ISSN: 2710-0588

A Stock Price Prediction Method based on LSTM and K-Means

Bo He[a], Enyu Gong[*], Longbing Li[b], Yongfen Yang[c]

School of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China

[a]hebo@cqut.edu.cn, [*]734770845@qq.com, [b]llbgod@163.com, [c]yangff0518@163.com

Abstract

Artificial neural network, as a nonlinear mapping or adaptive power system made by linking neurons,which can effectively resolve problems such as gradient explosion in stock price prediction processing. Recurrent neural network (RNN) is a common model for processing stock time-series data and is suitable for stock data involving sequential machine learning tasks, but the prediction results are poor when using long time span or nonlinear data for prediction. To address the problems of low prediction accuracy of ordinary neural networks for stock data with poor linearity and the inability of a single LSTM model to show the recommendation level of a target stock, the paper proposes a deep learning factor integrated prediction model based on LSTM-K-Means.On this basis, a stock price prediction method based on long-term and short-term memory network namely LSTM and K-means clustering algorithm is proposed. The method is not only designed to model stock ups and downs at different levels of combinations, but also more intuitively identifies stocks with better ups through returns and volatilities. Through experimental verification, the stock price prediction method based on LSTM with K-Means proposed in this paper is effective.

Keywords

Stock Price Prediction; LSTM Model; K-Means Algorithm; Machine Learning.

1. Introduction

Stock market's changes are often referred to as the "barometer" in the financial field, and research on stocks has been a hot topic in the financial field. With the rapid changes in the international situation under the epidemic, the changes in the world financial environment have become more and more dramatic, and the realization of stock price prediction has an important impact on the achievement of financial goals and stable economic development, and can also used to make or adjust investment decisions with this information[1]. In such a context, it is of great practical importance to apply deep learning related techniques to the stock industry to achieve more accurate stock price predictions. Stock market prediction has been of great interest, however, the process is very complex due to its inherent noisy environment and its large volatility relative to market trends, which are particularly constraining[2]. The long short-term memory network (LSTM) is a temporal recurrent neural network suitable for dealing with forecasting important events with relatively long intervals and delays in the time series. This technical feature has a high fit with the stock forecasting problem, and LSTM compensates the problem of GARCH being limited by linear models[3] and can provide a better solution to the stock forecasting problem. In this thesis, LSTM models are used to design and model the stock rise and fall, and in terms of model design, LSTM models with different combinations of levels are considered and K-Means algorithm is used to assist in the analysis to improve the prediction accuracy as a goal.

ISSN: 2710-0588

2. Related Theories

2.1. LSTM Model

RNN models refer to neural networks with "loops". Compared with traditional neural networks, RNN models incorporate the processing of time-series data. Taking multi-factor stock prediction as an example, traditional neural networks use a certain time cross-section of input factor data[4] to output subsequent excess return forecasts, while RNN models use a time series of long-term factor data of a specific stock as input. Since capital market information is continuous, RNNs can fully capture the opportunity of "history repeating".

However, traditional RNN models are prone to gradient vanishing of the prediction results due to the difficulty in handling long time series of data. In 1997, the long short-term memory network LSTM was proposed[5], which can alleviate the gradient vanishing problem of traditional prediction models by hidden layer neurons[6]. For the LSTM model, at each index position t, the sequence propagation direction is forward. It has another cell state Ct (Cell State) in addition to the same hidden state ht as RNN, as marked with black horizontal lines in Fig. 1. Ct essentially plays the role of the hidden layer state ht in RNN in the LSTM.

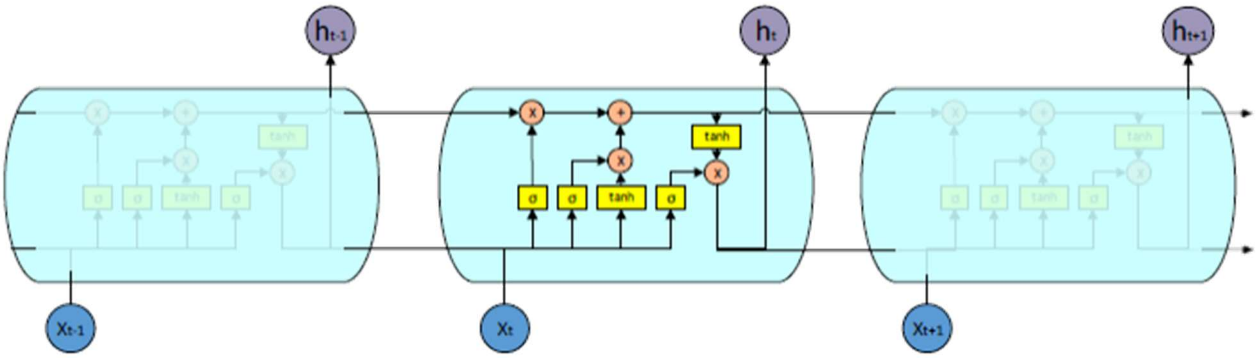


Fig. 1 LSTM model hidden state network model structure

In addition to cell states, there are other important structures in the cells of the LSTM, such as the gating structure (Gate)[7] . The gating structure of the LSTM model at each sequence index position t generally includes Input Gate, Output Gate, and Forget Gate, as shown in Fig. 2. These three gates are essentially weights. When the value is 1, it means that the gate is closed and the sequence passes normally; the value is 0, the gate is open and the sequence cannot pass; when the value is in between, it indicates the degree to which the sequence passes, and the value of this interval is often achieved through the activation function[8].

