IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 4, 2020, accepted December 14, 2020, date of publication December 24, 2020, date of current version January 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047109

Stock Prediction Based on Genetic Algorithm Feature Selection and Long Short-Term Memory Neural Network

SHILE CHEN AND CHANGJUN ZHOU
College of Mathematics and Computer Science, Zhejiang Normal University, Jinhua 321004, China
Corresponding author: Changjun Zhou (zhou-chang231@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 61672121, Grant 61702070, Grant 61751203, Grant 61772100, and Grant 61802040.

ABSTRACT In the financial market, there are a large number of indicators used to describe the change of stock price, which provides a good data basis for our stock price forecast. Different stocks are affected by different factors due to their different industry types and regions. Therefore, it is very important to find a multi factor combination suitable for a particular stock to predict the price of the stock. This paper proposes to use Genetic Algorithm(GA) for feature selection and develop an optimized Long Short-Term Memory(LSTM) neural network stock prediction model. Firstly, we use the GA to obtain a factors importance ranking. Then, the optimal combination of factors is obtained from this ranking with the method of trial and error. Finally, we use the combination of optimal factors and LSTM model for stock prediction. Thorough empirical studies based upon the China construction bank dataset and the CSI 300 stock dataset demonstrate that the GA-LSTM model can outperform all baseline models for time series prediction.

INDEX TERMS Deep learning, feature selection, genetic algorithm, machine learning, optimization methods, forecasting.

I. INTRODUCTION

With the rapid development of social economy, the number of listed companies is increasing, so the stock has become one of the hot topics in the financial field. The changing trend of stock often affects the direction of many economic behaviors to a certain extent [1], so the prediction of stock price has been paid more and more attention by scholars. The stock market data has the characteristics of non-linear, high noise, complexity and timing, etc., so scholars have done a lot of research on the stock prediction method [2]. The traditional stock prediction method is to build a linear prediction model based on the historical stock data, Bowden *et al.* [3] proposed to use ARIMA method to build autoregressive model to predict stock prices. Although this method has some advantages in computational efficiency, the assumption of statistical distribution and stability of the research data limits their ability to model the nonlinear and non-stationary financial time series, and the outliers in the research data also have a great impact on the prediction results. There are many factors affecting stock prices. With the increasing maturity of statistical techniques in the financial field, financial scholars have mined a large number of stock market impact factors and quantified them into specific data for the study of stock change trends. With the support of massive financial data, it provides the possibility for the implementation of machine learning algorithm. More and more researchers begin to use the non-linear prediction model of machine learning to predict stock prices. Nair *et al.* [4] proposed a decision tree system based on rough sets. This method combines the advantages of rough sets and decision trees, but this method is prone to overfitting when dealing with data sets with a large amount of noise, which will affect the trend of stock prediction. In theory, artificial neural network can learn any nonlinear relationship and is less disturbed by noise data, so it has been widely used in the field of time series prediction. Penman [5], Nottola *et al.* [6] have respectively carried out a series of prediction work by using neural network, and achieved better results in stock prediction accuracy than decision tree. However, neural networks are prone to local optimal problems in practical applications, and support vector machines (SVM) based on structural risk minimization

The associate editor coordinating the review of this manuscript and approving it for publication was Bilal Alatas.

greatly reduce the possibility of the model falling into local optimal problems. Cao *et al.* [7] established the SVM stock prediction model, which effectively improved the generalization ability of the model. Ensemble learning is characterized by fast operation rate, strong anti-interference ability and high accuracy rate. Deng *et al.* [8] proved in the experiment that, after parameter optimization, the random forests stock prediction model has a higher prediction accuracy than SVM.

With the development of artificial intelligence technology, deep learning has attracted extensive attention due to its excellent performance in machine translation [9], voice emotion recognition [10], image recognition [11] and other aspects. Compared with the traditional statistical model, the deep neural network (DNN) can analyze the deep and complex nonlinear relationship through the layered feature representation, which is suitable for the multi-factor influence, unstable and complex nonlinear problem of stock data analysis [12]. Tsantekidis *et al.* [13] proposed a stock prediction model based on convolutional neural network (CNN) and compared it with other classical models to verify the effectiveness of the convolution model in stock prediction. However, due to the timing of stock data, the convolutional neural network is not the most suitable neural network model for stock prediction. Selvin *et al.* [14] proposed three stock prediction models based on CNN, recurrent neural network (RNN) and LSTM deep learning networks respectively, and compared the performance of the three models by predicting the stock prices of listed companies. Finally, it was concluded that LSTM neural network is most suitable for forecasting the stock market with time series due to its long-term memorability.

