Digital Signal Processing 102 (2020) 102741

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

An improved deep learning model for predicting stock market price time series

Hui Liu *, Zhihao Long

Institute of Artificial Intelligence and Robotics (IAIR), Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, Hunan, China

A R T I C L E I N F O

Article history:
Available online 17 April 2020

Keywords:
Stock market price
Deep learning network
Time series forecasting
Data processing
Error correction

A B S T R A C T

As an important component of the economic market, the stock market has been concerned by many researchers. How to get the trend of the stock market and predict the stock price is a problem that many researchers are studying. In previous works, the prediction methods are mainly focused on statistical models and traditional neural network models which are relatively popular in recent years. Deep learning is not often used in the field of financial time series, but it has a strong learning ability and is suitable for complex time series such as financial time series. In particular, the LSTM network has the function of long-term memory because of its cyclic structure, so it is very suitable for financial time series prediction in theory. In the study, a novel stock closing price forecasting framework is proposed, which has a higher prediction than traditional models. The data processing part, the deep learning predictor part, and the predictor optimization method are the components of this deep hybrid framework. Data processing includes empirical wavelet transform (EWT) based preprocessing and outlier robust extreme learning machine (ORELM) model based post-processing. Long short-term memory (LSTM) network based deep learning network predictor, as the main part of the mixed frame, is jointly optimized by dropout strategy and particle swarm optimization (PSO) algorithm. Each algorithm in the hybrid framework can give full play to its own functions to achieve better prediction accuracy. In order to verify the performance of the model, three challenging datasets are selected for forecasting experiments. Some comparative models are also selected to prove the effectiveness of the proposed framework. Experimental results show that the hybrid framework proposed in the study has the best prediction accuracy and can be applied to stock market monitoring or financial data analysis and research.

1. Introduction

As one of the important components of the economic market, stock has always been concerned by the government and many investors [1–4]. However, due to the randomness and instability of the stock market, the law of stock price changes is difficult to grasp. It is often difficult for the government to effectively regulate the stock market in a timely manner. Moreover, people usually rely on subjective judgment or blind purchases to conduct stock trading and lack effective methods of assisting decision-making. It is the goal that economists are striving to analyze and realize the rules of the stock market and improve the security and efficiency of financial investment. In recent years, with the long-term research and analysis of stock trends by economists and data analysts, the predictable components of stock prices have gradually been unearthed [5], and different stock price characteristics have been used to verify their predictability, including the opening price, lowest price, highest price, closing price, and transaction price. At the same time, the stock market can be regarded as a weather vane for the financial industry because it is inseparable from the financial market. Stock price forecasts are an effective means of establishing early warning systems (EWS). The stock market has always been influenced by the national economic situation, the perceptions of investors and political events [6], and the price series is highly nonlinear and non-stationary. The above factors have brought great challenges to stock price forecasts. So, it is meaningful to study how to build a universal and effective model to predict stock market prices.

1.1. Related works

In earlier years, trapped by the Efficient Markets Hypothesis (EMH), stock market analysis was based solely on economic values, and predictions of price indexes were stagnant. With the continu-

* Corresponding author.
E-mail address: csuliuhui@csu.edu.cn (H. Liu).

https://doi.org/10.1016/j.dsp.2020.102741

ous research and update of related researchers on the economic market and stock market theory, the predictable components of the stock price index were gradually exposed, and the prediction of stock price became possible. Stock price fluctuation is a random process to some extent. At present, the main methods used in stock index prediction are mathematical-statistical models [7,8] and the fuzzy theory. However, the predictive performance of traditional statistical models is limited by the statistical laws of the predicted target variables [9]. Therefore, researchers combined the fuzzy theory with statistical methods to obtain more in-depth information on stock price. In addition, machine learning methods have also been applied in stock price prediction, such as random forest (RF), decision tree (DT) and common artificial neural network (ANN) algorithm. Although the machine learning method has achieved a certain effect in predicting stock price, its prediction performance and ability to grasp information at a deeper level are general, which needs to be further improved.

In recent years, deep learning has become increasingly popular and has been applied in image processing [10–12], speech recognition [13–15], fault diagnosis [16–18], wind power prediction [19,20] and other fields. At present, the application of deep learning methods in the stock market is not many. Compared with the statistical method and machine learning method, the deep learning method has a more complex learning network structure. Relying on its complex hidden layer structure, deep learning methods can learn the hidden information in massive data more accurately. For the intricate price characteristics in the stock market, deep learning is bound to play a very good prediction effect. Selvin et al. [21] used three kinds of deep neural networks containing recurrent neural network (RNN), long short-term memory (LSTM) model and convolutional neural network (CNN) to establish a hybrid model to predict the stock price of a listed company. Ritika Singh et al. [22] uses the principal components analysis (PCA) and Deep Neural Network (DNN) method to predict the stock market and achieves better results than the traditional neural network. Akira Yoshihara et al. [23] proposes a stock market trend forecasting method based on a recurrent depth neural network and takes the real data of ten Nikkei companies as an example to verify the effectiveness of the method. However, the deep network often has the problem of over-fitting because of its strong learning ability. For this reason, the application of the dropout strategy to the LSTM network may help to reduce the over-fitting phenomenon of the LSTM network [24]. The effectiveness of the LSTM depth network in the field of stock price prediction is worthy of further exploration.

Practices have proved that the performance of a single model was generally poor, and it usually integrated pre-processing, post-processing and some optimization algorithms to improve the performance of the hybrid model, including decomposition algorithm [25,26], error correction model [27] and heuristic optimization algorithm, etc. Wavelet theory is widely used in data preprocessing methods. Duarte et al. [28] studied the application of wavelet transforms (WT) in different situations. Lahmiri et al. [29] used variational mode decomposition (VMD) and backpropagation neural network (BPNN) to predict daily stock market prices. Compared to WT and VMD, the EWT algorithm can adaptively divide the Fourier spectrum and selects the appropriate wavelet filter Banks. Therefore, the empirical wavelet transforms (EWT) algorithm has a better decomposition effect for complex stock market price series. As for data post-processing, a common method is error modeling and many existing studies focus on the vector error correction model [30–32]. The outlier robust extreme learning machine (ORELM) model may be a good choice for error modeling due to its good robustness and real-time performance. Therefore, this paper utilizes the powerful information mining ability of the deep neural network and integrates preprocessing, optimization algorithm and post-processing to build a hybrid model to achieve accurate prediction of the stock closing price.