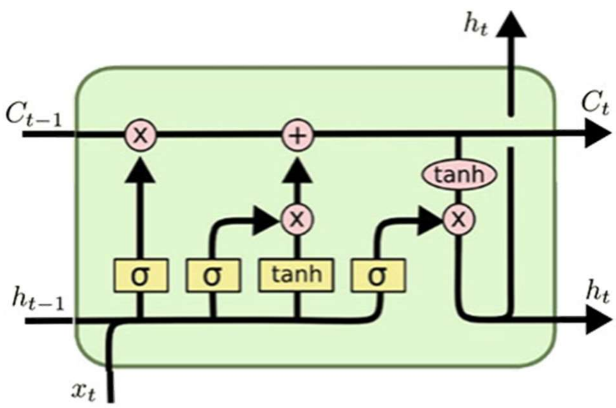


Fig. 2 Cell structure of LSTM model

2.2. ARIMA Model with K-Means Algorithm

ARIMA model is one of the time series forecasting analysis methods[9]. ARIMA(p, d, q), where AR is "autoregressive" and p is the number of autoregressive terms, MA is "sliding average", q is the number of sliding average terms, and d is the number of differences made to make it a smooth series[10] . The general form of the model is as shown in Eq.1 and Eq. 2.

$$\phi(B)*(1-B)^{d}*x(t)=\theta(B)*\varepsilon(t) \tag{1}$$

$$\begin{cases}\Phi(B)=1-\phi_1 B-\phi_2 B^2-\ldots-\phi_{p-1}B^{p-1}-\phi_p B^p\\ \theta(B)=1-\theta_1 B-\theta_2 B^2-\theta_{q-1}B^{q-1}-\theta_q B^q\\ x(t)*B=x(t-1)\end{cases} \tag{2}$$

where x(t) is the SOH series, ε(t) is the random error series of white noise[11] assuming zero mean and equal variance, B is the back-shift operator, p is the order of the AR part, d is the order of the difference computation (DC) operator, and q is the order of the MA part, {ϕp} and {ϕq} are the parameters of the AR model and the MA model, respectively, which mainly introduce their difference processes for the crawled data.

The k-means clustering algorithm is an iterative solution clustering analysis algorithm whose steps are, pre-dividing the data into K groups, then randomly selecting K objects as the initial cluster centers, and then calculating the Euclidean distance between each object and each seed cluster center as shown in Eq. 3, assigning each object to the cluster center nearest to it[12].

$$dis(X_i,C_j)=\sqrt{\sum_{t=1}^{m}(X_{it}-C_{jt})^2} \tag{3}$$

In the above equation Xi denotes the i-th object ∈ [1,n], Cj denotes the j-th cluster center∈ [1,k], Xit denotes the t-th attribute of the i-th object, and Cjt denotes the t-th attribute of the j-th cluster center.

3. Deep Learning Factor Synthesis Prediction Model based on LSTM-K-Means

3.1. Model Framework

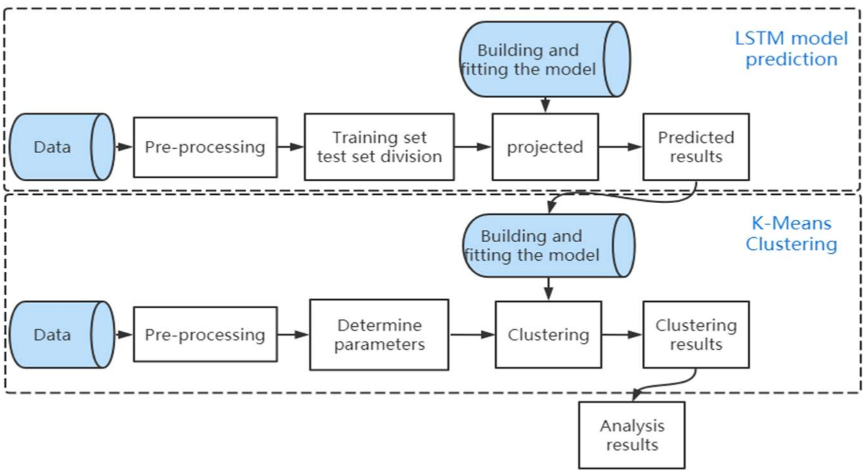


Fig. 3 Deep learning factor synthesis prediction model based on LSTM-K-Means

Considering the different analysis needs of stock data, this thesis combines LSTM model with K-Means algorithm to build a deep learning factor integrated prediction model based on LSTM-K-Means, which can satisfy the prediction of the closing price of the target stock as well as discover more intuitively, through data clustering, whether the stock is rising better and recommended for purchase. The overall framework of the model is shown in Fig. 3.

The basic workflow of the model is: the target stock data is first pre-processed, such as handling missing values and other processes, and the processed data is divided into training set and test set and input into the established LSTM model and predicted based on the relevant factors. The same data, if already suitable for clustering analysis, is directly input to the K-Means clustering process for determining the relevant parameters of the algorithm and can be clustered for analysis. After two layers of processing by LSTM model and K-Means algorithm, it is possible to visualize clearly the rise of the target stock and the degree of recommendation.

3.2. LSTM Layer

For the LSTM layer in this thesis, a two-layer structure is used to fit the data. The two-layer LSTM has a high degree of fit and the model complexity and training difficulty are within a reasonable range. Its main structure is shown in Fig. 4.