For multivariable financial timing prediction, effective feature selection is very important. Features selection has many benefits, such as: (i) it reduces the training time of the model; (ii) it helps in simplifying the complexity of forecasters; (iii) it improves the accuracy of the model; (iv) it also avoids over-fitting by eliminating unnecessary variables from the feature set [15]. The traditional feature selection methods mainly include filter methods, embedded methods and wrapper methods. Yu *et al.* [16] successfully improved the model prediction accuracy by using PCA for dimension-reduction extraction of feature data combined with SVM model. Qin *et al.* [17] proposed a dual-stage attention-based recurrent neural network (DA-RNN) for feature extraction and sequential prediction. In the first stage, they introduced an input attention mechanism to adaptively extract relevant driving series at each time step by referring to the previous encoder hidden state. In the second stage, they used a temporal attention mechanism to select relevant encoder hidden states across all time steps. With this dual-stage attention scheme, their model make predictions effectively. According to the changes of influence information in different time stages, Zheng *et al.* [18] designed a specific attention network and successfully learned the dynamic influence of the changes of multiple non-predictive time series on the target series over time. Although these methods can effectively capture temporal features, they cannot determine effective multi-factors combination. When the number of factors increases, the factors tend to have collinearity and interfere with each other. GA has a good effect in the application of feature selection problem. The application of population-based GA can effectively solve the problems of noise and collinearity [19]. Therefore, this paper proposes the use of GA for feature selection of multiple factors, and applies it to the LSTM neural network stock prediction model. Through experimental comparison, this method achieved remarkable results in improving the accuracy of stock prediction.

The remainder of this paper is organized as follows: Section 2 describes the methodologies that are used in this study. Section 3 describes the GA-LSTM two-stage stock price prediction model. Section 4 describes the whole experimental process. In this section, the best combination of feature factors was determined by the process of feature selection and experimental comparison. Section 5 summarizes the findings and provides suggestions for further research.

II. RESARCH METHODOLOGY

A. GENETIC ALGORITHM

GA [20] is an adaptive heuristic search algorithm based on the ideas of natural selection and genetic evolution, which is widely used to find the approximate optimal solution of optimization problems with large search space, and can be effectively used in the selection of optimization features. GA encodes a potential solution of a problem into an individual, and each individual is actually an entity with characteristics of chromosomes. The algorithm calls such a number of individuals together as a population, and the optimization process of GA is carried out on the population [21]. As the main carrier of genetic material, chromosome is a collection of multiple genes. Its internal expression is a combination of certain genes, which determines the external expression of individual shape. For example, the characteristics of black hair are determined by a combination of certain genes in the chromosome that control this characteristic. Therefore, the mapping from phenotype to genotype needs to be implemented at the beginning, i.e., encoding work. Because of the complexity of copying genetic code, we tend to simplify it, usually in the form of binary strings [22]. Chromosomes that closer to the optimal solution will have a better chance of reproducing. After the initial generation of the population, according to the principle of survival of the fittest and survival of the fittest, each generation of evolution produced better and better approximate solutions. In each generation, individuals are selected according to the fitness of individuals in the problem domain, and the population representing the new solution set is generated by combining crossover and variation with the help of genetic operators. This process will result in the population evolution, which is like the natural evolution of the population. Then the population would be more suitable to the environment than the previous generation [23]. After decoding, the optimal individual in the last generation of the population can be used as the approximate optimal solution to the problem.

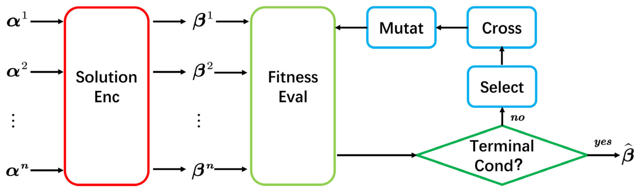

FIGURE 1. Flow chart of GA. Solution Enc: Solution encoding. Fitness Eval: Fitness evaluation. Mutat: Mutation. Terminal Cond: Termination condition.

