Team Control Number

202011151508

Problem Chosen

B

Research on Stock Forecasting Method Based on Hidden Markov Model and the Elastic Feedback Algorithm

Summary

In this paper,we verified the chaotic characteristics of the three stocks and predicted their ups and downs. Firstly, Lyapunov index is calculated based on chaos theory to verify the chaotic characteristics of stock data; Then Hidden Markov Model(HMM) is established to predict the rising and falling trend of stocks. Finally, in view of the randomness of stock market price fluctuation, the Elastic Feedback Algorithm is introduced to improve the prediction accuracy of the model.

For the first question, in the beginning, according to the daily, weekly and monthly charts of the three stocks in recent five years, it is found that, all the three stocks show a non-linear fluctuation upward trend in general. Then, the chaotic characteristics of three stocks are tested. After preprocessing the stock data logarithmically, the best delay time is determined by using autocorrelation data processing method, and the embedding dimension is obtained by combining with G-P algorithm. At last, the Lyapunov index is calculated by Wolf method, which verifies that the stock data is chaotic.

For the second question, at first, we established the HMM, the parameters needed for HMM are solved by using forward-backward algorithm, Baum-Welch algorithm and Viterbi algorithm. Secondly, we test the sensitivity of hidden states. It is found that there is no significant difference in the results under different hidden states. Finally, we tested it with the data of nearly 20 days.

For the third problem, in the beginning,based on the HMM established in Question 2, the Elastic Feedback Algorithm is adopted, in which HMM is used to predict the hidden state, and the Elastic Feedback Algorithm is used to predict the classified state. Therefore, a linear combination of the two can be used to further predict the trend of stocks. After that, the model was tested again, the results showed that the prediction accuracy of the three stocks increased by 15%, 10% and 5% respectively.

The innovation of this paper lies in the combination of HMM and Elastic Feedback Algorithm, thus the compound state is obtained. Based on this, the rising and falling trend of the stock can be predicted, which further improves the prediction accuracy of the model.

Key word: Stocks Forecast;Lyapuno;Hidden Markov Model;Elastic Feedback Algorithm

Content

1.Introduction

1.1 Background

The stock market is obviously influenced by many factors such as economy, politics and psychology, and its uncertainty is very high. The operating conditions of listed companies and market recognition can be measured by changes in stock prices. Because of the randomness and trend of stock price changes, the prediction and modeling of stock price has become a difficult problem. Obviously, the stock market is consistent with the characteristics of nonlinear complex systems, and chaos theory can be used to solve the modeling and application of nonlinear complex systems, and some achievements have been made in the fields of power and communication. Therefore, it is necessary for us to conduct relevant research with the help of chaos theory to provide theoretical basis for the prediction of stock prices.

1.2 Work

According to the understanding of the problem, this paper mainly solves the following three problems:

1.The daily, weekly and monthly trends of three stocks are quantitatively analyzed by descriptive analysis, and Lyapunov index model is established to analyze the chaotic characteristics of stock data.

2.Based on chaos theory, a prediction model of stock price rise and fall is established mainly by using HMM, and the prediction accuracy of the model is tested by using the data of the past 20 days.

3.On the basis of question 2' s model, according to the trend pointed out in the topic, the stock prices will fall if they rise for a long time, and will rise if they fall for a long time, we set up an elastic feedback algorithm to modify the prediction model, thus improving the prediction accuracy of the model. Besides, the model is tested with data.

2.Problem analysis

2.1 Analysis of question one

For the first question, we need to analyze the stock qualitatively and test its chaotic characteristics. First of all, we use Matlab to draw the daily K-line chart, weekly K-line chart and monthly K-line chart of these three stocks from January 2016 to October 2020, and draw the corresponding 60-day average line and 10-day average line in the charts. By analyzing the charts, it is concluded that all these three stocks have nonlinear fluctuation trends, among which the stock 600519 has increased by 730.8% in recent five years. Secondly, we use Lyapunov exponent model to run the relevant data of stock opening price, and draw the conclusion that the data of stock opening price is chaotic. At the same time, other stock data are also tested, and the results show that the other five types of stock data also have chaotic characteristics.

2.2 Analysis of question two

For the second problem, it is necessary to establish a prediction model of stock ups and downs, and use the data of the last twenty days to test it. Firstly, based on chaos theory, a prediction model of stock ups and downs is established mainly by using HMM, and the state transition parameters of the model are trained by using the data of three stocks provided, and finally the parameters needed by HMM are obtained. Then, according to the training model, the data of the last 20 days are used for back testing, and the prediction accuracy of the model is analyzed. Finally, according to the model accuracy, the problems of HMM are analyzed.

2.3 Analysis of question three

For the third problem, we need to consider the reverse phenomenon in the stock market, so as to further improve the prediction accuracy of the model. Therefore, firstly, according to the model established in Question 2, using the trend of stocks, the state of stocks is classified into three categories. Then, combining the state classification results of HMM, an elastic feedback algorithm is established by using linear combination. Finally, the prediction results of the model are analyzed more accurately.

3.Symbol and Assumptions

3.1 Symbol Description

Table 3.1 Symbol Description

Symbol	Meaning
$x(t)$	Time series
τ	Delay time
$C(\tau)$	Autocorrelation function value
m	Embedding dimension
$P(O\mid\lambda)$	Probability value of observation sequence
$b(k)$	Density function
L	Length of observation sequence
N	Number of hidden states

3.2 Fundamental assumptions

Hypothesis 1: In the process of forecasting, the volatility of the market itself is not considered, and only the fundamental data is used to forecast the model.

Hypothesis 2: In the process of analysis, the calculation method of stock rise and fall is the same as that commonly used at present.

Hypothesis 3: The data is reliable in the analysis process;

Hypothesis 4: The established model is effective within the acceptable error range.

Hypothesis 5: The state of Markov Chain at any moment is only related to the state at the previous moment, and has nothing to do with the state at other moments.

Hypothesis 6: The observed state only depends on the state of the Hidden Markov Chain at the current moment, and has nothing to do with other states.

4.Models and results

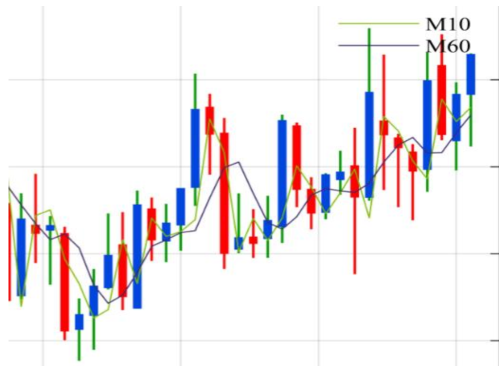
4.1 Model I Models and Results

4.1.1 Stock trend analysis

(a) Stock 002281.SZ daily trend chart

(b) Stock 002281.SZ weekly trend chart

(c) Stock 002281.SZ monthly trend chart



(d) Stock 002281.SZ monthly trend detail chart

Fig 4.1 002281.SZ Stock trend analysis chart

According to the data provided by the topic, the daily K-line chart, weekly K-line chart and monthly K-line chart of three stocks in recent four years are made by Matlab, in which we choose 002281.SZ stock to make qualitative and quantitative analysis, taking its daily opening price as the analysis basis, and the specific figures of the other two stocks will be displayed in the appendix. In figure (a), due to the large amount of stock data, the negative and positive lines in the figure are densely distributed. It can be seen that from September 2016 to October 2020, the stock as a whole showed a nonlinear upward trend, with the lowest opening price of 14.3 yuan per share in a single day and the highest opening price of 36.8 yuan per share in a single day. In Figure (b), the distribution of negative and positive lines is relatively dense, and the overall stock is still on the rise. The lowest average opening price in a single week is 15.3 yuan per share, and the highest average opening price in a single week is 35.7 yuan per share.

In Figure (c), due to the small amount of data, the negative line,positive line, 10-day average moving line and 60-day average moving line can be clearly distinguished. Figure (d) is a partial display of Figure (c), which is not difficult to see from June 2018 to March 2019. The average monthly opening price of stocks soared, and then it showed a nonlinear upward trend, reaching the highest value in October 2019,with the monthly average opening price of 33.9 yuan per share.

4.1.2 Establishment of lyapunov index model

Based on Packad et al's Coordinate delay phase space reconstruction method,for one-dimensional time series $\{x(t)\}, t=1,2,\cdots N$,we can construct a vector of m dimensions.

$$X_n=\{x(n),x(n+\tau),\cdots,x(n+(m-1)\tau)\}, n=1,2,\cdots,N-(m-1)\tau \tag{4.1}$$

Where m is the embedding dimension and τ is the delay time.The key of phase space reconstruction lies in the determination of embedding dimension and delay time. Taken theorem shows that we can simulate a phase space which is the same as the original dynamical system in topological sense from a one-dimensional chaotic time series, so that we can simulate the law of time series. We analyze all aspects of chaotic time series based on phase space reconstruction, so phase space reconstruction is the key to the study of chaotic time series. Next, we will discuss how to determine the delay time and embedding dimension.

The key to the choice of delay time τ is to make $x(n)$ and $x(n+\tau)$ independent, but it can't make it completely irrelevant in statistics. The main methods to determine the delay time are autocorrelation function method and mutual information method. What we mainly describe below is the autocorrelation function method, because we will use this method later.

Autocorrelation function method mainly investigates the linear correlation between the difference between the observed quantity $x(n)$ and the average observed quantity and the difference between $x(n+\tau)$ and the average observed quantity. Its definition is expressed mathematically as follows:

$$C(\tau)=\frac{1/N\sum_{n=1}^{N}(x(x+\tau)-\bar{x})(x(n)-\bar{x})}{1/N\sum_{n=1}^{N}(x(n)-x)^2} \tag{4.2}$$

Among them, $\bar{x}\frac{1}{N}\sum_{n=1}^{N}x(n)$,the delay time when the correlation function value $C(\tau)$ drops to zero for the first time is generally taken as the best delay time for phase space reconstruction.