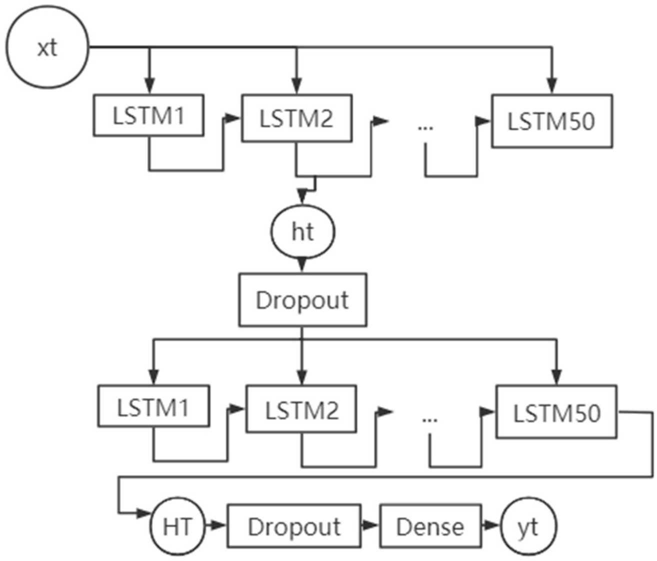


Fig. 4 Two-layer LSTM model structure

When the data is input into the two-layer LSTM model, it will pass through the set LSTM neurons in turn and output the vector of corresponding dimension. After the first layer of data processing is completed, 20% of the data is removed by the Dropout layer to prevent overfitting and then input into the second layer of LSTM, both layers of LSTM have the same number of neurons, The vector is output by the fully connected layer containing one neuron and can be used to plot the stock price of the target stock. The model optimizer is the Adam optimizer[13], and the root mean square error is used as the loss function to judge the prediction effect of the model.

3.3. K-Means Layer

In this thesis, we use the K-Means algorithm to cluster and analyze the data processed by the LSTM model, and additionally selects stocks in A-shares to calculate their volatility and return, and cluster them together with the target stocks, so as to get the recommendation degree of

ISSN: 2710-0588

stocks more intuitively. Firstly, the learning curve and contour coefficients are plotted using the target stock data processed by the LSTM layer[14], and the inflection points of the data can be obtained by the above two plots, which is convenient for determining the number of clusters of the K-Means clustering model. The K-Means algorithm is a good solution to the problem that the LSTM model can only analyze a single stock for a long time span and the recommendation degree is not obvious. The flow chart of its establishment is shown in Fig. 5.

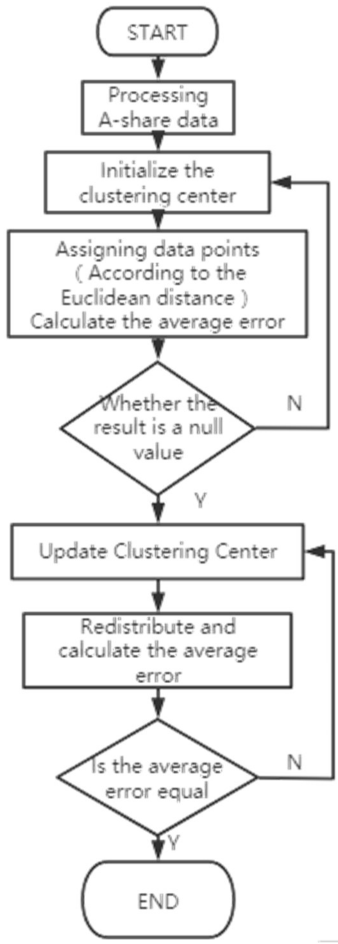


Fig. 5 Flow chart of K-Means algorithm establishment

4. Stock Price Prediction Method based on LSTM Model and K-Means Algorithm

4.1. General Description of the Method

This thesis proposes a stock price prediction method based on LSTM model with K-Means algorithm , which based on a deep learning factor integrated prediction model based on LSTM-K-Means. The method firstly crawls the data related to the target stock; secondly performs data pre-processing; then constructs a two-layer LSTM model to predict the stock; and finally uses the K-Means algorithm to discover the stock with a better upward trend.

This thesis uses TuShare's stock data interface to crawl data related to the target stock. The required modules for the program contain the read data module, data preprocessing module, building module, and graphing module. After the modules are prepared, data pre-processing is performed: (1) Missing value processing: check the missing values, if there are missing values, we need to do some corresponding processing, depending on the situation to see whether to delete or fill, if there are no missing values then proceed to the next step. (2) Data normalization: Perform Min-Max (normalization) operation on the data, map the data to the interval [0,1], and

ISSN: 2710-0588

scale the data, which can improve the convergence speed of the model as well as the accuracy of the prediction. (3) Reconstructing the data: The data set is split into x and y parts, x is used for the prediction of y. This step transforms the data into the time series required for the model as well as the supervised learning problem[15]. After the data processing is completed, the training and test sets are split. In this thesis, the first 80% of the data is selected as the training set to train the model, and the last 20% of the data is used as the test set to test the prediction ability of the model. The model uses a two-layer LSTM model, and the building process is as in Section 3.2. After the prediction is completed, the data needs to be inverse normalized, the prediction results are combined with the test set data and then invert the scaling, and also the expected values on the test set need to be scaled. Because the original data is pre-processed, the error loss calculation is performed on the processed data, and the data needs to be transformed in order to calculate the error on the original scale. Finally, the target stock closing price forecast is plotted.