GA processing can be divided into seven stages: solution encoding, initialization, fitness evaluation, termination condition checking, selection, crossover and mutation [24]. Fig.1 shows the whole process of GA. $\{\alpha^1, \alpha^2, \ldots, \alpha^n\}$ reprents the original feature set. First, it designs a binary encoding for each chromosome β that represents a potential solution to the problem, i.e., the binary encoding of each chromosome represents each feature combination. In the initialization phase, the population size is set for the population and a random original population $\{\beta^1, \beta^2, \ldots, \beta^n\}$ is generated. Then the fitness of each chromosome is calculated according to the pre-set fitness function. The fitness function is an evaluation index used to evaluate the chromosome performance. In GA, the definition of fitness function is a key factor affecting performance [25]. The process of calculating the fitness function will be used to retain the excellent solution for further reproduction. High-performing chromosomes are more likely to be selected multiple times, while low-performing ones are more likely to be eliminated. After several rounds of selection, crossover and mutation operation, we obtain the optimal chromosome $\hat{\beta}$. In this paper, we adopt r^2 determination coefficient as the fitness function of GA. The determination coefficient reflects how much percentage of the fluctuation of Y can be described by the fluctuation of X, i.e., the interpretation degree of the characteristic variable X to the target value of Y. The determination coefficient can be defined as follows:

$$r^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \tag{1}$$

where, the determination coefficient is represented by r^2, y is the label value, $\hat{y}$ is the predicted value, $\bar{y}$ is the average value, and the value range of r^2 is [0, 1]. The larger r^2 is, the stronger the ability of X to explain Y of this chromosome is, and the more likely it is to be passed on to the next generation. The process of chromosome crossing and mutation has great significance to GA. It is beneficial to increase the genetic diversity of the population to exchange the corresponding part of chromosome chain and change the gene combination to produce new offspring.

B. LONG SHORT-TERM MEMORY NEURAL NETWORK

RNN is a kind of recursive neural network which takes sequence data as input and performs recursion in sequence evolution direction and all nodes are connected by chain. Because of its memorability, RNN has achieved good results in the short sequence model. However, as the length of input sequence becomes longer, the number of layers in the network will increase greatly, which will easily cause problems such as gradient disappearance [26].

LSTM is a special deep RNN, LSTM greatly enhances the memory capacity of the model due to its special gate mechanism neural unit structure, and solves the problem of gradient disappearance caused by excessively long input sequence in the learning process of traditional cyclic neural network. Fig. 2 [27] shows the network structure of LSTM, the LSTM network saves all the information before each time step in the neural unit of the current time step, and each neural unit is controlled by the input gate, forgetting gate and output gate [28]. The input gate is used to control the input information of the neural unit at the current moment, the forgetting gate is used to control the historical information stored in the neural unit at the previous moment, and the output gate is used to control the output information of the neural unit at the current moment. The purpose of this design is to allow the LSTM model to selectively remember more important historical information.

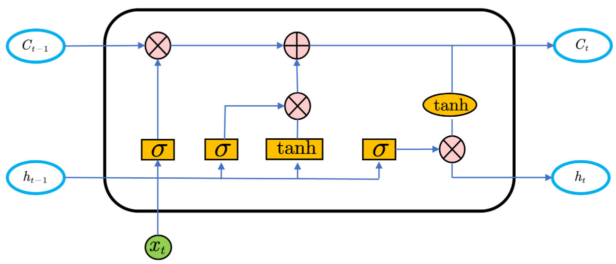

FIGURE 2. LSTM internal structure.

The update calculation method of LSTM is as follows:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{2}$$
$$\mathrm{f}_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{3}$$
$$\mathrm{o}_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{4}$$
$$\tilde{C}_\mathrm{t} = \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{5}$$
$$C_\mathrm{t} = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{6}$$
$$\mathrm{h}_t = o_t \times \tanh(C_t) \tag{7}$$

where, σ represents the sigmoid activation function in the LSTM network which is used to speed up the training process, W_f, W_i, W_c and W_o are weight matrices of forgetting gate, input gate, update gate and output gate respectively, b_f, $\mathrm{b_i}$, $\mathrm{b_c}$ and $\mathrm{b_o}$ are respectively the bias of forgetting gate, input gate, update gate and output gate. Finally, the output at the current moment and the updated cell state at the current moment are calculated.

