2;
        end
        Dea=[Dea,dea];
```

```
        end
        DEA=[DEA;Dea];
        Dea=[];
end
DEA=[DEA_0;DEA];
%MACD-z
Macdz=[];MACDz=[];
for i=1:length(DEA)
        for j=1:length(DEA_0)
                macdz=(DIF(i,j)-DEA(i,j))*2;
                Macdz=[Macdz,macdz];
        end
        MACDz=[MACDz;Macdz];
        Macdz=[];
end
%
HJ=[DIF,DEA,MACDz];
for i=1:4
        figure
        bar(HJ(:,i+12),'g');
        set(gca,'Fontsize',16,'FontName','Times New Roman');
        set(get(gca,'xlabel'),'string','Time series (Month)');
        set(get(gca,'ylabel'),'string','MACD column');
        figure
        plot(HJ(:,i+7),'b');hold on;
        plot(HJ(:,i),'r');
        set(gca,'Fontsize',16,'FontName','Times New Roman');
        set(get(gca,'xlabel'),'string','Time series (Month)');
        set(get(gca,'ylabel'),'string','MACD line');
        if i==1
                legend('DEA','DIF');
        end
        hold off;
end
%%
%Roc
AX=[];BX=[];
for i=13:length(d_000400)
        bx=d_000400(i-12,4);
```

```
    ax=d_000400(i,4)-d_000400(i-12,4);
    AX=[AX;ax];BX=[BX;bx];
    ROC=AX./BX;
end
%MAROC
ROC=[ROC;ROC(end,1);ROC(end,1)];
MAROC=[];
for i=1:194
    maroc=sum(ROC((1+(i-1)*6):i*6,1))/6;
    MAROC=[MAROC;maroc;maroc;maroc;maroc;maroc;maroc];
end
MAROC_x=[1:1:1164]';
figure;hold on;
plot(MAROC_x,ROC,'b');
for i=1:194
    ma=plot(MAROC_x((1+(i-1)*6):i*6,1),MAROC((1+(i-1)*6):i*6,1),'r');
    set(ma,'LineWidth',1.5);
end
hold off;
%%
%KDJ
RSV=[];
for i=1:length(d_000400)
    Cn=d_000400(i,1);Ln=d_000400(i,3);Hn=d_000400(i,2);
    rsv=((Cn-Ln)/(Hn-Ln))*100;
    RSV=[RSV;rsv];
end
KDJ=[];
for i=1:length(RSV)
    if i==1
        k=50;d=50;
    end
    k=(2/3)*k+(1/3)*RSV(i,1);
    d=(2/3)*d+(1/3)*k;
    j=3*k-2*d;
    kdj=[k,d,j];
    KDJ=[KDJ;kdj];
end
plot(KDJ(:,1),'b');hold on;
```

```
plot(KDJ(:,2),'r');
plot(KDJ(:,3),'g');hold off;
```

2. The Matlab code of the LSTM model

```
XTrainSize = length(d_000400);
XTrainData = dataStandardlized(1:XTrainSize,:);
XTrainLabel = dataStandardlizedLable(1:XTrainSize,:);
%XTrain
for i = 1:size(XTrainData,1)
    XTrain{i,1} = XTrainData(i,:);
end
%YTrain
TrainstrLable = num2str(XTrainLabel);% num to str
for i = 1:size(XTrainData,1)% str matrix to cell
    TraincellLable{i,1} = TrainstrLable(i,1);
end
YTrain = categorical(TraincellLable);%cell to categorical
XTestData = dataStandardlized(1+XTrainSize:end,:);
XTestLabel = dataStandardlizedLable(1+XTrainSize:end,:);
%XTest
for i = 1:size(XTestData,1)
    XTest{i,1} = XTestData(i,:);
end
%YTest
TeststrLable = num2str(TestLabel);
for i = 1:size(XTestData,1)
    TestcellLable{i,1} = TeststrLable(i,1);
end
YTest = categorical(TestcellLable);
inputSize = 1;
numHiddenUnits = 100;
numClasses = 6;
layers = [ ...
    sequenceInputLayer(inputSize)
    bilstmLayer(numHiddenUnits,'OutputMode','last')
    fullyConnectedLayer(numClasses)
    softmaxLayer
    classificationLayer]
```

```
maxEpochs = 100;
miniBatchSize = 100;
options = trainingOptions('adam', ...
    'ExecutionEnvironment','cpu', ...
    'GradientThreshold',1, ...
    'MaxEpochs',maxEpochs, ...
    'MiniBatchSize',miniBatchSize, ...
    'SequenceLength','longest', ...
    'Shuffle','never', ...
    'Verbose',0, ...
    'Plots','training-progress');
net = trainNetwork(XTrain,YTrain,layers,options);
miniBatchSize = 100;
YPred = classify(net,XTest, ...
    'SequenceLength','longest','MiniBatchSize',miniBatchSize);
acc = sum(YPred == YTest)./numel(YTest)
```