Using the same data set as in the previous step, the theoretical average annual return Returns and Volatility [16] are calculated and the data are formatted into numpy arrays to be provided to the K-Means algorithm. The data processed in the previous step are used to plot the learning curve and the contour coefficients, and the number of clusters of the algorithm can be obtained by judging the plot. and finally, after clustering, the degree of recommendation can be known by counting the clusters to which the stocks belong. The overall framework of the method is shown in Fig. 6.

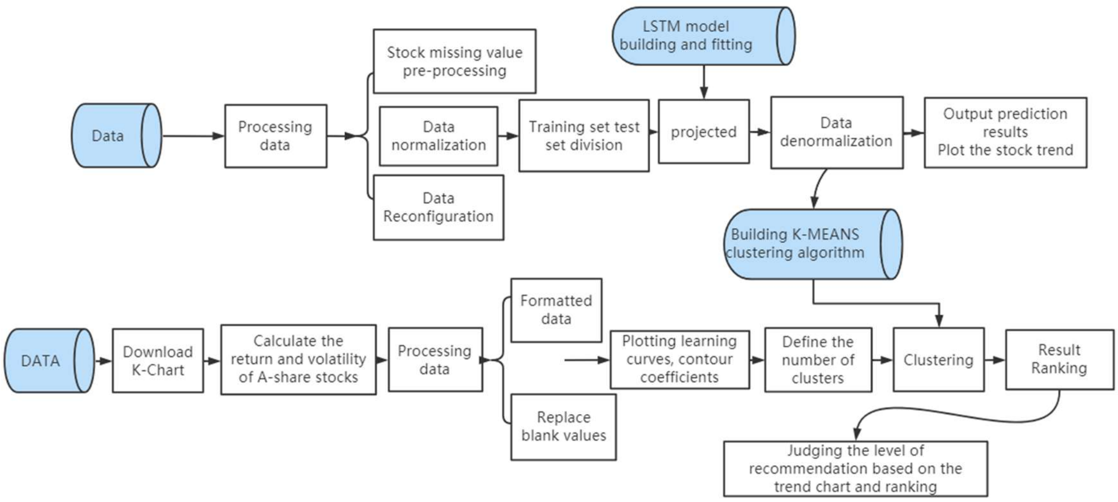


Fig. 6 Framework of stock price prediction method based on LSTM model and K-Means algorithm

4.2. Data Processing

4.2.1. LSTM Partial Data Processing

(1) Pre-processing of missing values of target stocks: Usually, stock market trading is suspended on holidays and weekends, and the data crawled down during this period may have missing values, so in order to improve the accuracy of prediction, it is necessary to process the missing values and delete or fill them according to the situation.

(2) Data normalization: Because of the difference in magnitude of different parameters, and the degree of gradient decline in the process of model training is proportional to the magnitude of the parameters, the data are first subjected to Min-Max (normalization) operation, which maps the data to the interval [0,1] and scales the data, so that the convergence speed of the model can be improved, and the accuracy of the prediction formula is shown in Eq. 4.

ISSN: 2710-0588

$$X_{\text{norm}} = X - \min(X) / \max(X) - \min(X) \tag{4}$$

(3) Data reconstruction: Firstly, the data set is split into x and y parts, x is used for y prediction, this step transforms the data into time series required for the model and supervised learning problem. The input to the LSTM requires a three-dimensional array (samples, timesteps, feature) referring to the number of samples, timesteps, and features, respectively. x_train and x_test will be used as input data for the LSTM, while they are two-dimensional data at this time. data (samples, timsteps=60), because this thesis only takes the closing price for analysis, so we need to add another one-dimensional FEATURE.

4.2.2. K-Means Partial Data Processing

(1) Data formatting: Format the data into numpy arrays to provide to the K-Means algorithm, see 4.2.1 for the normalization process.

(2) Blank value replacement: remove the NaN value and replace it with 0 to avoid insignificant clustering effect.

4.3. Stock Price Prediction and Clustering based on LSTM-K-Means Model

After the data processing is completed, a LSTM-K-Means model based on LSTM is built, the first part consists of two hidden layers of LSTM, containing: one input layer; one LSTM and one fully connected layer as the hidden layer, each constructed layer is dropout to disconnect some neurons to prevent overfitting; the last fully connected layer is used as the output layer (the output dimension is adjusted to 1). The parameter units=50 is chosen to mean that the layer has 50 LSTM neurons, and the output is a 50-dimensional vector. The parameter input_shape contains the number of time steps and the number of features, the parameter return_sequences sets whether to return a one-dimensional array containing the number of steps, the activation function uses tanh, the recurrent activation function uses hard_sigmoid. the cyclic kernel initialization method Orthogonal, the weights use glorot_ uniform initialization method, bias vectors use Zeros initialization method.

Set False to return a two-dimensional array (batch size, number of units) to prevent overfitting of the dropout layer: select the parameter Dropout(0.2) that is, the output of the previous layer, randomly remove 20% of the data, the use of dropout between different hidden layers can make the network more durable and avoid overfitting. Fully connected neural network layer: Dense() each node in this layer is connected to all nodes in the previous layer, and the parameter unites=1 is selected that there is one neuron in this layer. Finally the neural network is compiled and optimized with Adam, using the mean square error as the loss function to evaluate the model, and the overall structure of the model is shown in Table 1.