In the study, daily stock closing price is predicted by a hybrid forecasting framework, which is established to predict daily stock closing prices in the United States and China. The EWT algorithm and ORELM model are proposed together for data processing, which includes data preprocessing based decomposition and post-processing based error correction. The PSO algorithm and dropout strategy are utilized for model optimization.

1.2. Novelty of the study

Summarizing the above references, the hybrid model often has better performance in predicting. In this paper, LSTM deep learning network is used to predict the daily stock closing price series. The proposed hybrid framework is combined with four parts: EWT, dpLSTM, PSO and ORELM. The details of proposed framework can be described as follows: (a) the EWT algorithm which can divide raw stock closing price into serval sub-layers, is used for data preprocessing making training data more stable and regular; (b) the dropout strategy is employed for optimizing the training process of LSTM deep network and improving the generalization ability of the model; (c) PSO algorithm is utilized to choose appropriate hyperparameters of LSTM network; (d) ORELM is used for error correction based on previous forecasting results for each sub-layer.

The main contributions of the study can be summarized as follows:

(a) A novel hybrid deep learning framework is proposed for forecasting daily stock closing prices in the financial time series prediction field. The proposed hybrid framework consists of raw stock closing price series decomposition, the deep learning network with dropout strategy based forecasting computation and PSO algorithm based parameter optimization. The purpose of the proposed model is to combine the capacity of involved algorithms to realize the high-accuracy daily stock closing price prediction. Moreover, the proposed framework is flexible and extensible, which can provide some reference for other works. Experimental results from three data sets show that the proposed hybrid framework can effectively predict daily stock closing prices.

(b) EWT based decomposition and ORELM based error correction are employed together for data processing. Because the financial time series is complex and disorderly, direct modeling of the original data is often ineffective. The EWT decomposition turns the complex original financial sequence into several more stable sub-layers and the ORELM model is utilized for error correction to improve forecasting accuracy further. Experimental results show that EWT decomposition solves the problem of error superposition which may be caused by error modeling, and the ORELM model is used for error modeling to solve the problem of low prediction accuracy of low-frequency components.

(c) Dropout strategy and PSO algorithm are used to optimize LSTM deep learning predictor jointly. The dropout strategy solves the overfitting problem easily encountered by the deep network, and the PSO optimization algorithm selects the better super parameters of the LSTM. The combination of the two guarantees the precision and stability of the deep network predictor. In addition, this paper studies the relationship between dropout rate and the number of LSTM neurons, and the results show that the number of neurons corresponding to the optimal effect of dropout strategy is mostly identical to the optimal parameters of PSO algorithm, which reflected the superiority of the mixed frame in this paper.

2. Methodology

2.1. Empirical wavelet transform decomposition

In this paper, the EWT algorithm is used for data preprocessing. The EWT proposed by *Gilles* [33] is a novel signal processing technique, which builds the wavelets adaptively. The EWT is based on the theoretical framework of wavelet transform but overcomes the shortage of EMD theory and the problem of signal aliasing. The EWT adaptively divides the Fourier spectrum and selects the appropriate wavelet filter Banks. The empirical scaling function and the empirical wavelets can be expressed as Equations (1) and (2):

$$\hat{\phi}_n(\omega)=\begin{cases}1 \quad \text{if}|\omega| \le (1-\gamma)\omega_n \\ \cos[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_n}(|\omega|-(1-\gamma)\omega_n))] \\ \qquad \text{if}(1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n \\ 0 \quad \text{otherwise}\end{cases} \tag{1}$$

$$\hat{\psi}_n(\omega)=\begin{cases}1 \quad \text{if}(1+\gamma)\omega_n \le |\omega| \le (1-\gamma)\omega_{n+1} \\ \cos[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_{n+1}}(|\omega|-(1-\gamma)\omega_{n+1}))] \\ \qquad \text{if}(1-\gamma)\omega_{n+1} \le |\omega| \le (1+\gamma)\omega_{n+1} \\ \sin[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_n}(|\omega|-(1-\gamma)\omega_n))] \\ \qquad \text{if}(1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n \\ 0 \quad \text{otherwise}\end{cases} \tag{2}$$

where the approximation coefficients and the detail coefficients are computed according to Equations (3) and (4).

$$\begin{aligned} w_f^{\varepsilon}(0,t) &= \langle f,\phi_1\rangle = \int f(\tau)\overline{\phi_1(\tau-t)}d\tau \\ &= \left(\hat{f}(\omega)\overline{\hat{\phi}_1(\omega)}\right)^{\vee} \end{aligned} \tag{3}$$

$$\begin{aligned} w_f^{\varepsilon}(n,t) &= \langle f,\psi_n\rangle = \int f(\tau)\overline{\psi_n(\tau-t)}d\tau \\ &= \left(\hat{f}(\omega)\overline{\hat{\psi}_n(\omega)}\right)^{\vee} \end{aligned} \tag{4}$$

In addition, there are some constraints to keep the effectiveness of the proposed algorithm. For example, the ratio γ in Equations (1) and (2) is restricted to a small value as $\gamma < \min_n(\omega_{n+1}-\omega_n)/(\omega_{n+1}+\omega_n)$ to ensure the empirical scaling function and the empirical wavelets is a tight frame of $L^2(R)$. The $\beta(x)$ is usually defined as $\beta(x)=x^4(35-84x+70x^2-20x^3)$.

2.2. Long short-term memory network with dropout strategy

In the machine learning model, the trained model is prone to overfitting because of too many parameters of the model and too few training samples. Overfitting is a common problem in many machine learning systems. In order to solve this problem, the ensemble model is generally adopted, that is, multiple models are trained for combination. However, training and testing multiple models is time-consuming. Based on this consideration, the dropout strategy was first proposed by researchers in 2014 [24].