III. GA-LSTM TWO-STAGE STOCK PRICE PREDICTION MODEL

This paper mainly proposes a two-stage stock price prediction model combined with GA and LSTM deep learning network. The experiment is divided into the following two stages.

The first stage is to use GA to sort the importance of factors. The specific steps are as follows:

(i) Binary encoding of chromosomes, random initialization of the population. We denote the population of GA using pop as follows:

$$POP = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,k} \end{bmatrix} \tag{8}$$

In the above matrix, each row represents a chromosome (or a feature selection set), the chromosome length k represents the total number of features(40 in this paper), number m represents the population size(100 in this paper), the value of $a_{i,j}$ is 0 or 1, 1 represents selection, 0 represents non-selection, where are non-zero positive.

(ii) Taking roulette method for selection operation. We calculate the fitness of each chromosome in the population. The probability of each individual being selected is proportional to the fitness of chromosomes, and the sum of the probability of chromosome being selected is 1. During the algorithm, the population is updated once per cycle according to the probability.

(iii) The multi-point intersection method was used to carry out the crossover operation, and the chromosomes between two individuals will be exchanged, with the crossover probability set as 0.8. In the algorithm process, a cross operation is performed on each chromosome in each cycle. The method is to generate a random probability. If the random probability is less than the crossover probability, the exchange will be carried out, otherwise, there is no exchange.

(iv) The basic bit mutation method is used to carry out mutation operation. In contemporary individuals, a gene is altered with a small probability. The probability of variation is set to 0. 003. The algorithm produces one probability at a time. If the random probability is less than the crossover probability, variation will be carried out; otherwise, no variation will be carried out.

Loop steps (ii) through (iv) until the iteration is 100 times. At the end of the algorithm, we generate an optimal population close to the optimal solution. In this paper, the total number of occurrences of each factor in the population is statistically ranked as the factor importance ranking. The more times the factor appears, the more important it is. Table 1 shows the specific model parameters.

The second stage of this study is the stage of feature selection optimization of LSTM stock prediction model. Based on the factor importance ranking obtained in the previous stage, the top 40, 30, 20, 10 and 5 factors were taken as input features of the LSTM model. By comparing the prediction results, the optimal factor combination is determined, and the optimal model is compared with the base line models to verify the superiority of the proposed optimization model in improving the model accuracy.

TABLE 1. Parameters of the GA model.

parameter name	parameter value
chromosome size	40
Population size	100
crossover rate	0.8
mutation rate	0.003
iteration	100

TABLE 2. Parameters of the LSTM model.

parameter name	parameter value	parameter name	parameter value
network layers	3	loss function	MSE
hidden layer size	128	activation function	Elu
batch size	512	optimization	Adam
dropout	0.2	epochs	100
time step	5	train:test	8:2

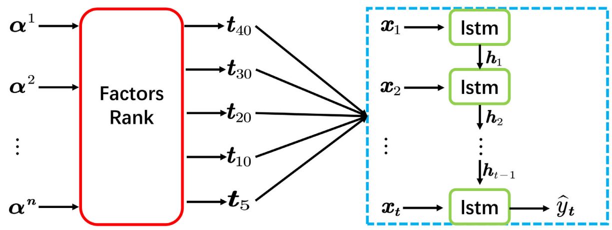

FIGURE 3. Experimental workflow. Factors Rank: Factor importance ranking. x_t: the input at time t. h_{t-1}: the hidden state of LSTM at time t-1.

Table 2 shows the specific model parameters. There are three network layers in the model, namely, the input layer, the hidden layer and the output layer. The number of neurons in the hidden layer and the output layer is 128 and 1 respectively, and the dropout parameter is set as 0.2 to randomly remove a part of neurons to avoid overfitting. The LSTM network time step is set to 5, i.e., this paper takes the historical data of the first five days as input to predict the stock price of the next day. Model gradient descent optimizer is Adam, and the number of model iterations is 100. In this paper, the data are divided into training set and test set in a scale of 8:2. The first 80% of the data is used for training, while the remaining 20% of the data is used to evaluate the model.