As for the embedding dimension m, the following calculation methods are adopted:

(1)On the basis of $\{x(t)\}, t=1,2,\cdots N$, the reconstructed phase space of m_0 with smaller value is given;

(2)In the calculation of correlation function $C(r)=\frac{1}{N^2}\sum_{i,j=1}^{N}\theta(r-|X_i-X_j|)$, $|X_i-X_j|$

represents the distance between phase points X_i and X_j, and $C(r)$ refers to the probability that the distance between two points of attractor in phase space is less than r;

(3)According to the formula $d(m)=\frac{\ln C(r)}{\ln r}$, the estimated value $d(m_0)$ of the correlation dimension corresponding to m_0 is obtained by fitting;

(4)Taking m_0 as the minimum value, increasing the embedding dimension in turn makes $m_1 > m_0$, $m_2 > m_1$...... calculate according to step (2) and step (3), until $d(m)$ no longer increases with m, and then d is the correlation dimension of attractor.

According to Takens theorem, if the value of d doesn't change with the increase of m after m increases to an appropriate dimension, it is proved that there is chaos in this system. Taking the m estimation when the first value of d reaches the stationary level as the embedding dimension, then $m = 2[d]+1$.

Because chaotic motion is sensitive to initial value, we calculate lyapunov index to quantify this phenomenon.Phase space needs to be reconstructed before extracting lyapunov index from univariate time series data. Wolf et al. proposed to determine lyapunov index directly based on phase plane, phase volume and phase trajectory. The main methods are as follows.The phase space also needs to be reconstructed before the lyapunov index is extracted from the time series data of a single variable. Wolf et al. proposed to determine lyapunov index directly based on phase plane, phase volume, phase orbit and other methods. Wolf method is suitable for systems with no noise and highly nonlinear evolution of small vectors in tangent space. The stock market meets these requirements well, so we adopt Wolf method.

The main methods are as follows：

Let the chaotic time series be $x_1, x_2, \ldots, x_n, \ldots$,If the embedding dimension is m and the delay time is τ, then the reconstructed phase space is ：
$Y(t_i)=\left(x(t_i), x(t_i+\tau), \ldots, x(t_i+(m-1)\tau)\right), (i=1,2,\ldots,N)$,The largest Lyapunov index is expressed as $\lambda=\frac{1}{t_M-t_0}\sum_{i=0}^{M}\log_2\frac{L_i'}{L_i}$.

4.1.3 Test of chaotic properties

We tested the chaotic properties of three stocks respectively. In this paper, we will show the chaotic properties of stock 002281.SZ as an example.See Fig 4.2 for the specific test.

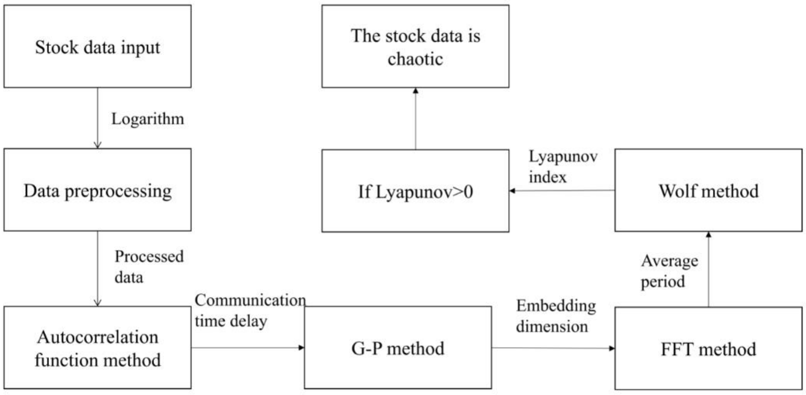


Fig 4.2 Flow chart of chaotic property detection of stock data

First, we process the data, and select the daily opening price of stocks from January 4, 2016 to October 30, 2020, with a total of 1174 sets of data. Secondly, the original time series data $\{x_t\}$ are smoothed by logarithmic difference method, and the difference formula is $Y_t = \ln(x_{t+1}) - \ln(x_t)$, in which the logarithmic Shanghai Composite Index is $\{Y_t\}$. After calculation,we get the logarithmic stock time series diagram, as shown in Fig 4.3.

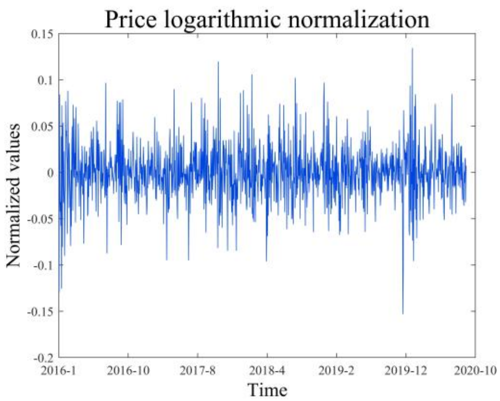


Fig 4.3 Logarithmic stock time series diagram

After obtaining the logarithmic time series data, we began to reconstruct the phase space. Firstly, the autocorrelation function method is used to calculate the delay time of the time series, and the autocorrelation function of the time series delay for 20 periods can be obtained in Matlab, as shown in Fig 4.4.

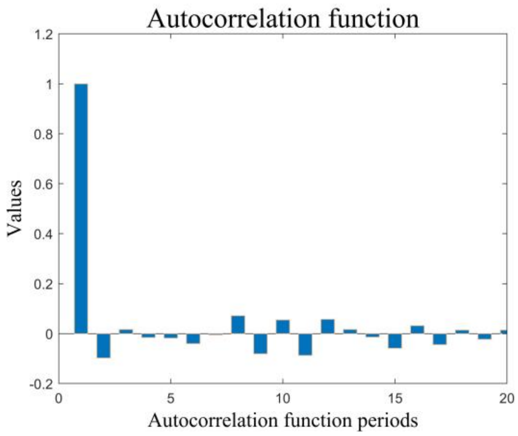


Fig 4.4 Autocorrelation diagram of time series

According to the calculation results, when the second phase is delayed, the value of autocorrelation function is -0.0025 and less than 0. Therefore, according to the definition of autocorrelation function method, we take $\tau=2$ as the best delay time for phase space reconstruction. After getting the best delay time, we continue to calculate the embedding dimension m, which is necessary to reconstruct the phase space. We use the G-P method, that is, d_m . We calculate embedding dimension with minimum embedding dimension 1, maximum embedding dimension 20 and delay time 2 as the standard, and the result is 9, as shown in Fig 4.5.

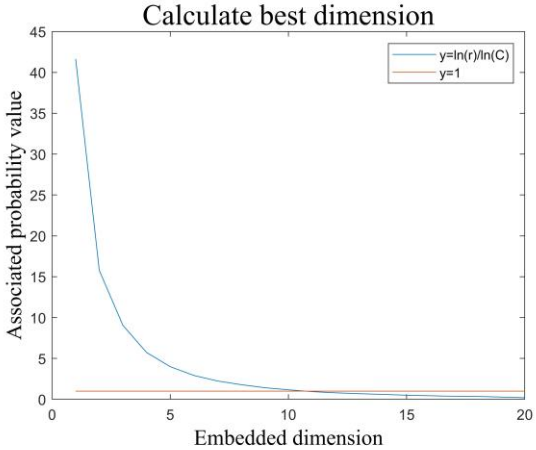


Fig 4.5 G-P method to find embedding dimension

After the embedding dimension m and delay time τ are obtained, in order to calculate lyapunov index, we also need to know the average period P of time series, and the commonly used method is FFT method. The average period of chaotic time series can be transformed from time domain to frequency domain by fast Fourier transform FFT time series, and the average period of original chaotic tim we series can be calculated according to the frequency information of the transformed sequence. The result calculated by FFT method is down to 2. In this way, the parameters needed to calculate lyapunov index are already available. Wolf method is used to calculate lyapunov index. After calculation, we get that the maximum lya-punov index is λ = 0.1260211. The index is greater than zero, which shows that the adjacent points of the nonlinear system where the stock we are looking for will eventually be separated, which leads to the local instability of the corresponding orbit. If the orbit has the overall stability factor, it will fold repeatedly under this action, and finally form a chaotic attractor. The Lyapunov index of the three stocks is shown in Table 4.1.

Table 4.1 Lyapunov index parameters of three stocks

Stock code	Parameter	Time delay	Average period	Embedding dimension	Lyapunov index
000400.SZ	Open	2	2	9	0.102
	High	3	4	11	0.136
	Low	3	2	11	0.106
	Close	2	2	11	0.078
	volume	2	2	3	0.679

	Turn	2	2	3	0.679
002281.SZ	Open	2	2	10	0.126
	High	4	9	10	0.156
	Low	3	3	11	0.085
	Close	2	9	11	0.086
	volume	2	3	2	0.943
	Turn	2	3	2	0.962
600519.SH	Open	2	4	12	0.095
	High	3	4	13	0.124
	Low	3	4	10	0.155
	Close	2	4	11	0.126
	volume	2	2	3	0.418
	Turn	2	2	3	0.418

4.2 Model II Models and Results

4.2.1 Establishment of Hidden Markov Model

HMM is a time series probability model, which can be described as a hidden Markov chains. Observation random sequence is generated by state random sequence . There is a corresponding relationship between observation random sequence and state random sequence. HMM can realize the combination of unobservable state probability and observable state probability, and study the observation sequence and state sequence separately. HMM usually contains three main factors: initial probability distribution, state transition probability distribution and observation probability distribution. Fig 4.6 for the specific test.

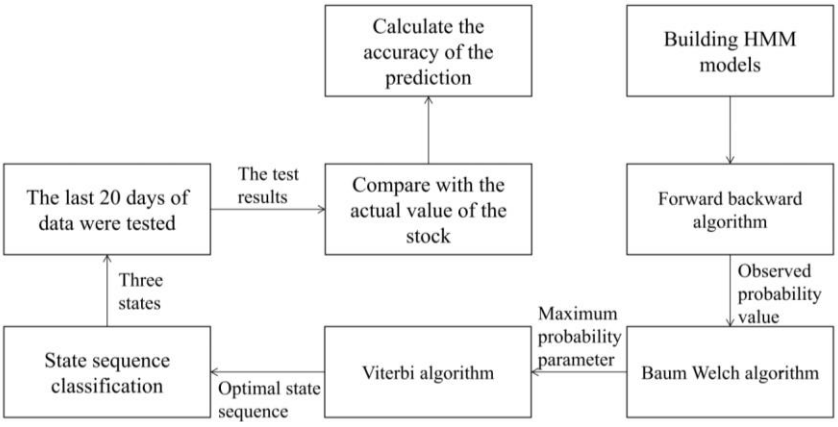


Fig 4.6 The flow chart of HMM stock ups and downs

Three problems of HMM and their solving algorithms:

(a)if the model is given an observation sequence, how to calculate the probability of the observation sequence of the model?