Table 1. Two-layer LSTM model structure

Model: "sequential"		
Layer(type)	Output Shape	Param #
Lstm(LSTM)	(None,60,50)	10400
Dropout(Dropout)	(None,60,50)	0
lstm_1(LSTM)	(None,50)	20200
Dropout_1(Dropout)	(None,50)	0
Dense(Dense)	(None,1)	51
Total params:30,651		
Trainabel params:30,651		
Non-trainabel params:0		

ISSN: 2710-0588

The second part consists of a K-Means clustering model. In order to build the model, firstly, set the variable n_clusters, i.e., the number of clusters, to 12. Secondly, determine the K value by plotting the learning curve and the contour coefficient, as shown in Fig. 7, and judge the trend graph inflection point at 11 when the contour coefficient is the smallest, so choose n_clusters=11, when the clustering effect is the best. The data processed in section 4.2 are used to calculate a theoretical one-year average annual return Returns and Volatility, which are used as labels for clustering.

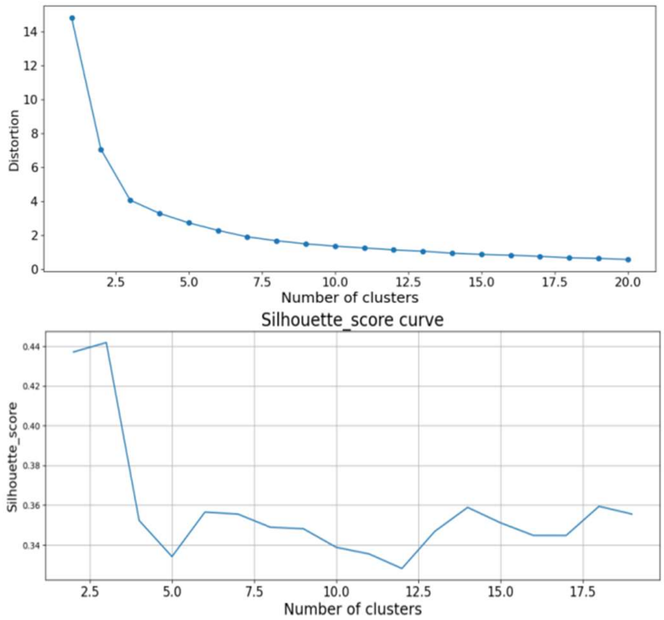


Fig. 7 Learning curve, contour coefficient graph

4.4. Stock Recommendation Degree Determination

After the LSTM model prediction and K-Means algorithm clustering is completed, we can get the closing price trend of the target stock, the result of the stock clustering analysis and its recommendation degree, combined with the trend chart to judge the rise and fall of the target stock, if the rise is good and the recommendation degree is high, then we can judge that the target stock is a quality stock worth buying according to this.

5. Experiment

5.1. Experimental Data Pre-processing

Table 2. 000001 Ping An Bank 2022 Selected Stock Data

Date	OPEN	HIGH	LOW	CLOSE	CHANGE	CHG	VOL	TURNOVER	AMPLITUDE	EXCHANGE
2022-03-31	15.1	15.57	15.06	15.38	0.17	1.12	1,163,747	178,921	3.35	0.6
2022-03-30	14.86	15.27	14.76	15.21	0.53	3.61	1,236,792	185,962	3.47	0.64
2022-03-29	14.88	14.95	14.6	14.68	-0.17	-1.14	615,643	90,917	2.36	0.32
2022-03-28	14.8	15.05	14.63	14.85	-0.13	-0.87	727,137	107,833	2.8	0.37
2022-03-25	15.15	15.25	14.82	14.98	-0.22	-1.45	742,933	111,635	2.83	0.38
2022-03-24	14.92	15.31	14.81	15.2	0.2	1.33	1,002,360	151,887	3.33	0.52
2022-03-23	15.11	15.19	14.9	15	-0.18	-1.19	1,050,102	157,756	1.91	0.54
2022-03-22	14.59	15.22	14.5	15.18	0.59	4.04	1,628,657	244,317	4.93	0.84
2022-03-21	14.5	14.82	14.39	14.59	-0.11	-0.75	1,208,868	176,567	2.93	0.62

ISSN: 2710-0588

This experiment uses TuShare's stock data interface to crawl the stock information related to the example stock 000001 (Ping An Bank), and some of the data are shown in Table 2.

Missing values are detected using the matrix function, and the data of this experiment are detected as complete data without missing values.

The data normalization uses the MinMaxScaler function, and the normalization results are shown in Fig. 8. Data reconstruction uses (1) x, y separation (2) list type transformation to array data (3) two-dimensional data conversion to three-dimensional data reconstruction process, the reconstruction results are shown in Fig. 9.