Nowadays, a dropout strategy is more likely to be used in the traditional forward propagation neural network. The dropout strategy could not play a good role in the RNN structure because of its special communication mode. The special gate structure in LSTM provides the possibility to use a dropout strategy. The dropout in LSTM cannot occur in the hidden layer like the traditional neural network, otherwise, the information message and gradient will disappear. In the study, the dropout strategy is used in the information transmission process of neurons at different moments. LSTM network with dropout strategy can be described as follows [34]:

(a) Generate a set of random vectors with values of 0 or 1, which is used to randomly delete some status information.

$$r \sim Bernoulli(p) \tag{5}$$

where p is dropout rate, r is random vectors.
(b) Calculate the value of the input gate i_t and the candidate state value of the cell input $\tilde{C}_t$. The state weight of the previous neuron should be dropout when entering the current input gate, which can be shown as follows:

$$i_t=\delta\big(W_i*(X_t,h_{t-1}*r)+b_i\big) \tag{6}$$

(c) Calculate the value of the input gate i_t and the candidate state value of the cell input $\tilde{C}_t$ at time t, and the formula is as follows:

$$i_t=\delta\big(W_i*(X_t,h_{t-1})+b_i\big) \tag{7}$$

$$\tilde{C}_t=\tanh\big(W_c*(X_t,h_{t-1})+b_c\big) \tag{8}$$

(d) The activation value of the forgetting gate f_t at time t is calculated as follows:

$$f_t=\delta\big(W_f*(X_t,h_{t-1})+b_f\big) \tag{9}$$

(e) The state value of the cell input C_t at time t can be obtained by results on (a) and (b).

$$C_t=i_t*\tilde{C}_t+f_t*C_{t-1} \tag{10}$$

(f) The value of the output gate can be gotten by follows:

$$O_t=\delta\big(W_o*(X_t,h_{t-1})+b_o\big) \tag{11}$$

$$h_t=O_t*\tanh(C_t) \tag{12}$$

(g) When the training at this moment is completed, the deleted status information needs to be restored. Then repeat the above process for the next iteration.

2.3. Variants of extreme learning machines

2.3.1. Extreme learning machine

Extreme learning machine (ELM) is a new learning algorithm based on a single hidden layer feedforward neural network, which was proposed by Huang et al. [35]. The structure of ELM is similar to that of Single Hidden Layer Feedforward Neural Networks (SLFNs). The training process of ELM is simple. The input weights and thresholds are randomly generated. The model training can be completed by using the generalized inverse matrix to obtain the output weight of SLFNs. As a result, the training speed and generalization ability are greatly improved.

2.3.2. Regularized extreme learning machine

The traditional ELM model ignores the structural risk. It is only based on the principle of empirical risk minimization and only considers empirical risk. Therefore, in practice, the traditional ELM model cannot be adjusted according to the characteristics of the data set, and the performance is poor. Based on this consideration, the regular extreme learning machine (RELM) introduces the principle of structural risk minimization on ELM by regularization factor, which was proposed by Deng et al. [36]. This is also the biggest difference between the regular extreme learning machine and the traditional extreme learning machine.

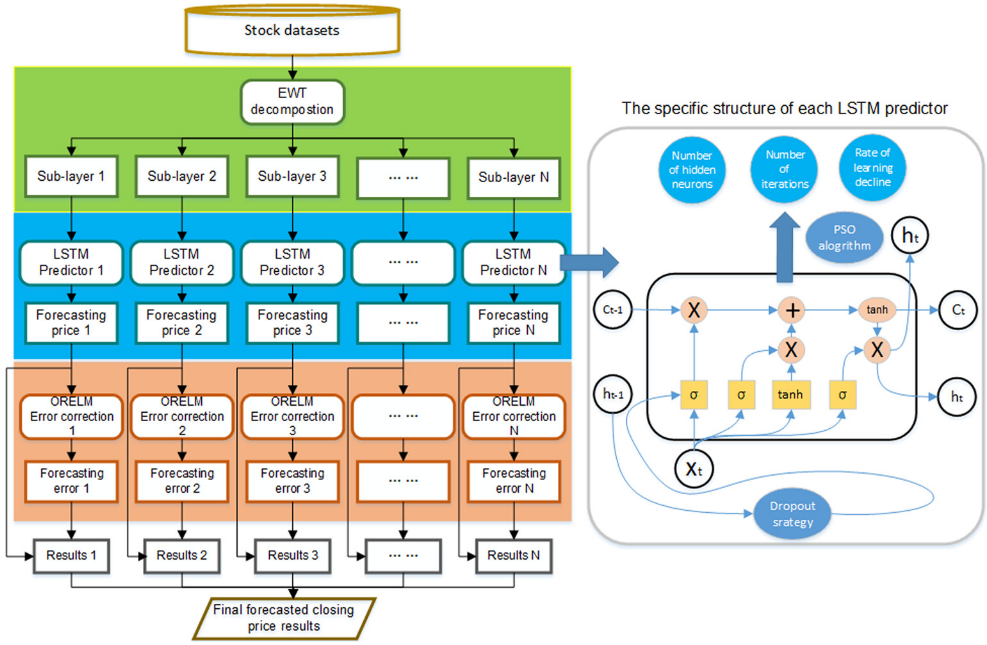


Fig. 1. The framework of the proposed model.

2.3.3. Weighted regularized extreme learning machine

Compared with the traditional ELM model, the weighted regularization extreme learning machine (WRELM) not only introduces the regularization factor but also adds the weight factor. The weight factor weights the error items in the objective function of the RELM model so that the training samples similar to the specific prediction objects can obtain higher accuracy in the training process [37].

2.3.4. Outlier robust extreme learning machine

According to the compressive sensing theory and the robustness analysis theory, ORELM effectively improves the robustness of the model by changing the norms of ELM in the training error of the solution model. It reduces the prediction accuracy decline problem caused by data fluctuation and the over-fitting of ELM training. The training process of the ORELM model is a process for solving the following linear system [38].

$$H\beta = Y \tag{13}$$

$$H = \begin{bmatrix} g(W_1 \bullet X_1 + b_1) & \cdots & g(W_L \bullet X_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(W_1 \bullet X_N + b_1) & \cdots & g(W_L \bullet X_N + b_L) \end{bmatrix}_{N\times L} \tag{14}$$

$$\beta = [\beta_1^T \cdots \beta_L^T]^T_{L\times m} \tag{15}$$

$$Y = [y_1^T \cdots y_L^T]^T_{N\times m} \tag{16}$$

where H is the output matrix of the hidden layer. β is the training target weight matrix. Y is the expected output.