(b)If the observation sequence is given in the model, how can we find that the potential state sequence is the maximum probability value?

(c)If the model gives the space of observation sequence and probability model, how can we find the optimal parameters and maximize the whole probability value?

For the problem (a), the forward-backward algorithm can be used. The forward algorithm defines the size of the forward probability. At time t , the probability of the state q is: $\alpha_t(t) = P\left(o_1, o_2, \ldots, o_t, t_t = q_i \mid \lambda\right)$ According to the recursive forward probability a , the probability value P of the observation sequence is obtained. The process is as follows:

①Initialization: $a_1(t) = \pi_1 b_1\left(o_1\right), \quad 1 \le i \le N$

②Recursion according to time t: $\alpha_{t+1}(i) = \left[\sum_{j=1}^{N} \alpha_t(j) a_{ji}\right] b_i\left(o_{t+1}\right), \quad 1 \le i \le N$

③Termination: $P(O \mid \lambda) = \Sigma_{i=1}^{N} \alpha_T(t)$

The definition probability value of backward algorithm is the observation sequence from time t+1 to T, and the probability of state q at this time is: $\beta_t(t) = P\left(o_{t+1}, o_{t+2}, \ldots, o_T, t_t = q_i \mid \lambda\right)$

According to the recursive backward probability a , the probability value P of the observation sequence is obtained, and the flow is as follows:

①Initialization: $\beta_T(t) = 1, \quad 1 \le t \le N$

②Recursion according to time t: $\beta_t(t) = \Sigma_{j=1}^{N} a_{ij} b_j\left(o_{t+1}\right) \beta_{t+1}(j), \quad 1 \le t \le N$

③ Termination: $P(O \mid \lambda) = \sum_{i=1}^{N} \pi_i b_i\left(o_1\right) \beta_1(i)$ Combining forward and backward probability formulas, P can be unified as:

$$P(O \mid \lambda) = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j\left(o_{t+1}\right) \beta_{t+1}(j), \quad t = 1, 2, \ldots, T-1 \tag{4.3}$$

For the problem (b), Baum-Welch algorithm can be used to learn the parameters of HMM, so as to maximize the probability value of the model in the prediction process. The probability value of its observation sequence can be expressed as: $P(O \mid \lambda) = \Sigma_l P(O \mid l, \lambda) P(I \mid \lambda)$

Using EM algorithm to realize the idea of recursion, the parameter learning problem of HMM is studied. Firstly, two probabilities are defined:

①Given model parameters and observation sequence, the probability of being in state q at time t is recorded as: $\gamma_t(t) = P\left(t_t = q_i \mid 0, \lambda\right) = \frac{P\left(i_t = q_i, O \mid \lambda\right)}{p(o \mid \lambda)}$, according to the definition of forward and backward probabilities, we can get: $\alpha_t(t) \beta_t(t) = P\left(t_t = q_i, O \mid \lambda\right)$, and $\gamma_t(t) = \frac{a_t(i) \beta_t(i)}{P(o \mid \lambda)} = \frac{a_t(i) \beta_t(0)}{\sum_{j=1}^{N} a_t(j) \beta_t(j)}$ can be obtained by analyzing the previous formula

②Given model parameters and observation sequence, the probability of being in state q at time t is recorded as: $\xi_t(t, j) = P\left(t_t = q_i, t_{t+1} = q_j \mid O, \lambda\right)$.

According to the definition of forward and backward probability, we can get:

$$\xi_t(i, j) = \frac{\alpha_t(\theta) a_{ij} b_j\left(o_{t+1}\right) \beta_{t+1}(j)}{\sum_{i=1}^{N} \Sigma_{j=1}^{N} a_t(i) a_{ij} b_j\left(o_{t+2}\right) \beta_{t+2}(j)} \tag{4.4}$$

Generally speaking, the Baum-Welch algorithm can basically realize parameter estimation:

① $\pi_i = \gamma_1(t)$

②

$$a_{ij} = \frac{\sum_{i=1}^{T-1} \xi_t(i, f)}{\sum_{i=1}^{2-1} \gamma_i(i)} \tag{4.5}$$

$$b_j(k) = \frac{\sum_{t=1,0t=w_k}^{T} \gamma_t(j)}{\sum_{i=1}^{T} \gamma_t(j)} \tag{4.6}$$

Through the above derivation, HMM can be established. However, in general, it is better to assume that the observation probability conforms to Gaussian mixture normal distribution, and its density function is:

$$b_j(k) = P\left(o_t = v \mid t_t = q_j\right) = \sum_{m=1}^{M} \omega_{jm} N\left(o_t, \mu_{jm}, \Sigma_{jm}\right) \tag{4.7}$$

At this time, the sample space of the observed variables is continuous. By establishing a continuous HMM, the observation probability matrix and Gaussian mixture model parameters are constructed.

As to question (c), Viterbi algorithm can be used. Define the maximum probability of all single paths at time t :

$$\delta_t(t) = \max_{i_2, f_2, \ldots = t_{t-2}} P\left(t_t = i, t_{t-1}, \ldots, t_1, o_t, \ldots, o_1 \mid \lambda\right), 1 \le t \le N \tag{4.8}$$

At time t , the path with the maximum probability of all single paths, the t-1th node is expressed as: $\psi_e(t) = \arg\max_{\mathrm{i} \le j \sin}\left[\delta_{t-1}(j) a_{ji}\right], 1 \le i \le N$. The main solving steps are:

①Initialization: $\delta_1(t) = \pi_i b_i\left(o_1\right), 1 \le t \le N$
$\psi_1(t) = 0, 1 \le t \le N$

②Recursion according to time t: $\delta_t(t) = \max_{1 sj \sin}\left[\delta_{t-1}(j) a_{ji}\right] b_i\left(o_t\right), \quad 1 \le i \le N$
$\psi_t(t) = \arg\max_{1 \le j \le N}\left[\delta_{t-1}(j) a_{ji}\right], 1 \le t \le N$

③Termination: $P^* = \max_{1 \le i \le N} \delta_T(i)$
$i_T^* = \arg\max_{1 \le i \le N}\left[\delta_T(i)\right]$

④Optimal path backtracking: $t_t^* = \psi_{t+1}\left(t_{t+1}^*\right)$

The optimal path is obtained by way of path backtracking, and the state sequence at this time is as follows: $I^* = \left(t_1^*, t_2^*, \ldots, t_\tau^*\right)$

4.2.2 Hidden Markov Model results

Combining the principle of HMM with three stocks, we can predict the rise and fall of stocks. According to the characteristics of stock prices, when stocks continue to rise or fall, there will be clear fundamental data correspondence.Stock 002281.SZ is used to train the state transition parameters of the model, and the parameters needed by HMM are obtained. According to the trained model, the data of the last 20 days are used for back testing, and the prediction accuracy of the model is analyzed. Finally, according to the results of model accuracy, the problems of HMM are analyzed.

4.2.2.1 Feature selection

In the aspect of feature selection, we mainly analyze the influence of features on the whole model according to the contribution degree of current factors. Firstly, the period of warehouse adjustment is set to 20 trading days, and HMM also uses 20 trading days to predict the whole ups and downs in the prediction process. After the closing of the trading day, it is predicted whether the stock price will rise and fall after 20 trading days. The data of three stocks provided by the topic are selected, and the selection of observed values is mainly based on six observed values (opening price, highest price, lowest price, closing price, trading volume and turnover rate).We trained the test set from January 4, 2016 to October 30, 2020, and determined the whole parameters by using grid search algorithm, and finally selected the probability parameter combination of test and sample evaluation:

①Number of hidden states: N=3.②Observation sequence length: Q=10.③The number of distribution models of Gaussian mixture model: M=2.

4.2.2.2 HMM training process

According to the hypothesis of this paper, it is found that there are corresponding models for the rise and fall of stocks, which can indicate whether stocks will rise or not. For the three stocks provided by the topic, if the observed probability value of stocks is large in the rising model, it shows that the development trend of this stock is great. Therefore, the prediction of stock rise and fall is realized by observing series prediction.

4.2.2.3 Strategy design and results

By observing the data of three stocks, we can get the ups and downs of stocks, so we can use two state sets to realize the classified prediction of stock states. Fig 4.7 shows stock 002281.SZ status classification results:

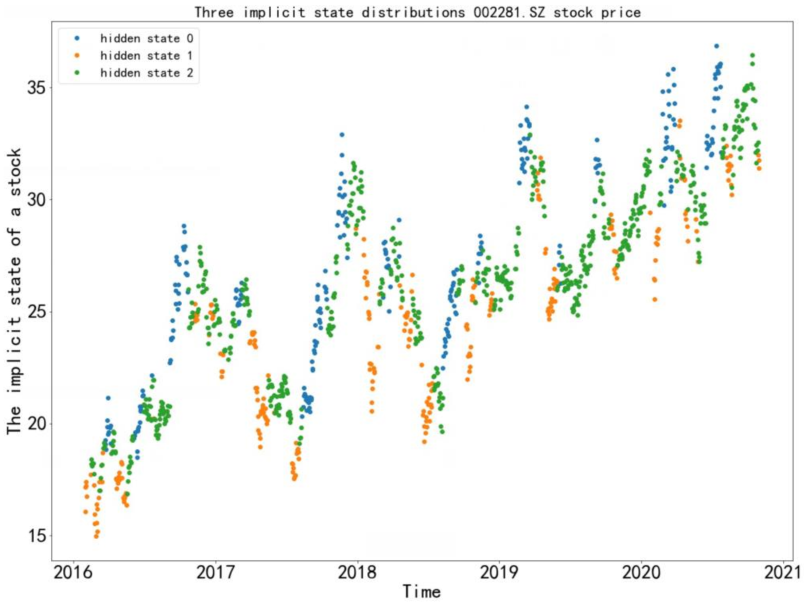


Fig 4.7 State classification results of stock 002281.SZ

Based on the principle of HMM and the data of stock 002281.SZ, this paper analyzes the trend of stock ups and downs. Think about the main reasons of the stock's rising from a macro point of view, build a selection strategy model by quantitative means, and analyze the state of the stock according to the current strategy model. Fig 4.8 shows three hidden states of stock 002281.SZ.