The model adopts mean square error (MSE) as the model evaluation index, and the formula is as follows,

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{9}$$

where, m is the number of samples, y_i is the stock price, and $\hat{y}_i$ is the model forecast stock price. Fig. 3 shows the experimental process of this study. We obtain 5 multi-factor combinations t_{40}, t_{30}, t_{20}, t_{10} and t_5 by ranking the importance

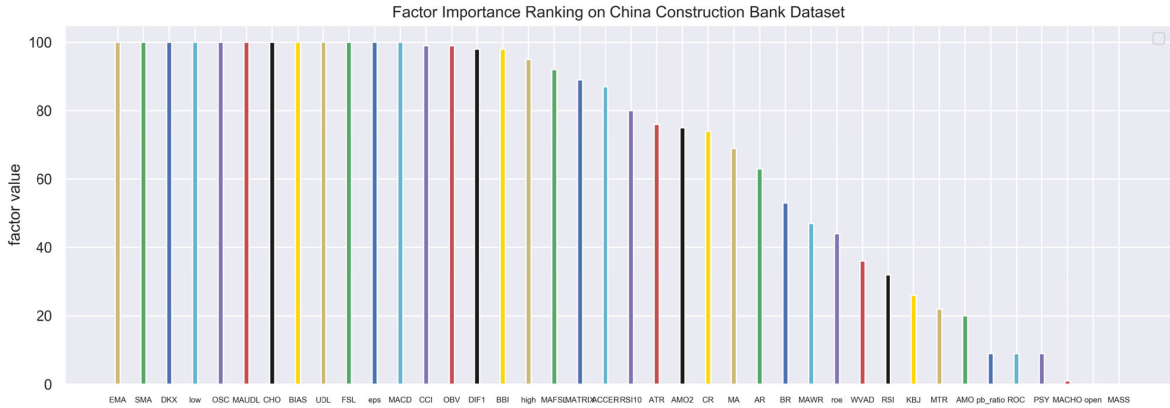

FIGURE 4. **Factor importance ranking on China construction bank dataset.**

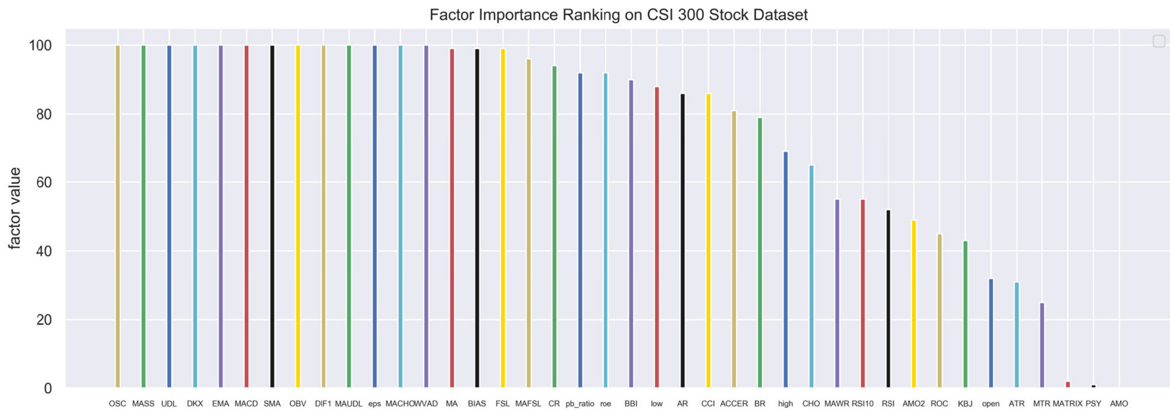

FIGURE 5. **Factor importance ranking on CSI 300 stock dataset.**

of the original feature set $\{\alpha^1, \alpha^2, \ldots, \alpha^n\}$. These multi-factor combinations are used as the input features of LSTM to predict stock price $\hat{y}_t$ at time t.

IV. EXPERIMENT

A. EXPERIMENTAL DATA

According to China's financial market, this paper determines the original factor combination consisting of 40 stock factors which including market factors, technical factors and financial factors. In this paper, 2490 pieces of historical data of China construction bank and CSI 300 stock from January 1, 2010 to April 1, 2020 are obtained through the JoinQuant quantitative platform [29]. In order to eliminate the dimensional influence between indexes and accelerate the speed of gradient descent to find the optimal solution, we normalized the data. The normalization principle is as follows,

$$\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{10}$$

where, $\min(x)$ and $\max(x)$ are the maximum and minimum values of x respectively.