In order to improve the robustness of the prediction model, the norm of the training error is replaced as follows:

$$\min_{\beta} \|e\|_1 + \frac{1}{c}\|\beta\|_2^2 \tag{17}$$

where $e = y - H\beta$ is training error. c is the regularization parameter.

Eventually, the output matrix β is obtained. The forecasting value can be circulated by $\hat{y} = H\beta$.

2.4. The framework of proposed methods

The structure of the proposed EWT-dpLSTM-PSO-ORELM hybrid framework is shown in Fig. 1. The specific details can be described as follows:

(a) The raw closing price time series data will be decomposed into several sub-layers by the EWT algorithm. The details of the EWT decomposition are described in Section 2.1.

(b) The sub-layers are divided into the training set 1 and the testing set 1. The data in each data set are used to construct the input and output according to the method explained in Section 3.1.

(c) LSTM models are built based on the training set 1 to forecast the decomposed closing price time series data; the constructed LSTM models are tested on the testing set 1. Each decomposition sub-layer is used to train an LSTM model accordingly. The details of the LSTM are described in Section 2.2.

(d) The dropout strategy is used to optimize the training process of the LSTM model. The values of optimal dropout are selected in Section 4.3.1 by experimental analysis. The details of the dropout strategy are given in Section 2.2.

(e) PSO algorithm is used to optimize the parameters of the LSTM model. The best value of the number of neurons in the hidden layer, the number of LSTM network iterations and the rate of learning decline will be calculated.

(f) ORELM method is utilized for error correction. The error series of forecasting results in step (c) is divided into training set 2 and testing set 2 further. The training set 2 is used for building the ORELM model. The testing set 2 is utilized for model validation. The details of the ORELM algorithm are given in Section 2.3.

(g) The forecasting results obtained in step (c) and the forecasting errors obtained in step (f) are added up to get an accurate prediction of closing price. Then the closing price predictions for all the sub-levels are added together to get the final forecasting results.

(h) The horizon of daily ahead predictions is investigated in the study. To investigate the prediction performance of the proposed EWT-dpLSTM-PSO-ORELM model, nine other prediction models are provided as the comparison models, which comprise of the single LSTM model, single BP model, dpLSTM model, dpLSTM-PSO model, EWT-dpLSTM-PSO model, dpLSTM-PSO-ORELM model,

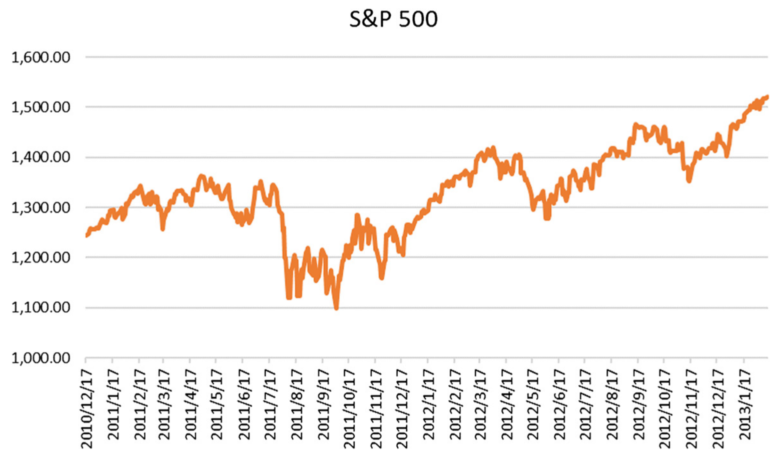


Fig. 2. Original data of the S&P 500 from 17 Dec 2010 to 17 Jan 2013.

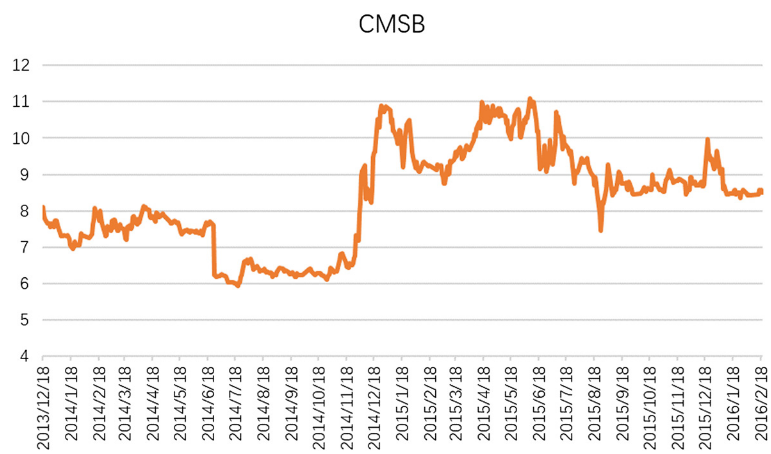


Fig. 3. Original data of CMSB from 18 Dec 2013 to 18 Jan 2016.

dpLSTM-PSO-RELM, dpLSTM-PSO-WRELM model and dpLSTM-PSO-ELM model.

3. Experimental setup and results

3.1. Data preparation and phase-space reconstruction

This paper is aimed to predict the day-ahead stock close price. The numbers of 10 historical samples are inputs of the trained model and the stock close price of the next day is output. Then the rolling recursive strategy is used to achieve the processing of training and test. Before experiments, the original series data will be transformed into an $11 \times N$ matrix by the phase space reconstruction, where N is the number of samples.

There are three experimental datasets set to ensure that the model was scalable. The dataset #1 is selected from the Standard and Poor's 500 Index (S&P 500) from 17 Dec 2010 to 17 Jan 2013 (totally 541 samples). Dataset #2 comes from China Minsheng Bank (CMSB) from 18 Dec 2013 to 18 Jan 2016 (totally 528 samples). The dataset #3 comes from Dow Jones Industrial Average (DJI) from 17 Dec 2014 to 17 Jan 2017 (totally 543 samples). Those datasets include all of stock close price among the range of selected time. The last 30 samples are utilized to test trained models in each dataset and others are used to be training sets.

