Associate Professor Jujie WANG, PhD (Corresponding author)
E-mail: wjj0913@126.com
Student Yu CHEN
E-mail: chenyu1566@163.com
Student Shiyao QIU
E-mail: shiyao_qiu@126.com
Student Quan CUI
E-mail: cq15212216737@126.com
School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, China

CUCKOO SEARCH OPTIMIZED INTEGRATED FRAMEWORK BASED ON FEATURE CLUSTERING AND DEEP LEARNING FOR DAILY STOCK PRICE FORECASTING

Abstract. *Under the background of the increasing prosperity of Internet finance, quantitative investment has become a hot topic, among which the prediction of stock price is the focus of research. In this paper, an optimized nonlinear integration framework based on feature clustering and deep learning is proposed to predict stock price daily data. Clustering algorithm is used to divide the complex and changeable stock price data into multiple clusters according to its characteristics, which can pave the way for the establishment of forecast model. Bidirectional long short-term memory (BiLSTM) network is introduced to construct the core of the proposed framework for accurately extracting timing information. Finally, the Radial basis function neural network based on Cuckoo search optimization (CSO-RBF) is constructed for final integration, which shows obvious advantages in improving generalization ability and adaptability of the model. The simulation results, compared with other benchmark models, demonstrate that the prediction performance of the proposed optimization integration framework is obviously better, which provides an effective method for stock price prediction.*

Keywords: *Stock price prediction, Feature clustering, Bidirectional long short-term memory network, Cuckoo search optimization, Nonlinear integration*

JEL Classification: O30

DOI: 10.24818/18423264/55.3.21.04

1. Introduction

High-risk and high-yield stocks have always been the focus of attention in the investment industry, and at the same time they have a major attraction for academic research. The changes in the stock market are inseparable from the market economy dynamics of the entire country. Due to the uncertainty of its own structure and the intricacies of external factors, it is difficult to analyze and forecast stock prices (Zhang et al.,2021). A small increase in the accuracy of stock price forecasts may save millions of dollars in investment costs. Therefore, various methods are constantly adopted to explore the mysteries of stock prices.

The current mainstream methods of stock price forecasting in the academic world are divided into two categories: econometric models and artificial intelligence methods (Wang, 2009; Baffour et al., 2019). Since the econometric model is based on the preconditions that the sample data meets the linear and normality assumptions, it has been proved to be inferior to the artificial intelligence model in the prediction of nonlinear financial time series data (Santos et al., 2007). However, it is difficult for a single model to fully recognize the information contained in the complex and diverse original data. In this case, the hybrid model emerged. It can combine the advantages of multiple models, improve the generalization ability of the model, and enhance the robustness (Chen et al., 2021).

Predictive ensemble method, as a popular hybrid model construction method, mainly consists of three parts. First, perform feature mining on the original data using clustering or decomposition methods, and then make predictions, and finally integrate the prediction results to obtain the ultimate prediction.

In view of the complex and changeable characteristics of financial time series data, most of the current researches use frequency decomposition methods to explore hidden data patterns and in-depth mining of internal laws. Commonly used methods such as empirical mode decomposition (EMD), complete ensemble empirical mode decomposition (CEEMD), etc. can effectively reduce the complexity of the data and are of great significance for analysis (Wang et al., 2019; Cao et al., 2019). Taking into account the diversity of data, the method of feature clustering on the original data, and then building a prediction model for each cluster separately also has a great effect on improving the prediction accuracy (Zhong and Enke, 2017), but clustering methods are currently insufficiently applied in the financial field.

The machine learning methods in artificial intelligence are commonly utilized in prediction part. Support vector machines (SVM) (Vilela et al., 2019), Decision Tree (Zhou et al., 2019) are widely used in nonlinear and non-stationary time series forecasting, and have shown that they are superior to traditional data processing capabilities. However, in the case of large data fluctuations and complex structures, simple machine learning models can no longer meet the needs. A method that can effectively learn data characteristics and grasp long-term fluctuation trends is needed to improve prediction accuracy. Therefore, deep learning technology was developed, which is based on neural networks, can learn and utilize patterns and relationships in data through a self-learning process (Wall, 2018). Among them, recurrent neural network (RNN) has been studied in depth due to its memory and good prediction effects in the financial field (Kim et al., 2018). As a variant of RNN, LSTM can overcome the problem of RNN gradient explosion and is a popular method in the time series prediction community (Kamal et al., 2020). Since in time series data, both past and future information can have impact on the current prediction output, Bidirectional long short-term memory (BiLSTM) can learn forward and backward data information at the same time without retaining information redundancy, which has been proven in research it’s effectiveness(Long et al.,2020).

The integration part mainly integrates the prediction results of weak learners. At the moment, the mainstream methods are Adaptive Boosting (Adaboost) (Zhang et al., 2016), Gradient Boosted Decision Tree (GBDT) (Zhou et al., 2019) and other methods. Due to its excellent nonlinear function approximation ability, RBF neural network is more suitable for the integration of weak learners than classic neural networks that tend to fall into local minima (Han et al., 2020).