B. FACTOR IMPORTANCE RANKING

In this subsection, GA is used to carry out 100 iterations for all 40 factors, and the factor importance ranking is obtained. Fig. 4 shows the details of the factor importance ranking on China construction bank dataset. The top factors EMA, SMA, DKX, low, OSC, MAUDL, CHO, BIAS, UDL, FSL, EPS, MACD, CCI, OBV, DIF1 and BBI have strong factor importance, while the factor importance decreases gradually after high. MACHO, open and MASS factor have almost no influence on the stock price. Fig. 5 shows the details of the factor importance ranking on CSI 300 stock dataset. The top factors OSC, MASS, UDL, DKX, EMA, MACD, SMA, OBV, DIF1, MAUDL, eps, MACHO, WVAD, MA, BIAS and FSL have strong factor importance, while the factor importance decreases gradually after MAFSL. MATRIX, PSY and AMO factor have almost no influence on the stock price.

C. ANALYSIS OF EXPERIMENTAL RESULTS

In this experiment, 2490 historical data of China construction bank and CSI 300 stock from January 1, 2010 to April 1, 2020 were substituted into the LSTM model, and the data were processed by mean filling and normalization. The experiment is trained according to the first 80% data and tested according to the last 20% data, and the MSE of test set was obtained for model evaluation. In particular, the original factor subset and the top 30, 20, 10 and 5 factor subsets in the factor importance ranking were used as the input features

TABLE 3. Performance comparison of different methods.

model	China construction bank	CSI 300 stock
	MSE	MSE
PCA-SVM	0.0080	0,0096
Random forests	0.0078	0.0083
DA-RNN	0.0067	0.0070
LSTM	0.0072	0.0073
GA-LSTM(K=30)	0.0066	0.0054
GA-LSTM(K=20)	0.0053	0.0039
GA-LSTM(K=10)	0.0042	0.0043
GA-LSTM(K=5)	0.0047	0.0050

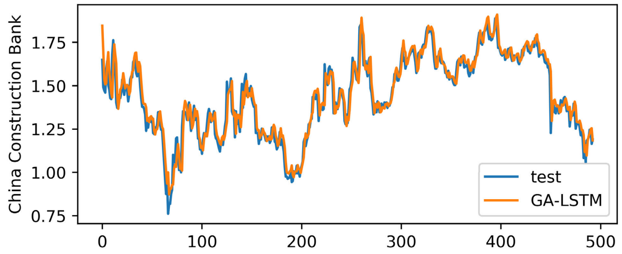

FIGURE 6. GA-LSTM(k = 10) model predicted results on China construction bank dataset.

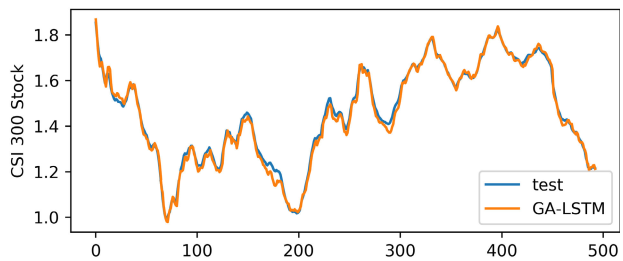

FIGURE 7. GA-LSTM(k = 20) model predicted results on CSI 300 stock dataset.

of the LSTM model for comparative experiments. To evaluate the performance of our proposed approach, the performances of four baseline models and our proposed model are compared. Among them, the LSTM model is very suitable for forecasting the stock market with time series due to its long term memorability [14]. The PCA combined with SVM (PCA-SVM) model [16] is considered as a classical nonlinear stock forecasting model which can effectively improve the generalization ability. The random forests stock prediction model [8] can effectively improve the prediction accuracy after parameter optimization. The dual-stage attention-based recurrent neural network(DA-RNN) [17] is an advanced feature extraction and sequential prediction model. Table 3 shows the performance comparison of different methods. In Table 3, k is used to refer to the factor number of feature set. As is shown in Table 3, the proposed GA-LSTM model can outperform all baseline models on both two datasets. On China construction bank dataset, when the top 10 factor subsets in the factor importance ranking are taken as the input features of the proposed GA-LSTM model, the prediction fitting degree reaches the highest. The MSE reaches a minimum of 0.0042. On CSI 300 stock dataset, when the top 20 factor subsets in the factor importance ranking are taken as the input features of the proposed GA-LSTM model, the prediction fitting degree reaches the highest. The MSE reaches a minimum of 0.0039. As is shown in Fig. 6 and 7, our proposed GA-LSTM model performs very well on the test set.