Figs. 2–4 show the distribution of the three experimental datasets. It can be seen that the three datasets include the rise, inflation, decline, plunge, smooth and other changes, which pose a huge challenge to the proposed forecasting model. Moreover, the research data cover domestic and foreign stock data from 2011 to 2016, which can well illustrate the practicability of the proposed model. In order to be on the same order of magnitude as #2, #1 will be scaled down 100 times and #3 will be scaled down 1000 times in the experiments.

3.2. The set of proposed methods

This section gives some necessary parameters of the model in experiments. The default cell numbers of LSTM are 30, the number of iterations is 60 and the rate of learning decline is 0.001. The traditional gradient descent algorithm is used in the processing of training without the backpropagation process. The default cell numbers of ORELM are 20, the Punish coefficient C equals to 2^{30}. As to the PSO algorithm, the optimization objectives include the number of neurons in the hidden layer, the number of LSTM network iterations and the rate of learning decline. The optimization range is 10∼200, 20∼150 and 0.001∼0.01 for each parameter to be optimized. The dropout rate is 0.4 and the reason will be shown in Section 4.3.1.

3.3. Evaluation indicators

In this study, four statistical evaluation indicators are utilized to compare performance of related models, including the MAE (Mean Absolute Error), the MAPE (Mean Absolute Percentage Error), the RMSE (Root Mean Square Error) and the SDE (Standard Deviation of Error), which can be calculated as follows:

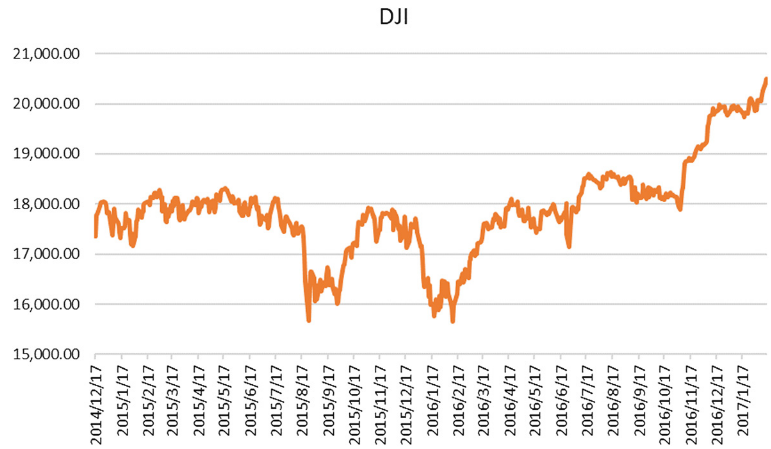


Fig. 4. Original data of DJI from 17 Dec 2014 to 17 Jan 2017.

$$MAE = \left(\sum_{t=1}^{N} \left| x(t) - \hat{x}(t) \right| \right) / N \tag{18}$$

$$MAPE = \left(\sum_{t=1}^{N} \left| \left(x(t) - \hat{x}(t)\right) / x(t) \right| \right) / N \tag{19}$$

$$RMSE = \sqrt{\left(\sum_{t=1}^{N} \left[x(t) - \hat{x}(t)\right]^2\right) / N} \tag{20}$$

$$SDE = \sqrt{\left(\sum_{t=1}^{N} \left[x(t) - \hat{x}(t) - \sum_{t=1}^{N} \left(x(t) - \hat{x}(t)\right) / n\right]^2\right) / n} \tag{21}$$

where N is the number of samples, $x(t)$ is the actual close price, $\hat{x}(t)$ is predicted close price.

In particular, the values of the above statistical indicators are related to the sample units. In order to facilitate the performance comparison of the model, this paper takes the data unit of CMBC as the standard and unifies the result indexes of other experimental data sets and those of other papers.

3.4. Numerical experiment

Three datasets #1∼#3, which is described in Section 3.1, are utilized for forecasting experiments. Daily stock close price is predicted by proposed models. Data between two years is used for the training set and data of firstly 30 days in next year is used for the testing set.

Some classic researches of stock market forecasting are selected and compared with the proposed model. Sun et al. [39] employed an improved backpropagation (BP) neural network to forecast the stock market. Basak et al. [40] uses the random forest algorithm integrated by the decision tree and selects the technical index as the input feature to forecast the stock market. Roondiwala et al. [41] utilized a single LSTM deep learning model for stock market prediction. The selected comparison models include classical neural network model, machine learning algorithm and deep learning algorithm, which is representative. As more, six benchmark models are selected by combining different algorithms to illustrate the superiority of the proposed hybrid model, including the dpLSTM-PSO model, EWT-dpLSTM-PSO model, dpLSTM-PSO-ORELM model, dpLSTM-PSO-RELM model, dpLSTM-PSO-WRELM model, and dpLSTM -PSO-ELM model. The last four models are selected to analyze the prediction accuracy of ORELM compared with other common variants of the ELM model, in order to prove the effectiveness of the ORELM model from the data level. In particular, all the above models are programmed based on MATLAB.

3.5. Forecasting results

Figs. 5–7 show the results of three datasets predicted by the proposed model. It can be simply seen from Figs. 5–7 that the proposed model has the same excellent prediction performance in the three data sets.

4. Comparison and analysis

4.1. Comprehensive analysis

Some classical researches are used to compare with the model proposed in this paper. Table 1 compares the pros and cons of this work. Tables 2–4 compare the accuracy of the model proposed in the study with that of these works.

From Tables 1–4, it can be seen that the proposed model has the best forecasting accuracy. In order to compare the prediction accuracy of each model more concretely, the evaluation indicators of other models mentioned in Section 3.3 are listed in Tables 5–7.

From Figs. 5–7 and Tables 5–7, some conclusions can be summarized as follows:

(a) The proposed model has the best performance in three experiments. These results strongly prove the accuracy, stability, and extensibility of the proposed model.

(b) Comparing to the traditional BP network, LSTM has higher accuracy. For instance, the MAPE, MAE, RMSE and SDE of BP for #1 is 1.3501%, 0.2029, 0.2427 and 0.1454, respectively. The MAPE, MAE, RMSE and SDE of LSTM for #1 is 0.3332%, 0.0496, 0.0675 and 0.0634, respectively. This shows that the deep learning strategy can be effectively applied in the field of financial time series.