In summary, the prediction integration framework is currently a hot research topic in the financial field owing to its strong generalization ability and high prediction accuracy. However, the application of clustering algorithms in data processing and feature extraction is still to be studied. In addition, the prediction result based on the LSTM model is only a continuation of the historical information of the data, and the prediction ability is limited. Moreover, the role of RBF in non-linear data processing is still worth exploring. There are defects in manual tuning, and the application of optimization algorithms to integrated algorithms is insufficient.

To overcome these shortcomings, this paper proposes a new prediction integration framework named K-means++-BiLSTM-CSO-RBF. This method uses BiLSTM as the main prediction model. Based on the K-means++ clustering of complex stock data, BiLSTM is established for each cluster which contains similar feature data generated by the clustering. The CSO-RBF model is introduced to integrate the prediction outputs of BiLSTMs for the ultimate prediction.

The innovations of this paper are as follows. Firstly, The K-means++ algorithm is utilized to efficiently cluster the chaotic and complex raw stock price data according to its potential characteristics, enhance the pertinence of the prediction models for diverse data, and improve the training speed. Secondly, awaring of the excellent characteristic of BiLSTM in extracting past and future features of time series data, BiLSTM is adopted to replace the LSTM model as the key prediction model. Thirdly, construct a CSO-RBF optimization integration method, which can improve the probability of the original RBF model entering the global optimum, to capture and learn the internal laws of prediction data, having stronger generalization capabilities when compared with simple linear integration.

The rest of the paper is organized as follows. Section 2 is an introduction to the theoretical methods involved. The construction of the proposed forecasting framework is shown in Section 3. The detailed description of the empirical process, result analysis and the comparative experiments is summarized in Section 4. Section

5 is a conclusion and outlook.

2. Methodology

This section thoroughly presents the methods utilized in the proposed model.

2.1. K-means++ clustering algorithm

K-means algorithm is a typical clustering algorithm which uses distance as the similarity evaluation index. The basic K-means model is easy to make the algorithm eventually converge to the local optimum, but cannot reach the global optimum. The proposal of K-means++ improves the setting of the initial value and improves the stability of the algorithm. The basic idea is that the distance between the initial cluster centers should be as far as possible.

Comprehensively, steps of K-means++ is listed in Algorithm 1.

Algorithm 1. *K-means++ algorithm*

Input: *Sample set :* $S=\{x_1, x_2, \cdots x_s\}$

Number of clusters: K

For $k = 1,2,\cdots K$:

1. Randomly select an initial center point P_1.

2. Calculate the distance $d(x_m)$ *between the remaining points and* P_1.

$$d(x_m) = \|x_m - P_1\|^2, m = 1,2,\cdots s-1$$

3. $Sum = \sum_{m=1}^{s-1} d(x_m)$

4. For (i=1, r>0, i++), $\forall r \in Sum$

$r = r - d(x_i)$

5. Choose x_i *as the next center point.*

Output: *Center points of K clusters*

After calculating the initial center points, the K-means algorithm is used to calculate the data of K clusters to complete the clustering.

2.2. Bidirectional long short-term memory network (BiLSTM)

LSTM is a method proposed on the basis of RNN (Hochreiter and Schmidhuber, 1997). It achieves a balance between information retention and discarding by adding multiple threshold gates, and exhibits excellent performance in time series modeling and prediction.

By connecting two LSTM networks to form BiLSTM, future and past information can be utilized well in time series processing, and outperformed unidirectional models (Liu et al., 2020). The BiLSTM structure consists of forward and reverse LSTMs, and each LSTM network is connected to an input layer and an output layer. Therefore, the information of the entire time series can be learned by BiLSTM in theory.

The main equations in BiLSTM are as below:

$$\overrightarrow{h_t} = f\left(M_{IF}x_t + M\overrightarrow{h_{t-1}} + \overrightarrow{b_h}\right) \quad (1)$$

$$\overleftarrow{h_t} = f\left(M_{IB}x_t + M_B\overleftarrow{h_{t+1}} + \overleftarrow{b_h}\right) \quad (2)$$

$$y_t = g(M_{FO}\overrightarrow{h_t} + M_{BO}\overleftarrow{h_t} + b) \quad (3)$$

where $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ represent the vector of forward propagation LSTM and back propagation LSTM respectively, y_t denotes the output.

$M_{IF}, M_F, M_{IB}, M_B, M_{FO}, M_{BO}$ represent corresponding weight matrices; $\overrightarrow{b_h}, \overleftarrow{b_h}, b$ are the bias vectors.

2.3. Radial basis function neural network (RBF)

RBF is an effective feedforward neural network, which has strong nonlinear function approximation ability and global optimal characteristics, strong learning convergence ability and fast training speed.