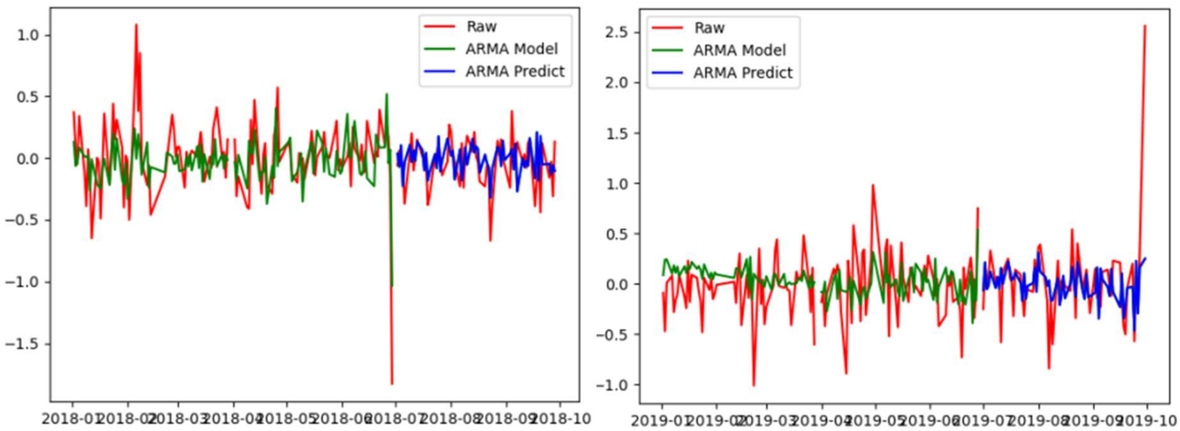
```
[[0.21632653 0.19609433 0.2269327  ... 0.21822332 0.39270073 0.7339405 ]
 [0.21224493 0.20341742 0.23441404 ... 0.20717523 0.43237716 0.7339405 ]
 [0.23591834 0.20423108 0.23940152 ... 0.12925549 0.40592617 0.5474771 ]
 ...
 [0.33795917 0.31407642 0.34912717 ... 0.02769644 0.5675419  0.6108578 ]
 [0.32326525 0.3034988  0.34081465 ... 0.02194686 0.5675419  0.62074625]
 [0.32734692 0.30512607 0.35162097 ... 0.01821499 0.5701644  0.6273559 ]]
```

Fig. 8 Data normalization results

	var1(t-1)	var2(t-1)	var3(t-1)	var4(t-1)	var5(t-1)	var6(t-1)	var7(t-1)	var7(t)
1	0.216327	0.196094	0.226933	0.198728	0.218223	0.392701	0.733940	0.733940
2	0.212245	0.203417	0.234414	0.222575	0.207175	0.432377	0.733940	0.547477
3	0.235918	0.204231	0.239402	0.206677	0.129255	0.405926	0.547477	0.628218
4	0.215510	0.191212	0.231920	0.208267	0.146830	0.408571	0.628218	0.583461
5	0.217143	0.192840	0.232751	0.200318	0.101189	0.395346	0.583461	0.537903

Fig. 9 Data reconstruction results

In order to improve the linearity of the experimental data, this experiment introduces the differencing process of ARIMA model to differentially process the data, and the processing results are shown in Fig. 10. It can be found through Fig. 10 that when the time span gradually decreases, the corresponding data are more volatile and the difference processing effect is worse.



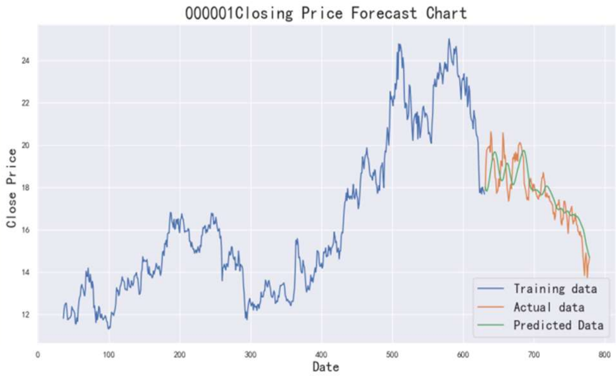
a) 2018~2022 b) 2019~2022

Fig. 10 Differential processing results of stock data with different time span

5.2. Experimental Results and Analysis

5.2.1. Experimental Results of Stock Closing Price Prediction based on LSTM Model

In this experiment, in order to verify the influence of data with different time spans on the experimental results, two sets of Ping An Bank stock data with different time spans are selected for modeling analysis, respectively, a) 2018 to 2022, a total of four years, and b) 2019 to 2022, a total of three years. The first 80% of the data are taken as the training set to train the model, and the last 20% of the data are taken as the test set to test the prediction ability of the model, setting the values of batch_size and epochs to 25 and 10, respectively, i.e., the sample size of each input is 25, and repeating 10 times to train the neural network. The prediction results are shown in Fig. 11. It can be found from Fig. 11 that the longer the time span, the more data in the training set, the stronger the model prediction ability, and the more accurate prediction results.



a) 2018~2022 b) 2019~2022

Fig. 11 Stock data prediction results for different time horizons

5.2.2. Stock Clustering Results based on K-Means Algorithm

Table 3. Selected Results of Stock Cluster Analysis, 2018 to 2022

	Average yield	Volatility	Recommendation level (1 is the highest level)
000001	-0.144822	0.335090	2
000002	-0.066997	0.414413	1
000004	-0.337801	0.291704	2
000005	-0.337562	0.364249	3
000006	-0.547141	0.380174	4
000007	-1.012188	0.587241	5
000008	-0.774166	0.356264	5
000009	-0.485928	0.382936	3
000010	-0.470558	0.484211	3
000011	-0.509757	0.438007	3
000012	-0.654360	0.337445	4
000014	0.027136	0.599191	1
000016	-0.518367	0.397534	4
000017	-0.284714	0.518430	3
000019	-0.433667	0.460587	3

Table 4. Selected Results of Stock Cluster Analysis, 2019~2022

	Average yield	Volatility	Recommendation level (1 is the highest level)
000001	0.561065	0.316350	1
000002	0.230610	0.294521	3
000004	0.454460	0.477491	1
000005	0.183086	0.401619	3
000006	0.029422	0.293103	4
000007	0.250829	0.529893	2
000008	-0.072522	0.317442	4
000009	0.518735	0.487592	1
000010	0.194792	0.388667	3
000011	0.005660	0.349082	4
000012	0.198975	0.342527	3
000014	0.036008	0.377535	4
000016	0.545746	0.425594	1
000017	0.015483	0.509242	4
000019	-0.139136	0.312834	4

ISSN: 2710-0588

5.2.3. Analysis of Results

The prediction results for different time spans are presented as shown in Fig. 11, where the root mean square error for the corresponding time span is shown in Table 5.