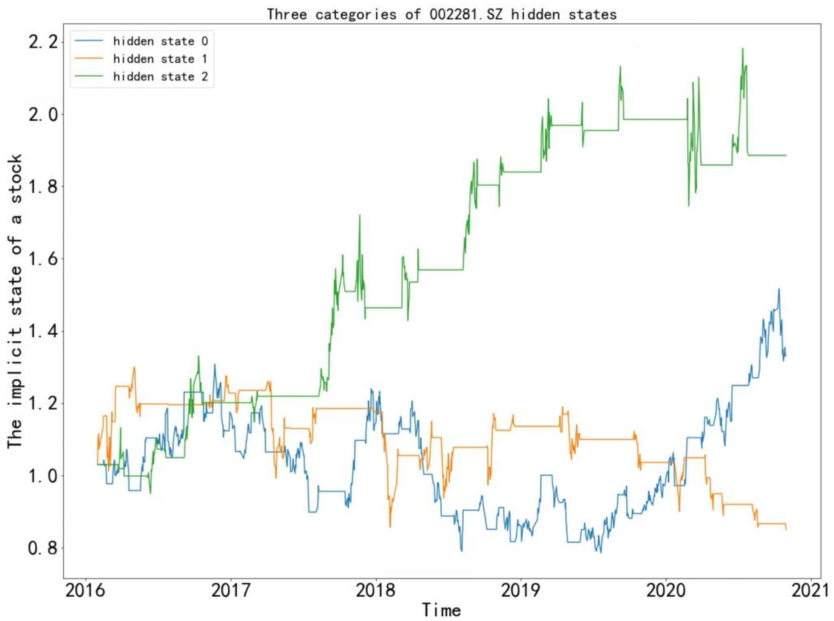


Fig 4.8 Three hidden states of stock 002281.SZ

According to the three states trained by HMM, it can be found that when the number of states is 2, the stock price is in a rising state, while when the number of states is 0 and 1, it is estimated to be in a fluctuating state, especially when the stock price is in a 1 state, the probability of falling is greater. Table 4.2 shows the observed values obtained from the training of HMM. According to the observed values, we can judge the specific days when the final prediction is accurate.

The hidden Markov model predicted results

Date	000400.SZ	Accurately predict	002281.SZ	Accurately predict	600519.SH	Accurately predict
2020-09-25	3	1	2.18	1	3.16	0
2020-09-28	3.02	1	2.18	1	3.16	0
2020-09-29	2.97	0	2.18	0	3.16	1
2020-09-30	2.99	1	2.18	1	3.16	0
2020-10-09	2.84	0	2.18	0	3.16	0
2020-10-12	2.97	0	2.18	0	3.16	0
2020-10-13	2.93	1	2.18	1	3.16	1
2020-10-14	2.95	1	2.18	1	3.16	1
2020-10-15	2.91	1	2.18	1	3.16	1
2020-10-16	2.89	1	2.18	1	3.16	1
2020-10-19	2.89	0	2.18	0	3.16	1
2020-10-20	2.91	1	2.18	1	3.16	0
2020-10-21	2.81	1	2.18	1	3.16	1
2020-10-22	2.78	0	2.18	0	3.16	0
2020-10-23	2.75	1	2.18	1	3.16	1
2020-10-26	2.82	0	2.18	0	3.16	1
2020-10-27	2.8	0	2.18	0	3.16	1
2020-10-28	2.82	1	2.18	1	3.16	0
2020-10-29	2.86	1	2.19	1	3.16	0
2020-10-30	2.77	1	2.22	1	3.16	1

It can be found from Fig 4.9 that according to the correct days predicted by HMM, the accurate days predicted by different stocks are relatively concentrated, and all three stocks are predicted correctly from the 7th to 10th day, which shows that the predicted values of HMM have continuity and can make long-term prediction for stocks. Looking back at the long-term prediction results of HMM in recent 20 days, we can see that they all conform to the long-term trend of stocks, which shows the effectiveness of HMM prediction.

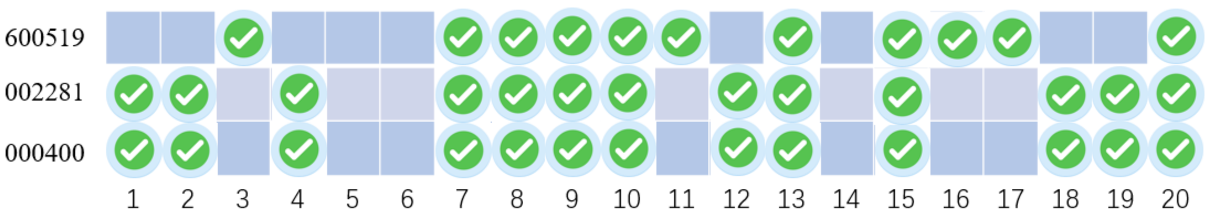


Fig 4.9 HMM Prediction Accuracy Diagram

According to HMM, the prediction results of each stock are shown in Table 4.3. It is found that HMM has different applicability in the prediction of different stocks, and the prediction accuracy of Optical Communication Technology (stock code: 002281.SZ) is the highest, which is 65%.

Table 4.3 Three stocks predictive accuracy

Stock name	Stock code	Accuracy
XJ electric	000400.SZ	65%
Accelink Technology	002281.SZ	60%
Kweichow Moutai	600519.SH	55%

4.3 Model III Models and Results

4.3.1 Data state classification processing

According to the data set of three stocks provided by the topic, because there are many characteristics of stock data in this paper, and because the influence of stock characteristics such as technical indicators on stock fluctuation trend is not clear, this paper only uses 6-dimensional stock basic trading data including opening price, highest price, closing price, lowest price, trading volume and turnover rate. The sample information of stock 002281.SZ data set is shown in Table 4.4.

Table 4.4 Stock fundamentals data

Date	Opening price/yuan	Maximum price/yuan	Closing price/yuan	Lowest price/yuan	Trading volume/hand	Turnover rate/%
2020-09-08	33.68	33.92	33.3	33.73	8686179	1.32
2020-09-09	33.29	33.74	32.7	33.01	11690977	1.78
2020-09-10	33.35	33.92	31.94	32.24	13601989	2.07

This paper intends to predict the rise and fall of its closing price according to the stock data training model.However, if the stock is predicted by two categories only according to the price increase and decrease, the forecast result contains more predicted values with smaller price increase and decrease, which is of little practical significance for investment guidance.Therefore, in order to make the forecast more suitable for the actual demand, this paper further digs the characteristics of stock data, makes sample statistics on the distribution of price increases and decreases, and determines the specific classification of price increases and decreases according to the statistical results.The Fig 4.10 show the distribution of the rise and fall of the overall stock data of the three stocks. The distribution of the price rise and fall of the three stocks is close to the Gaussian distribution, which conforms to the general distribution law of things.

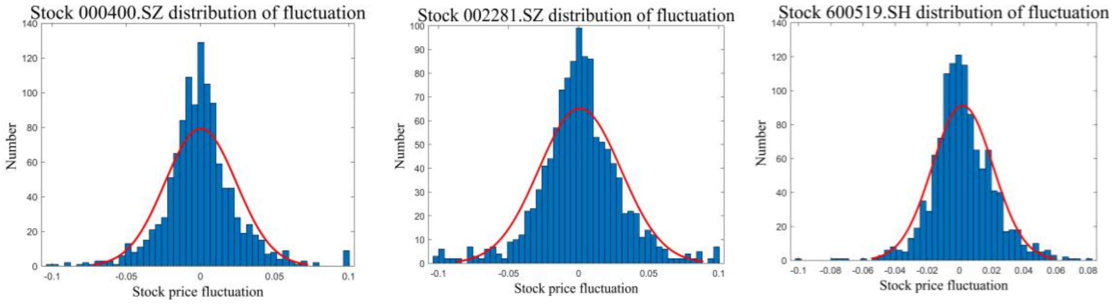


Fig 4.10 Test of normality of three stocks

We divide the trend of stock ups and downs into three categories based on the limit of

1%, and get the histogram shown in Fig 4.11. In the histogram of the three stock ups and downs, the proportion of each distribution interval is basically balanced, and the stock samples belonging to a small range of ups and downs (-1%~1%) are slightly more than the other two categories. In the multi-classification task, when the samples of each category are evenly distributed, the prediction effect is usually better. Therefore, in this paper, we choose the above-mentioned 3-category forecasting method to forecast stock ups and downs, so that we can distinguish small-scale ups and downs, big rises and big falls in the forecasting results, which is of more practical significance.

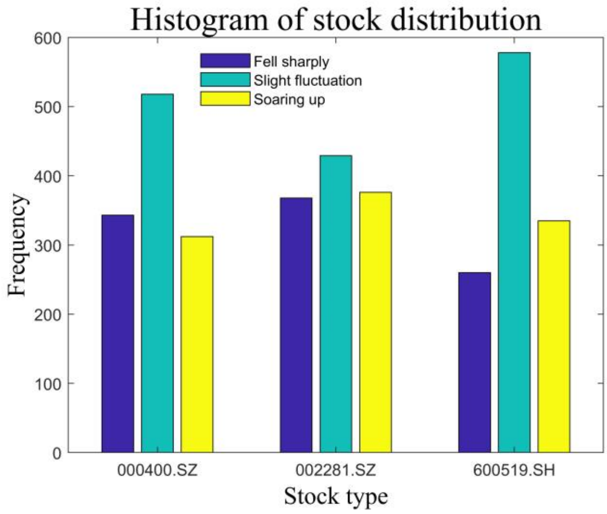


Fig 4.11 Histogram of 3 categories of three stocks

From the stock characteristic information in Table 4.4, we can see that stock data with different characteristics are quite different in magnitude, and the price characteristic values such as opening price, highest price, closing price and lowest price are consistent. The trading volume data is much larger in magnitude than other data types, and the change range of price fluctuation data is -10~10. Therefore, before the data set enters the model training, it is necessary to normalize the data to reduce the interference caused by the characteristic dimension difference. This paper uses zero-mean normalization to eliminate it.Mark the original data as X , $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of the original data respectively. The normalization formula is as follows:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{4.9}$$

In this paper, we use the classification method to predict the rise and fall of stocks. Mark the price series of a certain T consecutive trading day as P1,P2... ,PT. Based on the stock data of the first T days, this paper forecasts the rise and fall of the stock on the $T+1$ day . The definition of the rise and fall uses the previous analysis results, that is, taking the rise and fall of 1% as the limit, the rise and fall of stocks can be classified into three categories: a rise of more than 1% is a enormous rise; A drop of more than 1% is a huge drop; The rest of the stock data are in the category of keeping steady. Mark the range of stock price rise and fall on the $T+1$ day is y_{T+1}, and the quantitative value of the corresponding interval is Y_{T+1} , then the specific quantitative way of rise and fall is as follows:

$$Y_{T+1}=\begin{cases}0, & y_{T+1}>1\% \\ 1, & -1\%<y_{T+1}<1\% \\ 2, & y_{T+1}<-1\%\end{cases} \tag{4.10}$$

4.3.2 Establishment of elastic feedback algorithm

Elastic algorithm is an improved gradient algorithm based on the classical BP algorithm, which can solve the problems of slow training speed and easy to fall into local minima in the classical BP algorithm. A brief description is as follows:

(1)establish a BP neural network, determine an input layer, a hidden layer and an output layer, initialize that weights of the neural network, and initialize the initial adjustment size of each weight adjustment amount. Generally, it is set to a random small value, such as a random value between (0, 1). The initialization weight adjusts the direction of the last time, and +1 or -1 can be randomly selected for initialization.