(c) The dropout strategy and PSO algorithm can improve the performance of single LSTM. More specifically, the reasons are discussed later in Section 4.3.1

(d) The EWT-dpLSTM-PSO model is failed to achieve higher forecasting accuracy than the dpLSTM-PSO model. The reasons are discussed later in Section 4.2.

(e) The dpLSTM-PSO-ORELM is also failed to achieve higher forecasting accuracy than the dpLSTM-PSO model. The reasons are discussed later in Section 4.4.

(f) The dpLSTM-PSO-ORELM model has a better performance than the dpLSTM-PSO-ELM model, dpLSTM-PSO-RELM model, and dpLSTM-PSO-WRELM model.

(g) When EWT and error correction by ORELM are all combined with the dpLSTM-PSO model, the prediction accuracy of the proposed model will be improved remarkably.

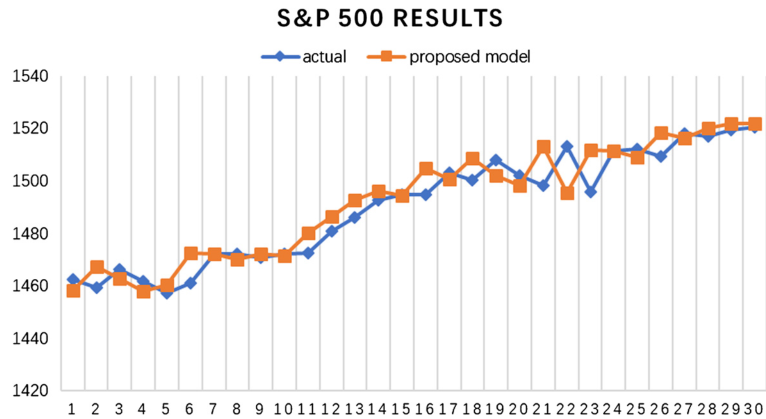


Fig. 5. Forecasting results of proposed models for S&P 500.

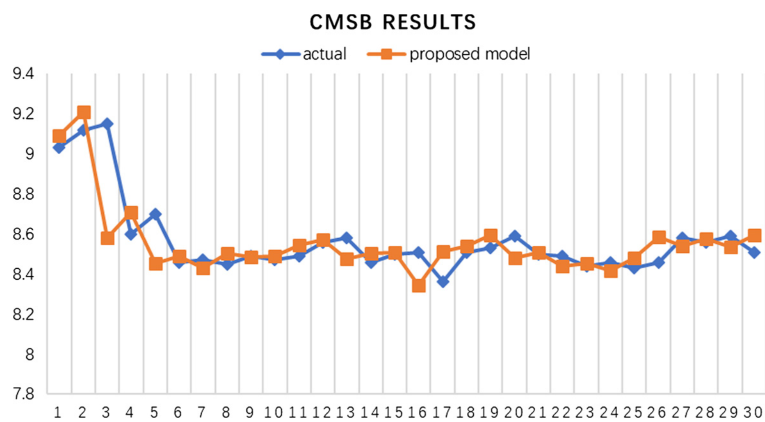


Fig. 6. Forecasting results of proposed models for CMSB.

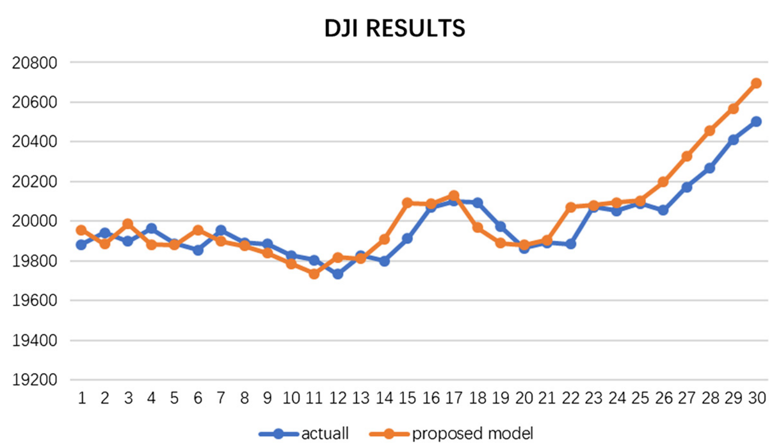


Fig. 7. Forecasting results of proposed models for DJI.

4.2. Analysis of the EWT decomposition

Tables 5–7 in Section 4.1 shows that using a deep learning network for prediction after EWT decomposition alone cannot improve the accuracy of prediction results. Specific decomposition predictions will be used for analyzing reasons.

Figs. 8–10 show all sub-layers obtained by the EWT algorithm. As can be seen from the figures, compared with the original financial time series, the time series of each sub-layer after decomposition is more stable and more regular.

In accordance with Tables 5–7 and Figs. 8–10, it can be found that:

(a) Raw stock close price data #1 is divided into 14 sub-layers. Raw stock close price data #1 is divided into 10 sub-layers. Raw stock close price data #1 is divided into 11 sub-layers. This shows the adaptability of the EWT decomposition algorithm.

(b) Compared with the original financial time series, the time series of each sub-layer after decomposition is more stable and more regular. From the perspective of input data, it is possible to obtain better training results by using the decomposed sub-layer data for model training.

(c) It can be seen from the figures that the main error source of EWT decomposition is the error of low-frequency components (S1∼S3). The prediction results of the sub-layer of the main trend

Table 1
The main disadvantages of previous research works.

Name	Published year	Major contributions	Disadvantages compared with the proposed model
Sun's model [39]	2015	• An improved BPNN is employed for stock market price forecasting	• Data pre-processing and post-processing are ignored, only model optimization is carried out
Roondiwala's model [41]	2017	• The LSTM deep learning method is utilized for stock market price forecasting	• The over-fitting effect of deep learning is ignored • The data processing process is ignored
Basak's model [40]	2019	• The RF model is utilized for stock market forecasting • Reasonable medium-and long-term forecasting indicators with higher precision are selected	• The input of the model is selected based on experience, and the theory is not strong

Table 2
Evaluation indicators of previous research works for #1.