TABLE 4. Factor definition.

factor name	defination
OSC	variable rate line index
MASS	relative strength index 6
UDL	gravity line index 1
DKX	more than empty line index
EMA	exponential moving average index
MACD	average of smooth similarities and differences index
SMA	moving average index
OBV	cumulative energy line index
DIF1	trend line index 1
MAUDL	gravity line index 3
eps	earnings per share index
MACHO	trend line index 2
WVAD	william's variable accumulation distribution index
MA	average index 1
BIAS	take away rate index
FSL	watershed index 1
MAFSL	watershed index 2
CR	relative strength index 5
pb_ratio	price value ratio index
roe	return on equity index
BBI	average index 2
low	lowest price index
AR	relative strength index 4
CCI	goods path index
ACCER	higher speed range index
BR	relative strength index 3
high	top price index
CHO	trend line index 3
MAWR	gravity line index 2
RSI10	relative strength index 2
RSI	relative strength index 1
AMO2	volume index 2
ROC	rate of change index
KBJ	random index
open	opening price index
ATR	true amplitude index 1
MTR	true amplitude index 2
MATRIX	true amplitude index 3
PSY	psychological line index
AMO	volume index 1

V. CONCLUSION

In this paper, the multi-factor model is introduced into stock forecasting. The stock market has a large number of stock factors which describe the change of stock price. In this research, a large number of typical stock factors are selected. However, typicality does not mean that it can be applied to all cases, so GA is proposed for feature selection to select the factors more suitable for the current scene. And combined with LSTM deep learning network model, the complex nonlinear relationship between factors and stocks is mined to predict stock prices.

Although the stock price prediction model proposed in this paper can effectively improve the prediction accuracy and has strong robustness, there are still some shortcomings as follow: Firstly, we only use Chinese stock data for experiment, so further research can include data from different stock markets. Secondly, in the design of model parameters in this paper, trial and error is usually adopted instead of systematic method to find the optimal size of parameters, such as the selection of number of factors. The improvement method is to combine with other machine learning technologies to find the optimal parameters and improve the interpretability of the model. In addition, when the control parameters of GA are set, such as crossover rate, mutation rate and number of factor combinations, a variety of suitable combinations can be derived to improve the performance of the research.

APPENDIX

Table 4 shows the 40 original factors definition in the experiment.

REFERENCES

[1] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *J. Financial Econ.*, vol. 33, no. 1, pp. 3–56, Feb. 1993.

[2] P. Ding, *Quantitative Investment: Strategy and Technology*. Beijing, China: Electronics Industry Press, 2012, pp. 122–127.

[3] N. Bowden and J. E. Payne, "Short term forecasting of electricity prices for MISO hubs: Evidence from ARIMA-EGARCH models," *Energy Econ.*, vol. 30, no. 6, pp. 3186–3197, Nov. 2008.

[4] B. B. Nair, V. P. Mohandas, and N. R. Sakthivel, "A decision tree-rough set hybrid system for stock market trend prediction," *Int. J. Comput. Appl.*, vol. 6, no. 9, pp. 1–6, Sep. 2010.

[5] H. Penman, "Application study of BP neural network and logistic regression for stock investment," *Expert Syst.*, vol. 5, no. 12, pp. 21–22, Dec. 1989.

[6] F. Li and C. Liu, "Application study of BP neural network on stock market prediction," in *Proc. 9th Int. Conf. Hybrid Intell. Syst.*, Beijing, China, 2009, pp. 2–5.

[7] L. J. Cao and F. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Trans. Neural Netw.*, vol. 14, no. 6, pp. 5–10 Nov. 2003.

[8] J. Deng and L. Li, "Application of parameter optimization stochastic forest in stock prediction," *Software*, vol. 41, no. 1, pp. 178–182, Jan. 2020.

[9] M. R. Costa-jussà, A. Allauzen, L. Barrault, K. Cho, and H. Schwenk, "Introduction to the special issue on deep learning approaches for machine translation," *Comput. Speech Lang.*, vol. 46, pp. 367–373, Nov. 2017.