(2)Determine training samples and test sample sets for training process evaluation.

(3)Input training samples, calculate hidden layer output, and calculate output layer output by using hidden layer output.

(4)Calculate the error between the training output and the real output, and calculate the derivative of each weight.

(5)Get the direction of this weight derivative and compare it with the previous direction. If the directions are the same, enlarge the weight adjustment amount, otherwise, reduce the weight adjustment amount.

(6)Input the test sample, calculate the test sample output and test error, and quit the training when the test error meets the requirements. Otherwise, repeat the training from (3).

According to the prediction results of the elastic feedback algorithm and combined with the predicted states of the Hidden Markov model, the final prediction results are obtained. The main flow Fig 4.12 is as follows.

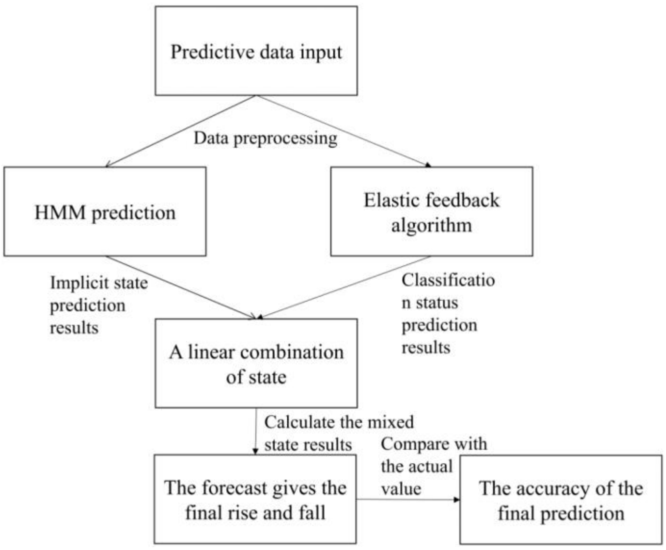


Fig 4.12 Stock rise and fall forecast flow chart

4.3.3 Testing of data

Due to the fluctuation of stock data, the hidden Markov model can only predict the overall trend in the process of prediction, but it fails to predict the daily fluctuation with high accuracy. Therefore, the elastic algorithm is used to modify the hidden Markov model, and the data of the last 20 days is also used to predict the state of ups and downs. Found that can improve the accuracy of model prediction, the stock 002281.SZ, on the basis of combination forecasting, the accuracy of forecasting significantly increased by 15%, the accuracy reached 75%. As can be seen from Table 4.5, according to the trend that stocks will always rise when they fall and combined with the results of prediction classification, both of the two classification methods have their own characteristics. The elastic feedback algorithm can have good prediction effect in the short-term prediction based on the periodicity of stocks.

Note:The first column of table 4.5 is the up and down state of stock data, the second column is the state predicted by HMM, the third column is the state predicted by elastic feedback algorithm, and the fourth column is the specific number of days with accurate prediction.

Table 4.5 Three stock hybrid algorithm results

Stock 000400.SZ				**Stock 002281.SZ**				**Stock 600519.SH**			
1.40%	1	1	1	-1.84%	1	-1	1	0.36%	1	0	0
-0.95%	1	0	1	-1.84%	1	-1	1	1.58%	1	1	0
1.98%	1	1	0	4.46%	1	1	1	-0.52%	1	0	1
-0.72%	1	0	1	-0.37%	1	0	1	0.98%	1	0	0
5.08%	0	1	0	0.51%	1	0	0	1.65%	1	1	0
4.56%	0	1	0	3.70%	1	1	0	3.22%	1	1	0
-1.39%	0	-1	1	-1.10%	1	-1	1	-0.66%	1	0	1
0.54%	0	0	1	-2.97%	1	-1	1	-0.58%	1	0	1
-1.13%	0	-1	1	-1.43%	1	-1	1	-0.23%	1	0	1
-0.81%	0	0	1	-3.28%	1	-1	1	-0.75%	1	0	1
0.07%	0	0	0	3.15%	1	1	0	-0.76%	1	0	1
0.48%	0	0	1	-3.52%	1	-1	1	2.06%	1	1	0
-3.24%	0	-1	1	-3.25%	1	-1	1	-0.04%	1	0	1
-1.26%	0	-1	0	0.97%	1	0	0	0.53%	1	0	0
-0.85%	0	0	1	-2.41%	1	-1	1	-1.56%	1	-1	1
2.57%	0	1	1	1.71%	1	1	1	-4.22%	1	-1	1
-0.70%	0	0	0	1.18%	1	1	0	-1.10%	1	-1	1
0.63%	0	0	1	-1.75%	2	-1	1	2.45%	1	1	1
1.18%	0	1	1	-0.78%	2	0	1	0.67%	1	0	0
-3.09%	0	-1	1	-1.07%	2	-1	1	-0.36%	1	0	1

According to Fig 4.13, it is found that the elastic feedback algorithm improves different stock data. From stock types, it can be found that the stocks with stronger periodicity, such as stock 000400.SZ, have a strong volatility, so the accuracy of prediction has been significantly improved. However, for stocks whose periodicity is not obvious, such as stock 600519.SH, the elastic feedback algorithm does not significantly improve the prediction accuracy in the process of calculation, but only increases the prediction accuracy by 5%.

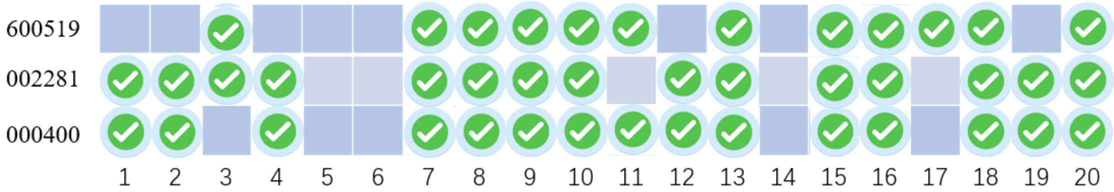


Fig 4.13 Precise diagram of mixed prediction model

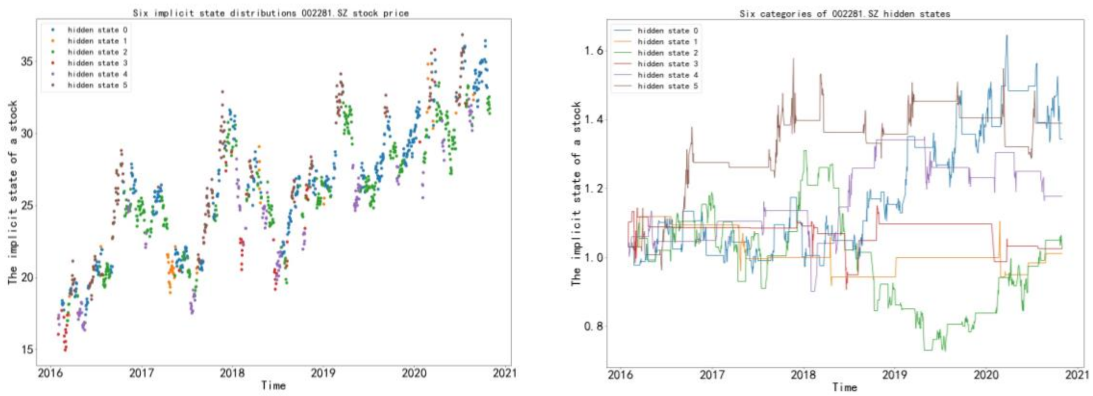
Therefore, the analysis of the different single algorithm can't good to predict stock status, through the combination of advantages and disadvantages of different algorithm, the better the analysis of the trend of the stock go up drop, can be seen from Table 4.6, the comprehensive prediction algorithm for periodic strong stock can better improve the accuracy of prediction, so different prediction model has its own characteristics, in the prediction process, need to complement each other, better predict the stock rise and fall of the state.