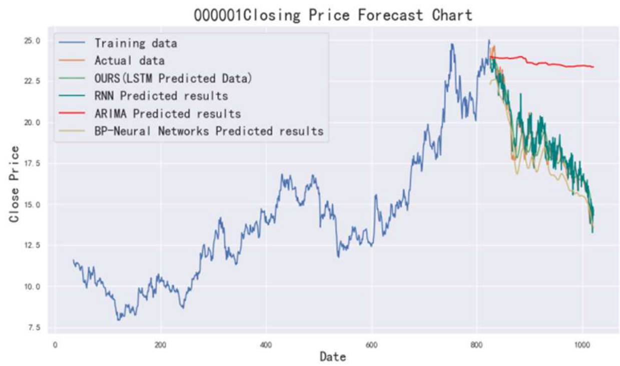
Table 5. RMSE values for different time spans

Time span	RMSE value
2018~2022	0.003
2019~2022	0.0046

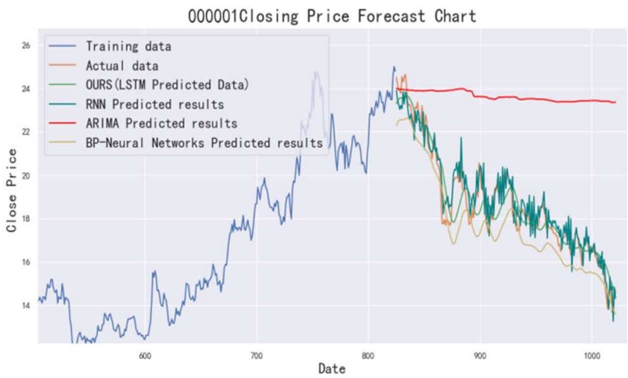
From Table 5 and Fig. 11, it can be seen that The prediction performance of the LSTM model with long time span is much better than that with short time span, and the rate of decline of the loss function of the model with long time span after several iterations is more rapidly converging compared with that with short time span, and the model fits better and the accuracy of the prediction is also higher. Combined with the difference processing in Fig. 10, it is also obvious that when the time span gradually decreases, the corresponding data are more volatile and the prediction is worse. By observing the average return and volatility of Ping An Bank stock and the degree of recommendation in Tables 3 to 4 under different time horizons, we can more quickly understand that Ping An Bank is a stock with smooth ups and downs and guaranteed returns.

5.2.4. Comparison of Methods

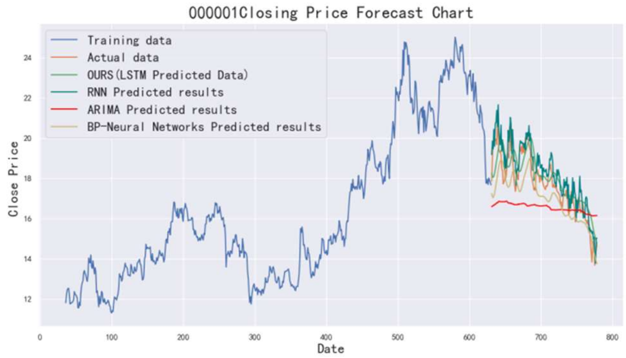
In order to make the prediction effect have more intuitive mention, this experiment uses RNN model, ARIMA model and BP neural network to predict the same experimental data, and the experimental results are shown in Fig. 12.



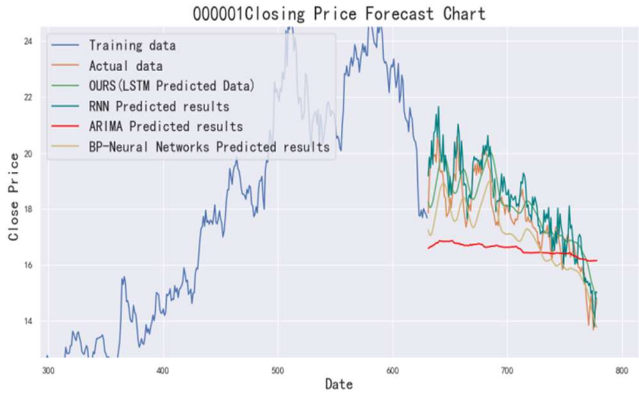
a) Comparison chart of forecast results from 2018 to 2022



b) 2018~2022 results chart model projection section



c) Comparison chart of forecast results form 2019 to 2022



d) 2019~2022 results chart model projection section

Fig. 12 Stock data prediction results for different time horizons

Table 6. RMSE values predicted by different models

Time span	RMSE			
	LSTM	RNN	ARIMA	BP
2018~2022	0.003	0.0042	0.2	0.0046
2019~2022	0.0046	0.0051	0.22	0.005

From Fig. 12 and Table 6, we can see that when using the same time span of experimental data for prediction, the LSTM model has the best prediction effect and smooth prediction curve, the BP neural network prediction results are similar to the LSTM model but with a slightly larger error, the RNN model prediction results are more extreme and change more drastically, which will lead to low stock returns when using the stock price prediction in practice with short position period, the ARIMA model is a time series model with poor ability to handle long time span of experimental data, and the prediction results tend to be a horizontal straight line.