A typical RBF network contains a three-layer structure, namely the input layer, hidden layer and output layer. The input signal is transmitted to the hidden layer through the input layer, and the number of nodes in the input layer is the same as the dimension of the input data. The hidden layer nodes are composed of RBF action functions, and Gaussian kernel functions are generally selected, namely:

$$G_j(z_m) = \exp\left(-\frac{1}{2\sigma_j^2}\left\|z_m - c_j\right\|^2\right), j = 1,2 \cdots k \tag{4}$$

where z_m is the m-dimensional input, c_j is the center of j^{th} Gaussian kernel function, σ_j is the normalization constant for the center of the j^{th} hidden layer. The distance between z_m and c_j is represented by the Euclidean norm $\left\|z_m - c_j\right\|^2$.

The input layer implement a non-linear mapping from $z_m \to G_j(z_m)$, and the hidden layer implement a linear mapping from $G_j(z_m) \to V_q$, the output of the RBF is the linear weighted sum of the output of the hidden unit neuron.

$$V_q = \sum_{j=1}^{k} w_{jq} G_j(z_m), q = 1,2,\cdots,s \tag{5}$$

Where V_q is the output of the q^{th} output node, s is the output dimension, the connection weight from the j^{th} hidden layer to the neuron to the q^{th} output neuron is represented as w_{jq}.

2.4. Cuckoo search optimization algorithm (CSO)

The CSO algorithm is a random global search algorithm that simulates the cuckoo's parasitic brooding behavior, which can effectively solve the optimization problem. At the same time, the Lévy flight search mechanism is adopted by CS to demonstrate the efficiency of search.

The algorithm is based on three idealized rules (Yang and Deb, 2010):

(1) Each cuckoo lays an egg and puts the egg in an arbitrary nest.
(2) The best high-quality egg nest will be inherited to the next generation.
(3) The number of egg nests is fixed as u, and the probability that the host finds cuckoo eggs is $P(A)$. If the host finds foreign bird eggs, it builds another nest.

Assuming that the solution $c_{u,n}$ already exists, the specific process of using the Lévy flight search to obtain the new solution $c_{u+1,n}$ is shown in Eq. 6.

$$c_{u+1,n} = c_{u,n} + \alpha \frac{\emptyset\mu}{|\delta|^{\frac{1}{\tau}}}\left(c_{u,n} - c_{u,best}\right) \tag{6}$$

where $c_{u,n}$ is the n^{th} solution of the u^{th} generation; $\alpha\left(c_{u,n} - c_{u,best}\right)$ is the step size to determine the size of the range during the search; $c_{u,best}$ represents the best solution of the u^{th} generation; $\frac{\emptyset\mu}{|\delta|^{\frac{1}{\tau}}}$ denotes the path of the Lévy flight search, δ and μ obey the standard normal distribution, τ is set 1.5. The calculation of $\emptyset$ is as follows.

$$\emptyset = \left\{\frac{\Gamma(1+\tau)\sin(^{\pi\tau}/_{2})}{\Gamma[(1+\tau)/2]\tau 2^{(\tau-1)/2}}\right\}^{\frac{1}{\tau}} \tag{7}$$

where: Γ is the standard Gamma function.

According to the probability $P(A)$ of the cuckoo egg being found, a part of the solutions are discarded, and the same many new solutions are generated according to Eq. (8).

$$c_{u+1,j} = c_{u,j} + \beta(c_{u,j} - c_{u,q}) \tag{8}$$

where β is the scaling factor, which obeys the uniform random distribution ranging from 0 to 1; $c_{u,j}$ and $c_{u,q}$ are both the random solutions of the u^{th} generation.

2.5. Cuckoo Search-optimized RBF model (CSO-RBF)

The main parameters that affect the effect of RBF regression are the number of hidden layer neurons and the connection weight of the hidden layer to the output layer. A proper parameter selection will increase the probability of RBF entering the global optimum. Therefore, the CSO algorithm which has the capability of jumping out of the local optimum is applied to optimize these two parameters.

The process of CSO-RBF generating parameters is given by Algorithm 2.

Algorithm 2. The process of CSO-RBF parameter optimization

1. Suppose the objective function $L(e), x = (e_1, e_2, \cdots, e_d)^T$
2. Generate n initial populations e_i containing hosts
While *(t<Max Generation) or (Stop Criterion)*
    3. Take a random cuckoo and produce a solution through Lévy flight
  4. Assess the fitness function E_i of the solution
  5. Randomly select one from n nests (assuming w)
    If *($e_i > e_w$)*
    6. Replace w with solution i
  End If
  7. Discard $P(A)$ terrible nests and new nests are produced.
    8. Keep the best solution.
    9. Sort the solution to find the current best solution.
End While
10. Post-processing and visualization