Table 4.6 Three stocks predictive accuracy

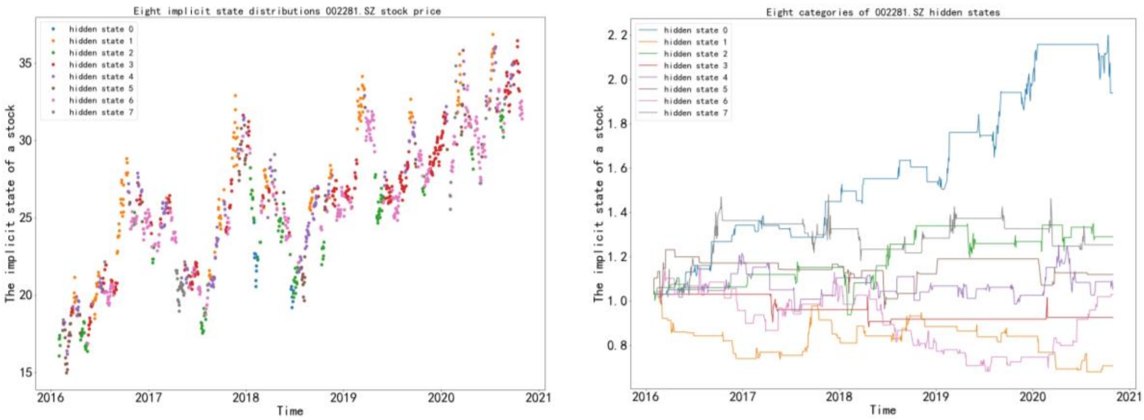
Stock name	Stock code	Accuracy
XJ electric	000400.SZ	75%
Accelink Technology	002281.SZ	75%
Kweichow Moutai	600519.SH	60%

5.Sensitivity Analysis

In the second part of the analysis process, we use the grid search algorithm to determine the whole parameters, and finally choose the probability parameter combination as follows: ① Number of hidden states: N=3. ② Length of observation sequence: Q=10. ③ Distribution model number of Gaussian mixture model: M=2. Therefore, for stock 002281.SZ, we choose the number of hidden states N=8 to analyze the adaptability and sensitivity of the model under different states. The specific analysis results are shown in Fig 4.14 and Table 4.7.



(a) 6 Classification States of Stock 002281.SZ (b) 6 Hidden States of Stock 002281.SZ



(c) 8 Classification States of Stock 002281.SZ (d) 8 Hidden States of Stock 002281.SZ

Fig 4.14 Different hidden states for stock 002281.SZ

Table 4.7 Hidden state result classification

State	Color	Trend
0	Blue	Rose sharply
1	Green	Slight decline
2	Red	Vibrate decline
3	Purple	Slight rise
4	Yellow	Cliff fall
5	Light blue	Fall sharply

It can be seen that when different hidden state numbers are selected, there is no significant difference in the accuracy of stock prediction by using HMM. It can be clearly seen from 6 state classifications and 8 state classifications that with the increase of hidden state numbers, although the prediction accuracy of the model has been improved to some extent, the complexity of analysis has also been improved. Therefore, we think that choosing different hidden state numbers has little effect on the conclusion, and the model is less sensitive to the hidden state numbers, and the results obtained are in line with reality.

6.Strengths and Weakness

6.1 Strengths

(1)In the first part, Matlab is used to describe the trend of stocks, and lyapunov index is used to judge whether stocks are chaotic, which can break the limitations of traditional inspection methods and improve the accuracy of verification.

(2) In the second part, HMM is established to predict the trend and randomness of stock price, and the model is tested with relevant data, which can improve the prediction accuracy.

(3) In the third part, the elastic algorithm is added to the problem 2 to establish a stock price reversal judgment model to predict the rise or fall of stock prices, and the model is tested to further improve the accuracy of prediction.

6.2 Weakness

Because the HMM needs more data to improve the prediction accuracy, and this paper only uses the data of three stocks to train the model, therefore, the advantages of the HMM are not better reflected in the final prediction accuracy.

7.Conclusion

According to the daily, weekly and monthly charts of the three stocks in recent five years, it is found that all the three stocks show a rising trend of nonlinear fluctuations in general. The chaos characteristic of the trading data of the three stocks is tested by Lyapunov index model. The results show that the Lyapunov index λ=0.126021521 of the stock 002281.SZ is greater than 0, and the chaotic characteristics of stocks have been successfully verified.

Based on chaos theory, the HMM is established, and different hidden states of stocks are successfully classified by using forward-backward algorithm, Baum-Welch algorithm and Viterbi algorithm. The results show that there is no significant difference in forecasting accuracy under different classification of hidden states. When the number of hidden states is 3, the prediction accuracy of three stocks is 60%, 65% and 55% respectively.

Finally, on the basis of the HMM established in question 2, the elastic feedback algorithm is adopted, and the prediction of hidden state by HMM is linearly combined with the prediction of classified state by elastic feedback algorithm, and the stock's rising and falling trend is predicted according to the final result. The results show that the forecasting accuracy of the stock price reversal model established in question 3 is 75%, 75% and 60%, respectively.

References

[1]Rachele Dini. Walter Benjamin's The Work Of Art in the Age of Mechanical Reproduction[M].Macat Library:2018-02-21.

[2]Anthony Wallace. The Work of Art in the Age of Mechanical Reproduction[J]. Southern Review,2016,52(3).

[3]Canan Akin,N. Sezin Kipcak. Art in the Age of Digital Reproduction: Reconsidering Benjamin's Aura in "Art of Banks"[J]. Journal of Communication and Computer,2016,13(4).

[4]Jonathan Davis. Questioning "The Work of Art in the Age of Mechanical Reproduction": A Stroll around the Louvre after Reading Benjamin1[J]. Contemporary Aesthetics,2008,6(1).

[5]Wickson. Works of Art in an Age of Mechanical Reproduction: ‘The Glamour of Italian Fashion 1945–2014’ at the Victoria and Albert Museum, London (5 April–27 July 2014)[J]. Italian Studies,2014,69(3).

[6]NICK PEIM. Walter Benjamin in the Age of Digital Reproduction: Aura in Education: A Rereading of ‘The Work of Art in the Age of Mechanical Reproduction’[J]. Journal of Philosophy of Education,2007,41(3).

[7]Croft, John. Markets chaos quashes air taxi firm stock sale[J]. Flight International,2008,173(5131).

[8].Mathematics - Chaos Research; Study Findings from Tsinghua University Provide New Insights into Chaos Research (Research On Application of Fractional Calculus In Signal Analysis and Processing of Stock Market)[J]. Journal of Robotics & Machine Learning,2020.
[9]Johann Tasker. Stock up on inputs to avoid Brexit chaos[J]. Farmers Weekly,2018,170(13).
[10]Aviral Kumar Tiwari,Rangan Gupta. Chaos in G7 stock markets using over one century of data: A note[J]. Research in International Business and Finance,2018.
[11]Karsten Webel. Chaos in German stock returns — New evidence from the 0–1 test[J]. Economics Letters,2012,115(3).
[12]S. Sujeesh Kumar, V. Nandamohan. Dynamics of randomness and efficiency in the Indian stock markets[J]. Int. J. of Financial Markets and Derivatives,2018,6(4).
[13]Salim Lahmiri. On fractality and chaos in Moroccan family business stock returns and volatility[J]. Physica A: Statistical Mechanics and its Applications,2017,473.
[14]Ata Ozkaya. A model of active trading by using the properties of chaos[J]. Digital Signal Processing,2015,39.
[15]Ritesh Kumar Mishra,Sanjay Sehgal,N.R. Bhanumurthy. A search for long-range dependence and chaotic structure in Indian stock market[J]. Review of Financial Economics,2011,20(2).
[16]Athina Bougioukou. Non-Linearity and Chaotic Behaviour in Cyprus Stock Market[J]. International Journal of Financial Management,2015,5(4).
[17]Alper Ozun,Mike P. Hanias,Panayiotis G. Curtis. A chaos analysis for Greek and Turkish equity markets[J]. EuroMed Journal of Business,2010,5(1).
[18]Melvin J. Hinich,Apostolos Serletis. Randomly modulated periodicity in the US stock market[J]. Chaos, Solitons and Fractals,2006,36(3).

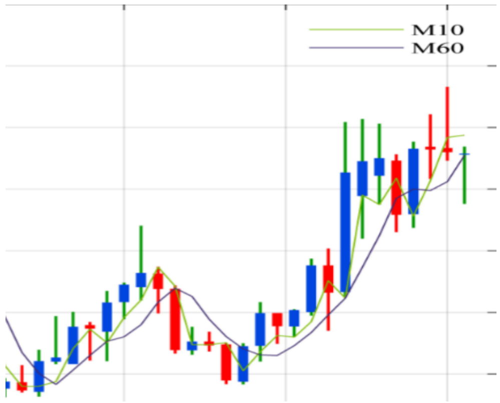
Appendix

1.Question 1. K chart for the other two stocks

(a) Stock 000400.SZ daily trend chart

(b) Stock 000400.SZ weekly trend chart

(c) Stock 000400.SZ monthly trend chart



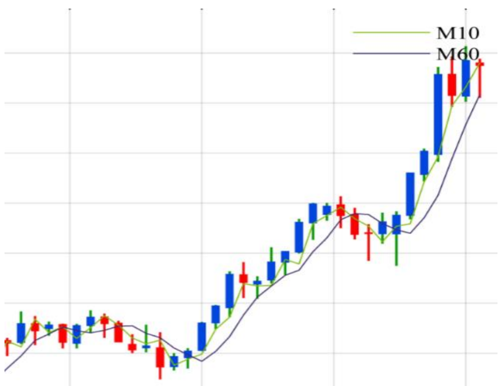
(d) Stock 000400.SZ monthly trend detail chart

Stock 000400.SZ trend analysis chart

(a) Stock 600519.SH daily trend chart

(b) Stock 600519.SH weekly trend chart

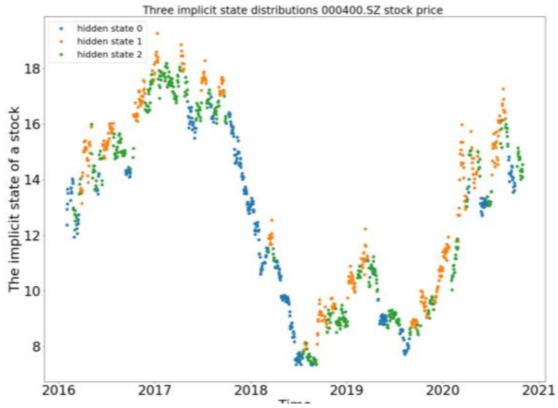
(c) Stock 600519.SH monthly trend chart



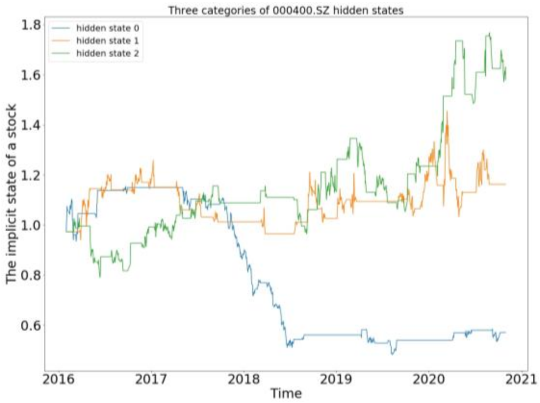
(d) Stock 600519.SH monthly trend detail chart