Name	MAPE (%)	MAE	RMSE	SDE
Sun's model [39]	1.3501	0.2029	0.2427	0.1454
Roondiwala's model [41]	4.3375	0.5800	0.7270	0.7070
Basak's model [40]	0.3332	0.0496	0.0675	0.0634
Proposed model	**0.0410**	**0.0061**	**0.0075**	**0.0072**

Table 3
Evaluation indicators of previous research works for #2.

Name	MAPE (%)	MAE	RMSE	SDE
Sun's model [39]	1.0274	0.0879	0.1464	0.1433
Roondiwala's model [41]	14.7578	1.3415	1.5689	1.2111
Basak's model [40]	0.5446	0.0473	0.0774	0.0720
Proposed model	**0.1552**	**0.0134**	**0.0176**	**0.0176**

Table 4
Evaluation indicators of previous research works for #3.

Name	MAPE (%)	MAE	RMSE	SDE
Sun's model [39]	0.6078	0.1221	0.1475	0.1194
Roondiwala's model [41]	4.1285	0.7372	0.9994	0.9994
Basak's model [40]	0.4471	0.0890	0.1081	0.0917
Proposed model	**0.1071**	**0.0215**	**0.0296**	**0.0252**

Table 5
Evaluation indicators of all related models for #1.

Models	MAPE (%)	MAE	RMSE	SDE
dpLSTM	0.2952	0.0439	0.0580	0.0562
dpLSTM-PSO	0.2082	0.0308	0.0433	0.0362
EWT-dpLSTM-PSO	0.7830	0.1172	0.1403	0.0892
dpLSTM-PSO-ORELM	1.1635	0.1738	0.1955	0.0975
dpLSTM-PSO-RELM	0.6562	0.0983	0.1155	0.0942
dpLSTM-PSO-ELM	0.4582	0.0682	0.0844	0.0807
dpLSTM-PSO-WRELM	1.6955	0.2550	0.3211	0.2114
PROPOSED MODEL	**0.0410**	**0.0061**	**0.0075**	**0.0072**

Table 6
Evaluation indicators of all related models for #2.

Models	MAPE (%)	MAE	RMSE	SDE
dpLSTM	0.5298	0.0458	0.0701	0.0619
dpLSTM-PSO	0.3894	0.0338	0.0538	0.0505
EWT-dpLSTM-PSO	0.9723	0.0845	0.1293	0.1293
dpLSTM-PSO-ORELM	1.2707	0.1087	0.1518	0.1518
dpLSTM-PSO-RELM	1.4617	0.1249	0.1744	0.1742
dpLSTM-PSO-ELM	1.4368	0.1227	0.1822	0.1765
dpLSTM-PSO-WRELM	1.3875	0.1186	0.1630	0.1626
PROPOSED MODEL	**0.1552**	**0.0134**	**0.0176**	**0.0176**

Table 7
Evaluation indicators of all related models for #3.

Models	MAPE (%)	MAE	RMSE	SDE
dpLSTM	0.2817	0.0564	0.0717	0.0612
dpLSTM-PSO	0.1514	0.0304	0.0385	0.0293
EWT-dpLSTM-PSO	0.4682	0.0939	0.120	0.1146
dpLSTM-PSO-ORELM	0.4383	0.0877	0.1002	0.0936
dpLSTM-PSO-RELM	0.4961	0.099	0.1218	0.1069
dpLSTM-PSO-ELM	0.4456	0.0892	0.1056	0.1027
dpLSTM-PSO-WRELM	0.3893	0.0779	0.0967	0.0967
PROPOSED MODEL	**0.1071**	**0.0215**	**0.0296**	**0.0252**

decomposition of the improved LSTM model are seriously deviated, which leads to the failure of the decomposition algorithm to improve the prediction accuracy of the original model. However, for the other decomposition sub-layers, the decomposition prediction is excellent.

(d) The reason for the gross prediction error of the main trend layer may be the irregularity of the main trend formation. The main trend layer may contain annual law data, so the training samples within two years cannot give a good training effect on the training model, resulting in low prediction accuracy. The possible solutions are to increase the time span of training samples or further data processing. From the perspective of benefit, this paper uses error modeling to solve this problem.

In order to better highlight the advantages of EWT algorithm, other traditional decomposition algorithms based on wavelet theory are used as comparative experiments, such as the wavelet decomposition (WD) [42], maximal overlap discrete wavelet transform (MODWT) [43] and wavelet packet decomposition (WPD) [44]. Dataset #1 is used as this part of the comparison experiment.

In this experiment, the wavelet mother function of all WA methods is selected as db6. Referring to the general literature [42–45], the number of decomposition layers is set to three layers. So, the original stock market price series is decomposed into 8 sub-layers by WPD and EWT, which are named as S1∼S8 in this experiment. The original stock market price series is decomposed into 4 sub-layers by WD and MODWT, which are named as S1∼S4 in this experiment. From the decomposition results of each sub-