3. Proposed model

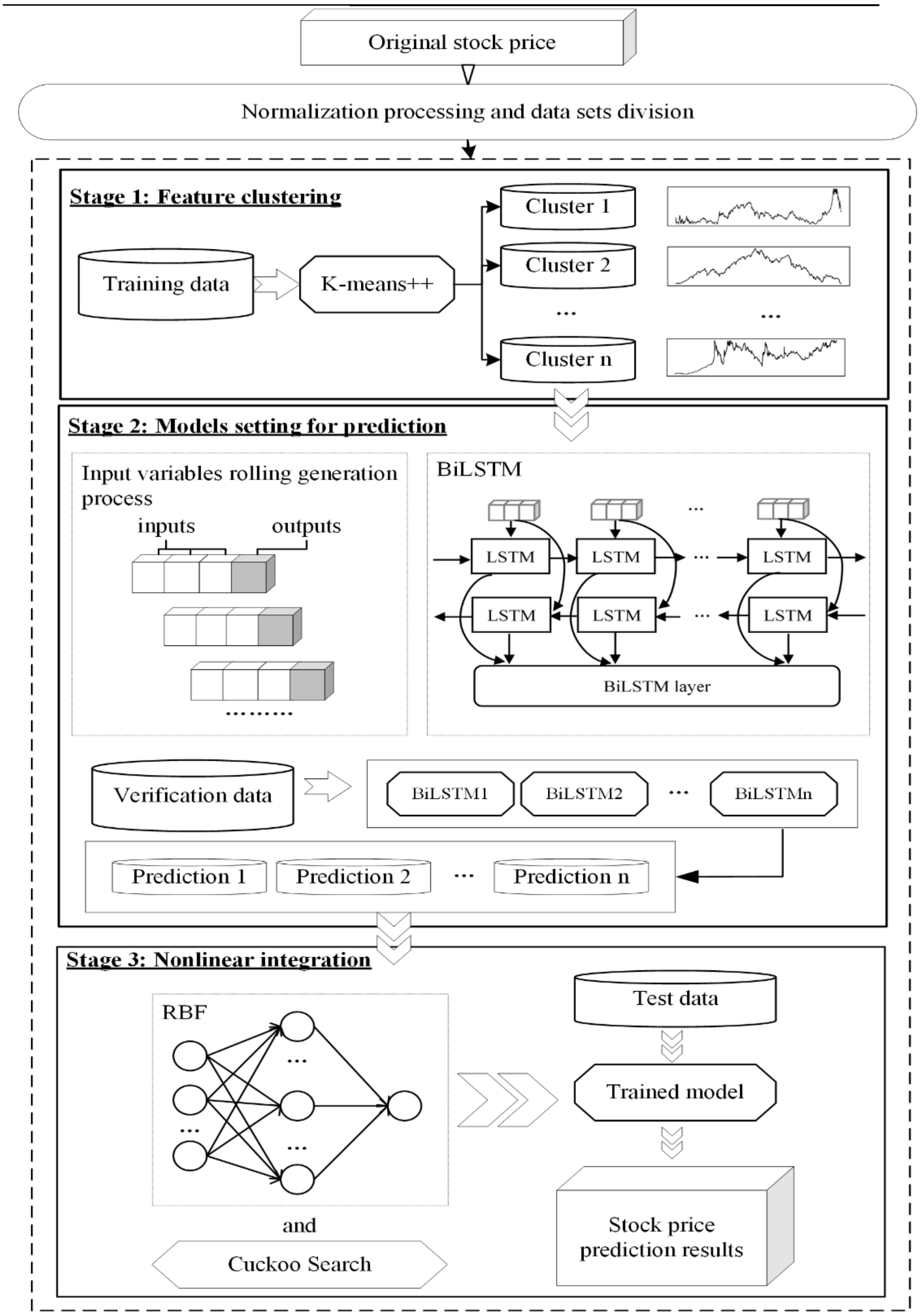

Figure 1. The structure of the proposed framework for prediction

In this paper, a novel forecasting framework is proposed to achieve better prediction performance on stock price based on K-Means++, BiLSTM, RBF and CSO optimization. The novel framework mainly consists of three stages.The detailed structure is shown in Fig. 1.

Stage 1: After the data is normalized and divided, first perform clustering operations. Through the K-means++ clustering algorithm, a large amount of data of the training dataset is divided into multiple clusters according to its potential characteristics, so that the data with high similarity are gathered into one category.

Stage 2: Establish a suitable BiLSTM prediction model for each cluster separately. To improve the utilization of the data, the method of continuously updating the estimation window is adopted to recursively generate the inputs of the BiLSTM, the step size is set to 3, and the rolling window is set to 1. The trained BiLSTM models are retained and applied to the prediction on the verification set and test set.

Stage 3: Optimizing nonlinear integration is the final stage of the prediction framework. The proposed CSO-RBF model with strong nonlinear fitting ability and fast learning convergence is adopted for the final integration. The optimization of RBF parameters selection through the Cuckoo search algorithm not only has capability of preventing the RBF model from falling into the local optimal problem, but also has a significant effect in saving the cost of manual tuning and avoid the artificial error.

4.Case study

In this section, two case studies and a comparison with the base methods are used to confirm the effectiveness of the proposed framework.