Stock 600519.SH trend analysis chart

2.Question 2. The hidden state diagram of the other two stocks

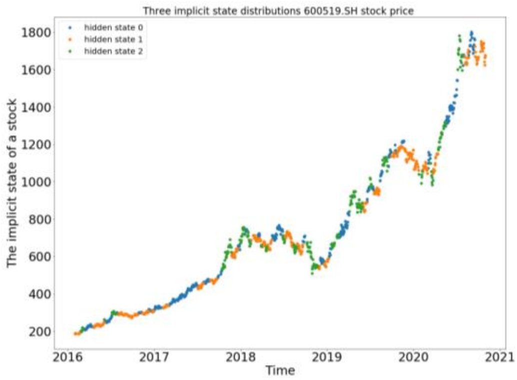


(a) State classification results of stock 000400.SZ

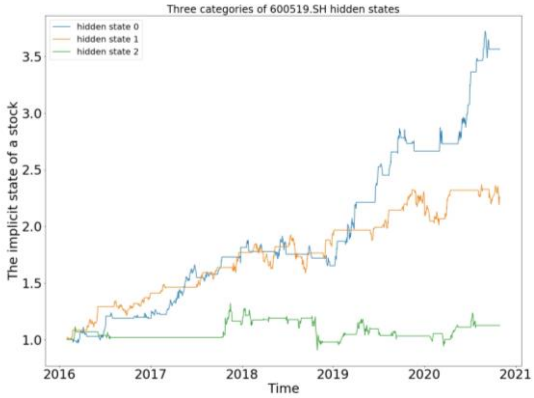


(b) Three hidden states of stock 000400.SZ

Stock 000400.SZ three hidden status diagram



(a) State classification results of stock 600519.SH



(b) Three hidden states of stock 600519.SH

Stock 600519.SH three hidden status diagram

3.Appendix code:

3.1 Question 1 code:

K chart analysis (Only with daily K chart)

```
%% 清空工作区
clc
clear
close all
%% 读取数据
data1 = xlsread('000400.SZ.SZ.xls');
data2 = xlsread('002281.SZ.SZ.xls');
data3 = xlsread('600519.SH.SH.xls');
%% 混沌因子分析
data1=data3;
```

```
%% 箱线图
%箱线图
set(0,'defaultfigurecolor','w')
opengl hardwarebasic
opengl software
Date=[];
i=1;
while i<size(data1,1)
    Date(i,1)=i;
    i=i+1;
end
OpenPrice=data1(:,1); %开盘
HighPrice=data1(:,4); %收盘
LowPrice=data1(:,3); %最低
ClosePrice=data1(:,2); %最高
Vol=data1(:,5); %成交量
% save Data Date OpenPrice HighPrice LowPrice ClosePrice Vol; %存储 mat 文件 方便下次使用
candle(HighPrice,LowPrice,ClosePrice,OpenPrice,'r')%高 低 收 开 红色 时间 时间格式
hold on
h1=plot(data1(:,7),'Color',[124/255,187/255,0/255]);
h2=plot(data1(:,8),'Color',[62/255,43/255,109/255]);
h=legend([h1,h2],'M10','M60');
legend('boxoff')
dates = { '2016-1' '2016-10' '2017-8' '2018-4' '2019-2' '2019-12' '2020-10' }';
set(gca,'xtick',0:200:1200) ;%y 轴的增长幅度
set(gca,'xticklabel',dates(1:7));
set(gca,'fontname','Times New Roman','FontSize',10);
xlabel('Time','fontname','Times New Roman','Color','k','FontSize',15)
ylabel('Price','fontname','Times New Roman','Color','k','FontSize',15)
title('Stock 600519.SH.SH prices change','fontname','Times New Roman','Color','k','FontSize',20);%图形的标题
print('3-year','-djpeg','-r1000')
```

Lyapunov index calculation method

1.mian.m

```
%% 清空工作区
clc
clear
close all
%% 读取数据
data1 = xlsread('000400.SZ.SZ.xls');
data2 = xlsread('002281.SZ.SZ.xls');
```

```
data3 = xlsread('600519.SH.SH.xls');
%% 混沌因子分析
% 第一步对数据进行处理，对数平滑化
set(0,'defaultfigurecolor','w')
X1=data2(:,1);
for i =1:size(X1,1)-1
    Y1(i,1)=log(X1(i+1))-log(X1(i));
end
figure(1)
plot(Y1,'Color',[0/255,70/255,222/255])
dates = { '2016-1' '2016-10' '2017-8' '2018-4' '2019-2' '2019-12' '2020-10' }';
set(gca,'xtick',0:200:1200) ;%y 轴的增长幅度
set(gca,'xticklabel',dates(1:7));
set(gca,'fontname','Times New Roman','FontSize',10);
xlabel('Time','fontname','Times New Roman','Color','k','FontSize',15)
ylabel('Normalized values','fontname','Times New Roman','Color','k','FontSize',15)
title('Price logarithmic normalization','fontname','Times New Roman','Color','k','FontSize',20);%图形的标题
print('first2','-djpeg','-r500')
%%
% 第二步对根据自相关函数，求时延参数
%使用 autocorr 函数
figure(2)
[ACF,lags,bounds] = autocorr(Y1,20) ;
plot(lags(1:end),ACF(1:end));
title('autocorr 求自相关') ;
bar(ACF,0.6,'FaceColor',[0,0.45,0.74],'EdgeColor',[237/255,177/255,131/255])
axis([0 20 -0.2 1.2])
xlabel('Autocorrelation function periods','fontname','Times New Roman','Color','k','FontSize',15)
ylabel('Values','fontname','Times New Roman','Color','k','FontSize',15)
title('Autocorrelation function','fontname','Times New Roman','Color','k','FontSize',20);%图形的标题
print('Second','-djpeg','-r500')
%根据上述计算的成果，可以知道时延参数为 2，第二期对应的数值为-0.0566
%% 第三步计算平均周期，利用 FFT 方法
T1 = period_mean_fft(Y1);
%% 第四步计算嵌入维数，利用 G-P 方法,计算得出维数为 11
[ln_r,ln_C]=G_P(Y1(1:50,1),50,2,1,20,20);
figure(3)
for i=1:20
    ln_chu(i,1)=ln_r(1,i)/ln_C(1,i);
end
figure(4)
```

```
p1=plot(ln_chu(:,1));
hold on;
p2=plot(ones(20,1));
legend([p1,p2],'y=ln(r)/ln(C)','y=1')
xlabel('Embedded dimension','fontname','Times New Roman','Color','k','FontSize',15)

ylabel('Associated probability value','fontname','Times New Roman','Color','k','FontSize',15)
title('Calculate best dimension','fontname','Times New Roman','Color','k','FontSize',20);%图形的标题
print('Third','-djpeg','-r500')
lambda_1=lyapunov_wolf(Y1,size(Y1,1),11,2,4);
```

2.fx: period_mean_fft.m

```
function T_mean=period_mean_fft(data)
%该函数使用快速傅里叶变换 FFT 计算序列平均周期
%data：时间序列
%T_mean：返回快速傅里叶变换 FFT 计算出的序列平均周期
Y = fft(data); %快速 FFT 变换
N = length(Y); %FFT 变换后数据长度
Y(1) = [];%去掉 Y 的第一个数据，它是 data 所有数据的和
power = abs(Y(1:N/2)).^2;%求功率谱
nyquist = 1/2;
freq = (1:N/2)/(N/2)*nyquist; %求频率
period = 1./freq; %计算周期
[mp,index] = max(power);%求最高谱线所对应的下标
T_mean=period(index);%由下标求出平均周期
```

3.fx: reconstitution.m

```
function X=reconstitution(data,N,m,tau)
%该函数用来重构相空间
% m 为嵌入空间维数
% tau 为时间延迟
% data 为输入时间序列
% N 为时间序列长度
% X 为输出,是 m*n 维矩阵
M=N-(m-1)*tau;%相空间中点的个数
for j=1:M          %相空间重构
    for i=1:m
       X(i,j)=data((i-1)*tau+j);
    end
end
```

4.fx: correlation_integral.m

```
%程序中的 correlation_integral 函数如下:
function C_I=correlation_integral(X,M,r)
%the function is used to calculate correlation integral
%C_I:the value of the correlation integral
%X:the reconstituted state space,M is a m*M matrix
%m:the embedding demention
%M:M is the number of embedded points in m-dimensional sapce
%r:the radius of the Heaviside function,sigma/2<r<2sigma
%calculate the sum of all the values of Heaviside
%skyhawk
sum_H=0;
for i=1:M
%     fprintf('%d/%d\n',i,M);
    for j=i+1:M
       d=norm((X(:,i)-X(:,j)),inf);
       sita=heaviside(d-r);%calculate the value of the heaviside function
       sum_H=sum_H+sita;
    end
end
C_I=2*sum_H/(M*(M-1));%the value of correlation integral
```

5.fx: G_P.m

```
function [ln_r,ln_C]=G_P(data,N,tau,min_m,max_m,ss)
% the function is used to calculate correlation dimention with G-P algorithm  %计算关联维数的 G—P 算法
for m=min_m:max_m
    Y=reconstitution(data,N,m,tau);
    M=N-(m-1)*tau;
    for i=1:M-1
        for j=i+1:M
            d(i,j)=max(abs(Y(:,i)-Y(:,j)));
        end
    end
    max_d=max(max(d));%the max distance of all points
    d(1,1)=max_d;
    min_d=min(min(d));
    delt=(max_d-min_d)/ss;
    for k=1:ss
        r=min_d+k*delt;
        C(k)=correlation_integral(Y,M,r);%calculate the correlation
        ln_C(m,k)=log(C(k));%lnC(r)
```

```
        ln_r(m,k)=log(r);%lnr
    end
end
```