[10] H. M. Fayek, M. Lech, and L. Cavedon, "Evaluating deep learning architectures for speech emotion recognition," *Neural Netw.*, vol. 5, no. 3, pp. 23–28, Aug. 2017.

[11] J. Xing, K. Li, W. Hu, C. Yuan, and H. Ling, "Diagnosing deep learning models for high accuracy age estimation from a single image," *Pattern Recognit.*, vol. 66, no. 1, pp. 106–116, Jun. 2017.

[12] D. Lv, D. Wang, M. Li, and Y. Xiang, "DNN models based on dimensionality reduction for stock trading," *Intell. Data Anal.*, vol. 24, no. 1, pp. 19–45, Feb. 2020.

[13] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Forecasting stock prices from the limit order book using convolutional neural networks," in *Proc. IEEE 19th Conf. Bus. Informat. (CBI)*, Beijing, China, Jul. 2017, pp. 10–15.

[14] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Beijing, China, Sep. 2017, pp. 234–239.

[15] U. F. Siddiqi, S. M. Sait, and O. Kaynak, "Genetic algorithm for the mutual information-based feature selection in univariate time series data," *IEEE Access*, vol. 8, pp. 9597–9609, 2020.

[16] H. Yu, R. Chen, and G. Zhang, "A SVM stock selection model within PCA," *Procedia Comput. Sci.*, vol. 31, no. 1, pp. 406–412, May 2014.

[17] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. W. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Melbourne, VIC, Australia, Aug. 2017, pp. 2627–2633.

[18] J. Hu and W. Zheng, "Multistage attention network for multivariate time series prediction," *Neurocomputing*, vol. 18, no. 383, pp. 122–137, Mar. 2020.

[19] G. Li, P. Liu, C. Le, and B. Zhou, "A novel hybrid meta-heuristic algorithm based on the cross-entropy method and firefly algorithm for global optimization," *Entropy*, vol. 21, no. 5, p. 494, May 2019.

[20] Y. C. Li, L. N. Zhao, and S. J. Zhou, "Review of genetic algorithm," *Adv. Mater. Res.*, vol. 179, no. 180, pp. 365–367, Aug. 2011.

[21] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI, USA: Univ. of Michigan Press, 1975, p. 183.

[22] G. Armano, M. Marchesi, and A. Murru, "A hybrid genetic-neural architecture for stock indexes forecasting," *Inf. Sci.*, vol. 170, no. 1, pp. 3–33, Feb. 2005.

[23] I. Sekaj and V. Veselý, "Robust output feedback controller design: Genetic algorithm approach," *IMA J. Math. Control Inf.*, vol. 22, no. 3, pp. 257–265, Sep. 2005.

[24] S. Jadhav, H. He, and K. Jenkins, "Information gain directed genetic algorithm wrapper feature selection for credit rating," *Appl. Soft Comput.*, vol. 69, no. 1, pp. 35–37, Aug. 2018.

[25] K. Deb, "Multi-objective genetic algorithms: Problem difficulties and construction of test problems," *Evol. Comput.*, vol. 7, no. 3, pp. 205–230, Sep. 1999.

[26] D. Wang and J. Fang, "Research on optimization of big data construction engineering quality management based on RNN-LSTM," *Complexity*, vol. 15, no. 2, pp. 15–20, Jun. 2018.

[27] Y. Kim, J.-H. Roh, and H. Kim, "Early forecasting of rice blast disease using long short-term memory recurrent neural networks," *Sustainability*, vol. 10, no. 2, p. 34, Dec. 2017.

[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[29] *JoinQuant Quantitative Platform*. Accessed: Apr. 15, 2020. [Online]. Available: https://www.joinquant.com/

SHILE CHEN received the bachelor's degree in electrical engineering from the Zhejiang University of Media and Communications, in 2017. He is currently pursuing the master's degree in computer Science and Technology with Zhejiang Normal University, Jinhua, China. His research interests include machine learning, deep learning, social networking, natural language processing, and finance. He has won scholarships from Zhejiang Normal University twice.

CHANGJUN ZHOU was born in Shangrao, China, in 1977. He received the Ph.D. degree in mechanical design and theory from the School of Mechanical Engineering, Dalian University of Technology, Dalian, in 2008. He is currently a Professor with Zhejiang Normal University. His research interests include pattern recognition, intelligence computing, and DNA computing. He has published 60 papers in these areas.