4.1 Data description

The Shanghai Composite Index (SSE) and Shenzhen Component Index (SZI) daily stock closing prices are selected as experimental data, and the data comes from the Wind Database (http://www.wind.com.cn/). SSE data range from 1990/12/20 to 2020/11/06, and SZI data is from 1991/1/24 to 2020/11/06. A wide time range can overcome the unpredictability of cycles or unexpected situations caused by insufficient sample size.Divide the first 60% of each dataset into the training set, the middle 20% into the validation set, and the last 20% as the test set. The data trend and division are shown in Figs. 2 and 3. The basic statistical indicators of stock prices are displayed in Table 1.

Table 1. Statistic indicators of SSE and SZI

	Max	Min	Mode	Mean	Media	Range	S.D.
SSE	6092.06	104.39	134.24	1983.00	1903.45	5987.67	1072.230
SZI	19531.15	930.06	3266.01	7270.02	7520.28	18601.09	4053.89

Figure 2. SSE stock price trend and dataset division

Figure 3. SZI stock price trend and dataset division

4.2 Data preprocessing

The data are necessary to be preprocessed before the training starts to make it dimensionless. In this paper, Minmax Scaler is adopted to normalize the data to reduce the influence of outliers on training and achieve better iterative effects. The normalization formula is as follows:

$$\breve{x} = [x - \min(x)]/[\max(x) - \min(x)] \tag{9}$$

where x represents the input vector, the $\breve{x}$ is the normalized results of x.

4.3 Evaluation indicators

In this experiment, root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), which are widely adopted in literatures, as predictive evaluation indicators. The formulas are indicated respectively as follows:

$$RMSE = \sqrt{\sum_{i=1}^{H}(y_i^{pred} - y_i^{true})^2/H} \tag{10}$$

$$MAE = \frac{1}{H}\sum_{i=1}^{H}\left|y_i^{true} - y_i^{pred}\right| \tag{11}$$

$$MAPE = \frac{1}{H}\sum_{i=1}^{H}\left|y_i^{pred} - y_i^{true}/y_i^{true}\right| \times 100\% \tag{12}$$

where H denotes the number of values of test dataset, y_i^{true} is the true observation, y_i^{pred} is the final prediction values obtained by the proposed model.

4.4 Empirical process and results analysis

4.4.1 Feature clustering

According to the potential characteristics of stock price data, K-means++ is adopted to cluster the data to achieve the maximum homogeneity of data in the same cluster and the maximum heterogeneity of data in different clusters. How to determine the key number of clusters —— the value of K, this paper adopts the method of minimizing the cost function to determine. The cost function involved is the sum of the squared errors (S_E). The formula is indicated as Eq. 13.

$$S_E = \sum_{k=1}^{K} \sum_{p \in C_k} |p - m_k|^2 \quad (13)$$

where C_k is the k^{th} cluster, p is the sample point of C_k, the centroid of C_k is represented as m_k.

Before reaching the optimal value of K, as the number of clusters increases, the degree of aggregation of each cluster increases, and S_E will greatly decrease. When the number of real clusters is reached, the change of the aggregation degree obtained by increasing K will significantly reduced, hence the rate of decrease in S_E also rapidly decreases.

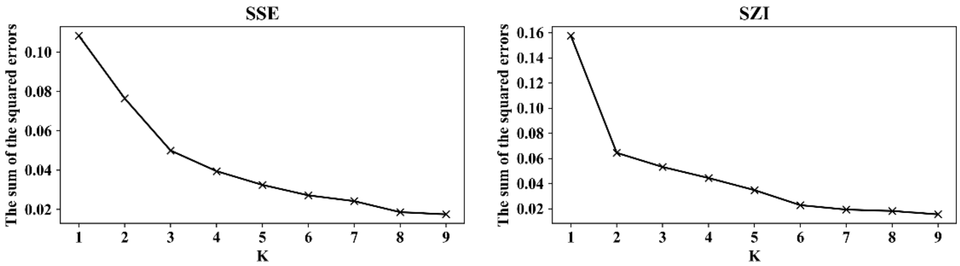

Figure 4. Optimal K value selection

The method of selecting the best K value through S_E is shown in Fig. 4. The abscissa is the K value and the ordinate is the S_E value. It can be seen clearly that as K increases, S_E continuously decreases, and the point at which the rate of decline changes from fast to slow is the optimal K value point. In this way, the SSE training set is divided into three categories, and SZI is divided into two categories.

4.4.2 BiLSTM models settings for prediction

For multiple datasets obtained by clustering, different BiLSTM models are established respectively.

In the BiLSTM algorithm, the main hyperparameters are: epochs, batch size and neurons. To suppress the overfitting problem, the dropout parameter is also introduced. All models uniformly use the 'tanh' function as the activation function, which can achieve faster convergence in practice. The choice of hyperparameters is of great significance for neuron network training. To achieve the best prediction accuracy, each BiLSTM model is tested by trail and error to obtain suitable hyperparameters. Finally, the hyperparameters of each BiLSTM model and prediction accuracy on the testing sets are shown in Table 2.