6. lyapunov_wolf.m

```
function lambda_1=lyapunov_wolf(data,N,m,tau,P)
%  该函数用来计算时间序列的最大 Lyapunov 指数--Wolf 方法
min_point=1  ; %&&要求最少搜索到的点数
MAX_CISHU=5 ;  %&&最大增加搜索范围次数
%FLYINGHAWK
%   求最大、最小和平均相点距离
    max_d = 0;
    min_d = 1.0e+100;                        %最小相点距离
    avg_dd = 0;
   Y=reconstitution(data,N,m,tau);                %相空间重构
   M=N-(m-1)*tau;
    for i = 1 : (M-1)
       for j = i+1 : M
          d = 0;
          for k = 1 : m
             d = d + (Y(k,i)-Y(k,j))*(Y(k,i)-Y(k,j));
          end
          d = sqrt(d);
          if max_d < d
            max_d = d;
          end
          if min_d > d
            min_d = d;
          end
          avg_dd = avg_dd + d;
       end
    end
    avg_d = 2*avg_dd/(M*(M-1));
    dlt_eps = (avg_d - min_d) * 0.02 ;
    min_eps = min_d + dlt_eps / 2 ;
    max_eps = min_d + 2 * dlt_eps  ;
    DK = 1.0e+100;
    Loc_DK = 2;
    for i = (P+1):(M-1)
       d = 0;
       for k = 1 : m
          d = d + (Y(k,i)-Y(k,1))*(Y(k,i)-Y(k,1));
       end
```

```
   d = sqrt(d);
   if (d < DK) && (d > min_eps)
     DK = d;
     Loc_DK = i;
   end
end
sum_lmd = 0 ;                        % 存放前 i 个 log2（DK1/DK）的累计和
for i = 2 : (M-1)                     % 计算演化距离
   DK1 = 0;
   for k = 1 : m
      DK1 = DK1 + (Y(k,i)-Y(k,Loc_DK+1))*(Y(k,i)-Y(k,Loc_DK+1));
   end
   DK1 = sqrt(DK1);
   old_Loc_DK = Loc_DK ;              % 保存原最近位置相点
   old_DK=DK;
   if (DK1 ~= 0)&&( DK ~= 0)
     sum_lmd = sum_lmd + log(DK1/DK) /log(2);
   end
   lmd(i-1) = sum_lmd/(i-1);
   point_num = 0  ; % &&在指定距离范围内找到的候选相点的个数
   cos_sita = 0  ; %&&夹角余弦的比较初值 ——要求一定是锐角
   zjfwcs=0    ;%&&增加范围次数
    while (point_num == 0)
      for j = 1 : (M-1)
         if abs(j-i) <=(P-1)    %&&候选点距当前点太近，跳过！
          continue;
         end
         dnew = 0;
         for k = 1 : m
           dnew = dnew + (Y(k,i)-Y(k,j))*(Y(k,i)-Y(k,j));
         end
         dnew = sqrt(dnew);

         if (dnew < min_eps)||( dnew > max_eps )  %
         continue;
         end

        %*计算夹角余弦及比较
         DOT = 0;
         for k = 1 : m
           DOT = DOT+(Y(k,i)-Y(k,j))*(Y(k,i)-Y(k,old_Loc_DK+1));
         end
         CTH = DOT/(dnew*DK1);

```

```
            if acos(CTH) > (3.14151926/4)    %&&不是小于 45 度的角，跳过！
            continue;
            end

            if CTH > cos_sita  %&&新夹角小于过去已找到的相点的夹角，保留
              cos_sita = CTH;
              Loc_DK = j;
              DK = dnew;
            end
            point_num = point_num +1;

        end

         if point_num <= min_point
           max_eps = max_eps + dlt_eps;
           zjfwcs =zjfwcs +1;
           if zjfwcs > MAX_CISHU
               DK = 1.0e+100;
               for ii = 1 : (M-1)
                 if abs(i-ii) <= (P-1)
                 continue;
                 end
                 d = 0;
                 for k = 1 : m
                   d = d + (Y(k,i)-Y(k,ii))*(Y(k,i)-Y(k,ii));
                 end
                 d = sqrt(d);

                 if (d < DK) && (d > min_eps)
                   DK = d;
                   Loc_DK = ii;
                 end
               end
             break;
           end
           point_num = 0        ;
           cos_sita = 0;
         end
       end
   end
%取平均得到最大李雅普诺夫指数
lambda_1=sum(lmd)/length(lmd);
```

3.2 Question 2 code:

The hidden Markov model predicts the status of stock 002281.SZ.SZ

```
from hmmlearn import hmm
import numpy as np
from matplotlib import pyplot as plt
import pandas as pd
n = 2 # 6 个隐藏状态

data = pd.read_csv('6.csv', index_col=0)
volume = data['volume']
close = data['close']

logDel = np.log(np.array(data['high'])) - np.log(np.array(data['low']))
logRet_1 = np.array(np.diff(np.log(close)))
logRet_5 = np.log(np.array(close[20:])) - np.log(np.array(close[:-20]))
logVol_5 = np.log(np.array(volume[20:])) - np.log(np.array(volume[:-20]))

# 保持所有的数据长度相同
logDel = logDel[20:]
logRet_1 = logRet_1[19:]
close = close[20:]

Date = pd.to_datetime(data.index[20:])
A = np.column_stack([logDel, logRet_5, logVol_5])
#print(A)
model = hmm.GaussianHMM(n_components=n, covariance_type="full", n_iter=2000).fit(A)
hidden_states = model.predict(A)

fig1=plt.figure(figsize=(20, 15))
ax = plt.subplot(111)
ax.set_title(..., fontsize=30)
ax.set_xlabel(..., fontsize=30)
ax.set_ylabel(..., fontsize=30)          # 设置轴标题字体大小
for i in range(n):
    pos = (hidden_states == i)
    plt.plot_date(Date[pos], close[pos], 'o', label='hidden state %d' % i, lw=2)
    plt.xticks(fontsize=30)
    plt.yticks(fontsize=30)                  # 设置坐标标签字体大小
    plt.title('Six implicit state distributions 002281.SZ.SZ stock price ')
    plt.legend(loc='upper left')
    plt.rcParams.update({'font.size': 20})      #设置图例字体大小
    #plt.grid(True)                           #是否要显示网格线
    plt.xlabel('Time')
```

```
    plt.ylabel('The implicit state of a stock')
plt.show()
fig1.savefig("C:/Users/19766/Desktop/third/stocks.jpg")

res = pd.DataFrame({'Date': Date, 'logReg_1': logRet_1, 'state': hidden_states}).set_index('Date')
series = res.logReg_1

templist = []
fig2=plt.figure(figsize=(20, 15))
ax = plt.subplot(111)
ax.set_title(..., fontsize=30)
ax.set_xlabel(..., fontsize=30)
ax.set_ylabel(..., fontsize=30)            # 设置轴标题字体大小
for i in range(n):
    pos = (hidden_states == i)
    pos = np.append(1, pos[:-1])
    res['state_ret%d' % i] = series.multiply(pos)
    data_i = np.exp(res['state_ret%d' % i].cumsum())
    templist.append(data_i[-1])
    plt.plot_date(Date, data_i, '-', label='hidden state %d' % i)
    plt.xticks(fontsize=30)
    plt.yticks(fontsize=30)                     # 设置坐标标签字体大小
    plt.title('Six categories of 002281.SZ.SZ hidden states')
    plt.legend(loc='upper left')
    plt.rcParams.update({'font.size': 20})      #设置图例字体大小
    #plt.grid(True)                              #是否要显示网格线
    plt.xlabel('Time')
    plt.ylabel('The implicit state of a stock')
plt.show()
fig2.savefig("C:/Users/19766/Desktop/third/predict.jpg")


templist = np.array(templist).argsort()
long = (hidden_states == templist[-1]) + (hidden_states == templist[-2])  # 买入
short = (hidden_states == templist[0]) + (hidden_states == templist[1])  # 卖出
long = np.append(0, long[:-1])
short = np.append(0, short[:-1])

plt.figure(figsize=(20, 15))
res['ret'] = series.multiply(long) - series.multiply(short)
plt.plot_date(Date, np.exp(res['ret'].cumsum()), 'r-')
plt.show()
```

3.3 Question 3 code:

Elastic feedback algorithm

```
%% 清空工作区
clc
clear
close all
%% 读取数据
data1 = xlsread('000400.SZ.SZ.xls');
data2 = xlsread('002281.SZ.SZ.xls');
data3 = xlsread('600519.SH.SH.xls');
%% 对股票数据进行处理
% 区分三类指标，大涨，大跌，不涨不跌
X1=data3(1:1173,4);  %前一天收盘价
X2=data3(2:1174,4);  %后一天收盘价
X3=(X2-X1)./X1;   %储存涨跌幅
Y1=[];  %储存状态
% 区分状态
% histfit(X3,50)
%%
a=0;b=0;c=0;
for i=1:size(X3,1)
    %涨
    if X3(i,1)>0.01
        Y1(i,1)=1;
        a=a+1;
    %跌
    elseif X3(i,1)<-0.01
        Y1(i,1)=-1;
        b=b+1;
    else
        Y1(i,1)=0;
        c=c+1;
    end
end
A=[a,b,c];
%% 绘制正态分布图，股票涨跌的幅度
histfit(X3,50)
xlabel('Stock price fluctuation','fontname','Times New Roman','Color','k','FontSize'
,15)
ylabel('Number ','fontname','Times New Roman','Color','k','FontSize',15)
title('Stock 600519.SH.SH distribution of fluctuation','fontname','Times New Roman',
'Color','k','FontSize',20);%图形的标题
print('3-zheng','-djpeg','-r500')
```

```
%% 画直方图
%初始化数据
A=[343  518 312;
368 429 376;
260 578 335;
];
X = categorical({'000400.SZ.SZ','002281.SZ.SZ','600519.SH.SH'});
X = reordercats(X,{'000400.SZ.SZ','002281.SZ.SZ','600519.SH.SH'});
b = bar(X,A,'FaceColor','flat');
for k = 1:size(A,2)
    b(k).CData = k;
end
xlabel('Stock type','fontname','Times New Roman','Color','k','FontSize',15)
ylabel('Frequency','fontname','Times New Roman','Color','k','FontSize',15)
title('Histogram of stock distribution','fontname','Times New Roman','Color','k','FontSize',20);%图形的标题
legend(b,'Fell sharply','Slight fluctuation','Soaring up')
legend boxoff
```