6. Conclusion

In this paper, a prediction study of stocks, a time series, was conducted using a neural network approach. The K-Means algorithm was analyzed and incorporated to build a comprehensive stock forecasting model based on LSTM neural network to obtain more accurate forecasting and recommendation results. The research work of this paper is focused on the following aspects.

(1) The LSTM model was used to study the stock data of Ping An Bank, and the stock data data of different time spans were trained, and then the closing price of the next day was predicted, and finally a more accurate prediction effect was achieved. Experimenting with different

ISSN: 2710-0588

models for the same data, we can see that the LSTM model has the best prediction effect based on the prediction results, and the prediction effect is closely related to the selection of the time step.

(2) K-Means algorithm is proposed to assist in the analysis of the prediction results of the LSTM model, which fully exploits the correlation between different features (yield, volatility) and the dependence between the time-series information. The degree of recommendation of Ping An Bank under different time spans can be obtained more intuitively based on the final clustering results.

Acknowledgments

This research is supported by the humanities and social science research project of chongqing municipal education commission under grant No. 23SKGH252 and the chongqing university of action plan for high quality development of postgraduate education of chongqing university of technology under grant No.gzlcx20233255.

References

[1] Cui H,Yin X,Wen X.:Application of TWSVR Models in Stock Price Forecast[C].In Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence,pp.29-32(2019).https://doi.org/10.1145/3366194.3366200.

[2] Qu Y,Zhang Z,Qin Z.:Wavelet-Aided Stock Forecasting Model based on Ensembled Machine Learning[C].2020 The 3rd International Conference on Machine Learning and Machine Intelligence, pp.37-39(2020).https://doi.org/10.1145/3426826.3426834.

[3] Sun L,Xu W,Liu J.:Two-channel Attention Mechanism Fusion Model of Stock Price Prediction Based on CNN-LSTM[J]. Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 20(5): 1-12(2021).https://doi.org/10.1145/3453693.

[4] He Y,Mao Y,Xie X,; et al.:Multi-factor Data Mining Analysis of Stock Index Volatility[C].2021 2nd International Conference on Artificial Intelligence and Information Systems(ICAIIS), pp.1-9(2021).https://doi.org/10.1145/3469213.3470231.

[5] Huang B,Ding Q,Sun G,;et al.:Stock prediction based on bayesian-lstm[C].In Proceedings of the 2018 10th International Conference on Machine Learning and Computing(ICML), pp.128-133(2018). https://doi.org/10.1145/3195106.3195170.

[6] Lin M,Chen C.:Short-term prediction of stock market price based on GA optimization LSTM neurons[C].In Proceedings of the 2018 2nd International Conference on Deep Learning Technologies (ICDLT'18), pp.66-70(2018).https://doi.org/10.1145/3234804.3234818.

[7] Xu H,Xu B,He J,; et al.:Deep transfer learningbased on LSTM model in stock price forecasting[C]. In 5th International Conference on CrowdScience and Engineering(ICCSE'21),pp.73-80(2021). https://doi.org/10.1145/3503181.3503194.

[8] Hu C,Zheng K,Liu L.:Layered exchange rate prediction model based on LSTM[C].In Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence(ICMAI 2020), pp.124-127(2020).https://doi.org/10.1145/3395260.3395288.

[9] Huang T.:Analysis and Discussion of Stock DataBased on Arima-Grach Model Take Dabinong as an Example[C].In The Sixth International Conference on Information Management and Technology (ICIMTECH'21),pp.1-4(2021).https://doi.org/10.1145/3465631.3465867.

[10] Chen Z.:Asset Allocation Strategy with Monte-Carlo Simulation for Forecasting Stock Price byARIMA Model[C].In 13th International Conference on E-Education, E-Business, E-Management, and E-Learning (IC4E 2022),pp.481-485(2022).https://doi.org/10.1145/3514262.3514331.

[11] Sun Z.:Comparison of Trend Forecast Using ARIMA and ETS Models for S&P500 Close Price[C].In The 4th International Conference on E-Business and Internet(ICEBI 2020),pp.57-60(2020). https://doi.org/10.1145/3436209.3436894.

ISSN: 2710-0588

[12] Li C,Zhang Y,Jiao M,; et al.:Mux-Kmeans: multiplex Kmeans for clustering large-scale data set[C].In Proceedings of the 5th ACM workshop on Scientific cloud computing(Science Cloud'14),pp.25-32(2014).https://doi.org/10.1145/2608029.2608033.

[13] Ta V D,Liu C M.:Stock market analysis using clustering techniques: the impact of foreign ownership on stock volatility in Vietnam[C].In Proceedings of the Seventh Symposium on Information and Communication Technology(Soict'16),pp.99-106(2016).https://doi.org/10.1145/3011077. 30111 40.

[14] Cho C H,Lee G Y,Tsai Y L,;et al.:Toward stockprice prediction using deep learning[C].In Proceedings of the 12th IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC' 19 Companion),pp.133-135(2019).https://doi.org/10.1145/3368235.3369367.

[15] Abe M,Nakagawa K.:Cross-sectional stock priceprediction using deep learning for actual investment management[C].In Proceedings of the 2020Asia Service Sciences and Software Engineering Conference (ASSE'20), pp.9-15(2020).https://doi.org/10.1145/3399871.3399889.

[16] Koratamaddi P,Wadhwani K,Gupta M,;et al.:Market sentiment-aware deep reinforcement learningapproach for stock portfolio allocation[J]. Engineering Science and Technology, an International Journal(CODS COMAD 2021),24(4): 848-859(2021).https://doi.org/10.1145/ 34309 84.3431045.