Table 2. BiLSTM models construction and prediction performance

	SSE			SZI	
	BiLSTM1	BiLSTM2	BiLSTM3	BiLSTM1	BiLSTM2
Epochs	200	500	100	100	50
Batch size	32	16	32	16	16
Neurons	200	50	50	100	300
Dropout	0.1	0.2	0.2	0.1	0.2
MAE	32.8631	25.0802	56.3175	119.4776	127.2785
RMSE	53.1274	34.9425	73.1129	171.4361	182.3428
MAPE (%)	1.0203	1.0114	1.7682	1.156983	1.2216

According to the performance of the evaluation indicators in Table 2, it is obvious to see that the BiLSTM models show quite good prediction performance on the stock price sample sets, which prove the effectiveness of selecting the BiLSTM model as a predictive model.

4.4.3 CSO-RBF models for nonlinear integration

After the test set is separately predicted by the established multiple BiLSTM models, the results obtained will be finally passed to the CSO-RBF model as input variables.

Through the ensemble process, the deterministic prediction results of the stock price can be obtained. Clearly comparison of ensemble prediction results and the true values on two sample sets are shown in Fig. 5 and 6. Meanwhile, the quantitative performance of the proposed framework is summed up in Table 3. It is obvious that the prediction result can successfully simulate the true value of stock price trends and fluctuations, and has a fairly high accuracy rate. Each evaluation indicator performs well, the MAPE of SSE and SZI data sets is only 0.9333% and 1.0849% respectively.

Table 3. Evalution indicators of ultimate prediction results

Dataset	MAE	RMSE	MAPE (%)
SSE	29.8886	46.1923	0.9333
SZI	111.9627	161.9426	1.0849

This result shows that the CSO-RBF model can further improve the forecasting capability of the model on base of the high precision of BiLSTM, all indicators have been improved, and reflects excellent nonlinear fitting performance. At the same time, it can be found that the prediction effect on the SSE dataset is generally superior to that on the SZI dataset. This phenomenon can be explained by that the data of SZI is more volatile than SSE and has more uncertain factors, which increases the difficulty of fitting.

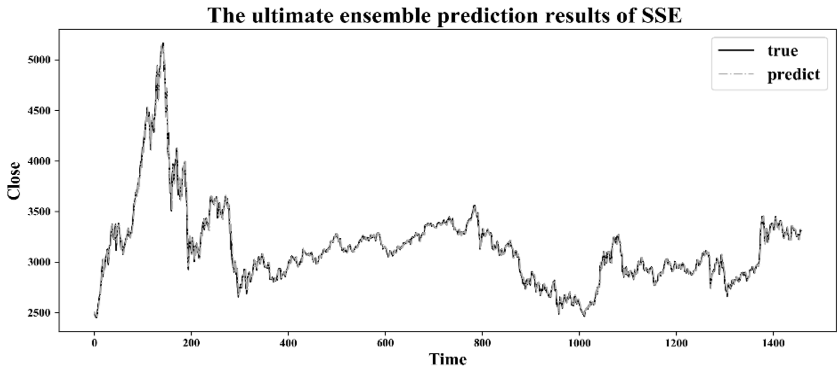

Figure 5. The ultimate prediction results of SSE

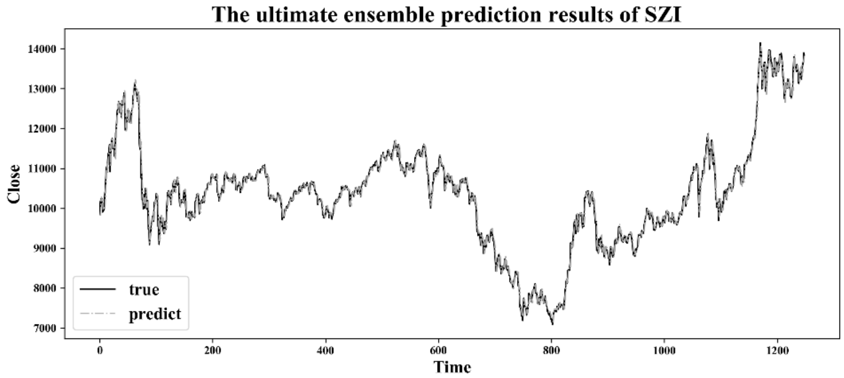

Figure 6. The ultimate prediction results of SZI

4.5 Comparison experiment

Table 4. The performance of comparison models

Dataset	Model	MAE	RMSE	MAPE(%)
SSE	LSTM	60.4601	79.7285	1.9015
	BiLSTM	43.6977	66.0323	1.3643
	BP	150.0329	181.3095	4.3794
	K-means++-BiLSTM	32.9089	53.3575	1.0129
	K-means++-BiLSTM-RBF	32.6074	52.0806	1.0079
SZI	LSTM	258.3140	306.7166	2.5259
	BiLSTM	141.7633	201.8440	1.3373
	BP	540.1586	597.7475	5.0664
	K-means++-BiLSTM	123.4931	176.5613	1.2004
	K-means++-BiLSTM-RBF	124.2284	179.5661	1.1968

To further investigate the effectiveness of the proposed prediction framework, this section conduct a comparative experiment among combined models (K-means++-BiLSTM equal weight linear integration, K-means++-Bilstm-RBF) and single models (BP, LSTM, BiLSTM) by using the same datasets. The quantitative evaluation indicators of the prediction effect of each model on the SSE and SZI datasets are summarized in Table 4. The visualization of the prediction results is displayed in the Fig. 7.

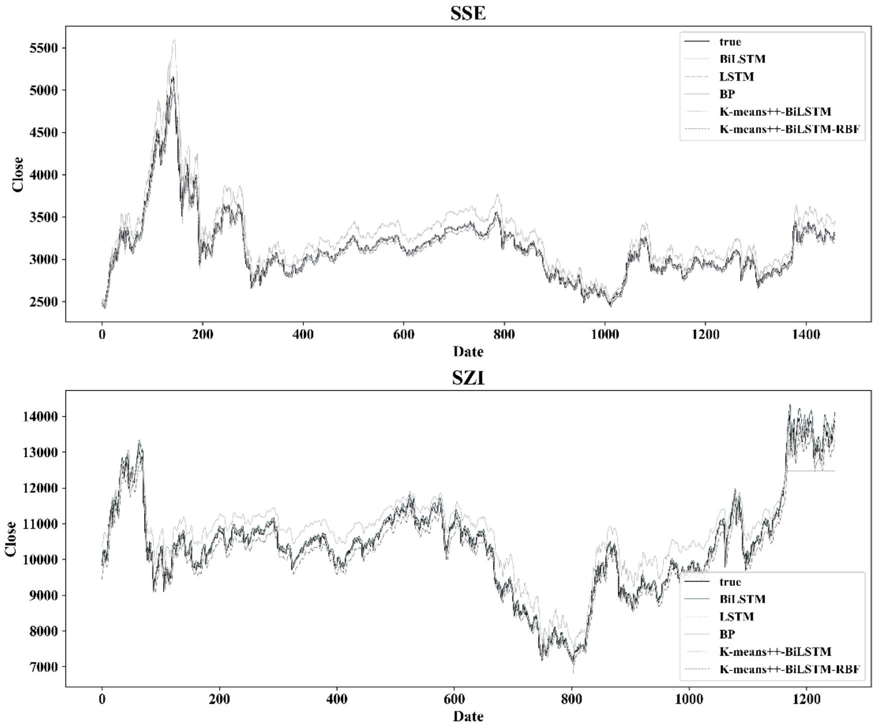

Figure 7. Visualization of comparison models' prediction results

By observing the prediction performance of the comparison model, the following findings can be obtained.

(a) In the prediction part, BP and LSTM are proposed to compare with BiLSTM. The observations can be seen that BiLSTM has the optimum prediction effect, MAPE is only 1.3643% and 1.3373%, followed by LSTM, which MAPE is 1.9015% and 2.5259%. BP has the weakest predictive ability which indicators on the two data sets are nearly three times evaluation that of BiLSTM, and it is difficult to achieve accurate prediction in the stage of relatively severe fluctuations. This shows that deep neural networks such as LSTM and BiLSTM have better prediction effects on time series data than shallow neural networks such as BP. And the combination of forward and backward LSTM can achieve more accurate prediction, and has lower error than using standard LSTM network.

(b) The combined models have better prediction accuracy than a single prediction model. The MAPE values of K-means++-BiLSTM on the two datasets are 0.3514% and 0.1269% lower than that of BiLSTM respectively, reflecting the adoption of the combined models has stronger robustness for datasets with complex features.

(c) The nonlinear integration method can further reduce the error of the combined model. When comparing the K-means++-BiLSTM-RBF with the K-means++-BiLSTM equal-weight linear integration model, taking SSE as an example, MAE, RMSE, and MAPE are reduced by 0.3015, 1.2769, and 0.0050% respectively. The advantage of RBF in fitting nonlinear problems is fully demonstrated.

(d) The CSO optimized RBF has stronger generalization ability than the original RBF. Taking MAPE as an example, the proposed model has reduced errors by 0.0746% and 0.1119% respectively, compared with K-means++-BiLSTM-RBF.

All in all, the prediction framework recommended in this paper outperforms the comparison models, revealing its contribution to stock price prediction.

5. Conclusions

This paper proposes an effective cluster prediction optimization integrated framework named K-means++-BiLSTM-CSO-RBF for application in stock price forecasting. In view of the non-stationary changes of stock market data, K-means++ unsupervised feature learning is introduced to effectively extract the hidden features of stock prices. The improvement of the prediction performance of the BiLSTM models and the enhancement of the framework robustness are all due to it. Considering the complexity of nonlinear subsets, BiLSTM is introduced as the basic prediction model. The predicted values achieved by BiLSTMs are integrated adopting the RBF optimized by the CSO algorithm to get the ultimate stock price forecasting results. On base of the experiment results, it can be confirmed that compared with other baseline models, the stock price prediction framework proposed in this paper has achieved the best prediction results on both SSE and SZI data sets, hence it is very promising to apply it to stock price prediction.

However, this study only utilizes the stock price as input, and does not take into account influencing factors such as technical indicators and related markets. In addition, with the in-depth development of computer technology, more accurate and suitable time series forecasting models will emerge, that are worthy of study.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 71971122 and 71501101).

REFERENCE

[1] **Zhang, D.H., Lou, S.(2021),** ***The Application Research of Neural Network and BP Algorithm in Stock Price Pattern Classification and Prediction.*** *Future Generation Computer Systems, 115: 872-879;*

[2] **Wang, Y.H.(2009).** ***Nonlinear Neural Network Forecasting Model for Stock Index Option Price: Hybrid GJR–GARCH Approach.*** *Expert Systems with Applications, 36 (1): 564–570;*

[3] **Baffour, A.A., Feng, J.C., Taylor, E.K.(2019),** ***A Hybrid Artificial Neural Network-GJR Modeling Approach to Forecasting Currency Exchange Rate Volatility.*** *Neurocomputing, 365: 285-301;*

[4] **Santos, A.A.P., Costa, N.C.A.D., Coelho, L.D. S.(2007),** ***Computational Intelligence Approaches and Linear Models in Case Studies of Forecasting Exchange Rates***. *Expert Systems with Applications, 33(4): 816-823;*

[5] **Chen,W., Zhang, H.Y., Mehlawat, M.K., Jia, L.F.(2021),** ***Mean–variance Portfolio Optimization Using Machine Learning-Based Stock Price Prediction.*** *Applied Soft Computing Journal, 100: 106943;*

[6] **Wang, J.N., Du, J.Z., Jiang, C.H., Lai, K.K.(2019),** ***Chinese Currency Exchange Rates Forecasting with Emd-Based Neural Network.*** *Complexity, 2: 1-15;*

[7] **Cao, J., Li, Z., Li, L. (2019),** ***Financial Time Series Forecasting Model Based on CEEMDAN and LSTM.*** *Physica A: Statistical Mechanics and its Applications, 519: 127–139;*

[8] **Zhong, X., Enke, D.(2017),** ***A Comprehensive Cluster and Classification Mining Procedure for Daily Stock Market Return Forecasting***. *Neurocomputing, 267: 152–168;*

[9] **Vilela, L.F.S., Leme, R.C., Pinheiro, C.A.M., Carpinteiro, O.A.S.(2019),** ***Forecasting Financial Series Using Clustering Methods and Support Vector Regression.*** *Artificial Intelligence Review, 52(2): 743-773;*

[10] **Zhou, F., Zhang, Q., Sornette, D., Jiang, L.(2019),** ***Cascading Logistic Regression onto Gradient Boosted Decision Trees for Forecasting and Trading Stock Indices.*** *Applied Soft Computing, 84: 105747;*

[11] **Wall, L.(2018),** ***Some Financial Regulatory Implications of Artificial Intelligence.*** *Journal of Economics and Business, 55–63;*

[12] **Kim, H.Y., Won, C.H.(2018),** ***Forecasting The Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple GARCH-Type Models.*** *Expert Systems With Applications, 103: 25-37;*

[13] **Kamal, I., Bae, H., Sunghyun, S., Yun, D. H.(2020),** ***DERN: Deep Ensemble Learning Model for Short-and Long-Term Prediction of Baltic Dry Index.*** *Applied Sciences, 1504;*

[14] **Long, J.W., Chen, Z.P., He, W.B., Wu, T.W., Ren, J.T.(2020),** ***An Integrated Framework of Deep Learning and Knowledge Graph for Prediction of Stock Price Trend: An Application in Chinese Stock Exchange Market.*** *Applied Soft Computing Journal, 91: 106205;*

[15] **Zhang X., Li, A., Pan, R.(2016),** ***Stock Trend Prediction Based on a New Status Box Method and Adaboost Probabilistic Support Vector Machine.*** *Applied Soft Computing, 49: 385–398;*

[16] **Han, Y.M., Fan, C.Y., Geng, Z.Q., Ma, B. Cong, D., Chen, K., Yu, B.(2020),**

Energy Efficient Building Envelope Using Novel RBF Neural Network Integrated Affinity Propagation. *Energy, 209: 118414;*

[17] **Hochreiter, S., Schmidhuber, J.(1997),** ***Long Short-term Memory.*** *Neural Computation, 9: 1735–1780;*

[18] **Liu, H., Duan, Z., Chen, C.(2020),** ***A Hybrid Multi-Resolution Multi-Objective Ensemble Model and Its Application for Forecasting of Daily PM2.5 Concentrations.*** *Information Sciences, 516: 266–292;*

[19] **Yang, X.S., Deb, S.(2010),** ***Engineering Optimization by Cuckoo Search.*** *International Journal of Mathematical Modeling and Numerical Optimization. 11(4):330-343.*