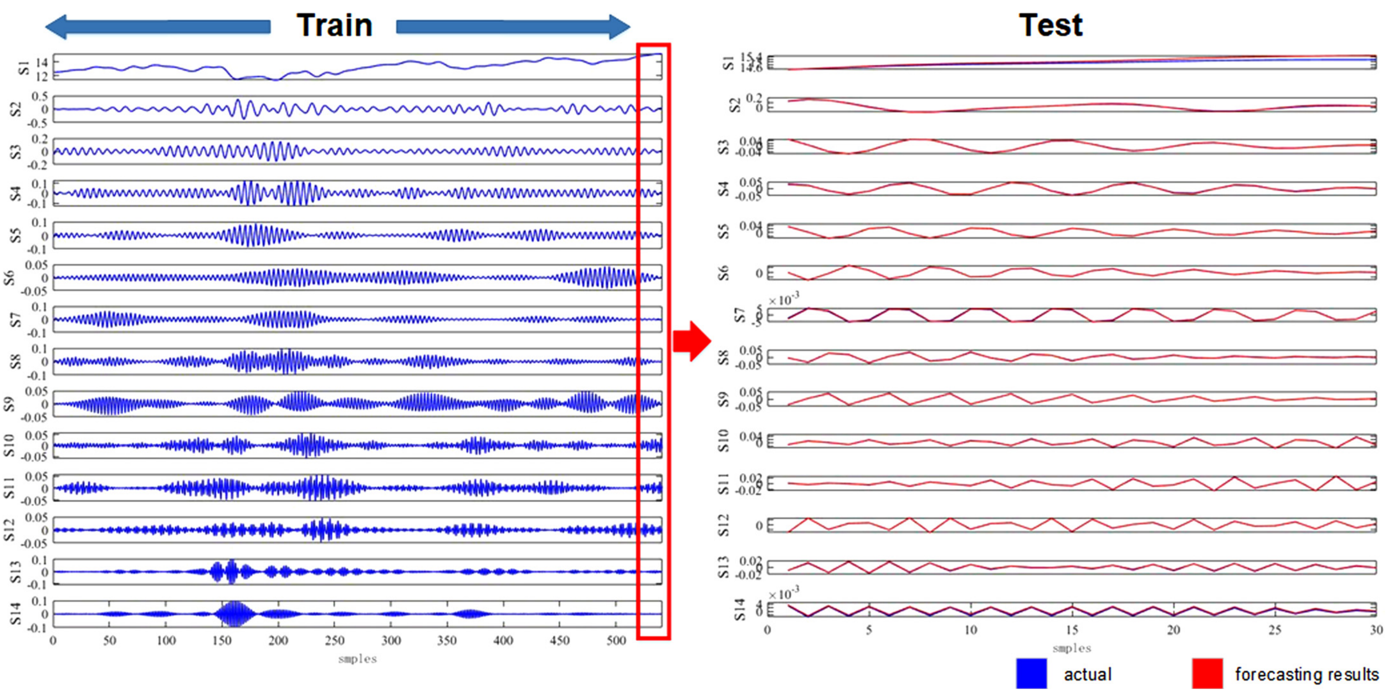


Fig. 8. Results of EWT decomposition and forecasting results for each sub-layer by EWT-dpLSTM-PSO for #1.

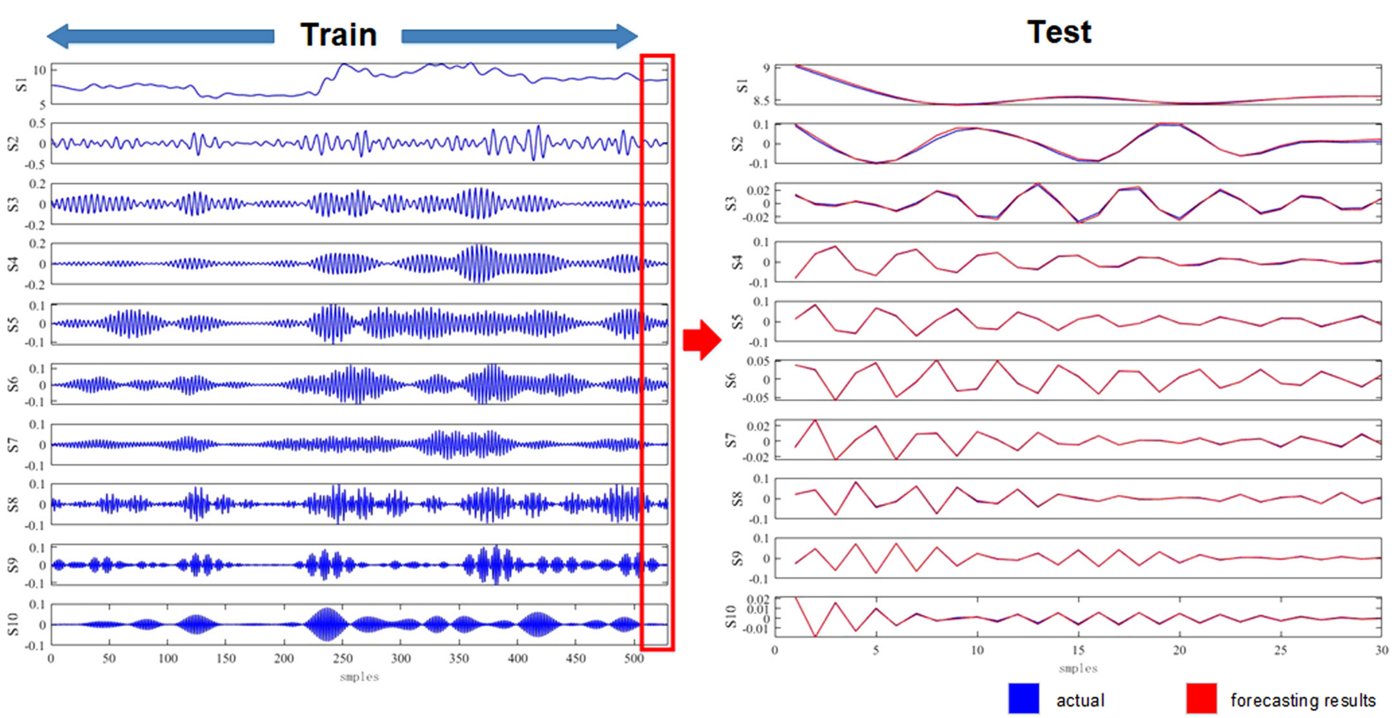


Fig. 9. Results of EWT decomposition and forecasting results for each sub-layer by EWT-dpLSTM-PSO for #2.

layer in Fig. 11, it can be seen that the decomposition results of WPD and EWT algorithms are obviously better than those of WD and MODWT algorithms. Compared with the decomposition results of the WPD algorithm, the sub-layers obtained by the EWT algorithm has better symmetry and compact support, and there is almost no signal aliasing. Generally speaking, because EWT is adaptive, there are more sublayers of decomposition, and the decomposition effect of each subsequence is the best of the four methods.

4.3. Analysis of the improved LSTM

4.3.1. Comparison of dropout strategy

From the comprehensive analysis in Section 4.1, it can be seen that the dropout strategy can improve the prediction accuracy of the LSTM model to some extent, but the optimization effect of this method is affected by some parameters. This section will carry out further experimental discussions about whether it can play a benign role under various parameter conditions.

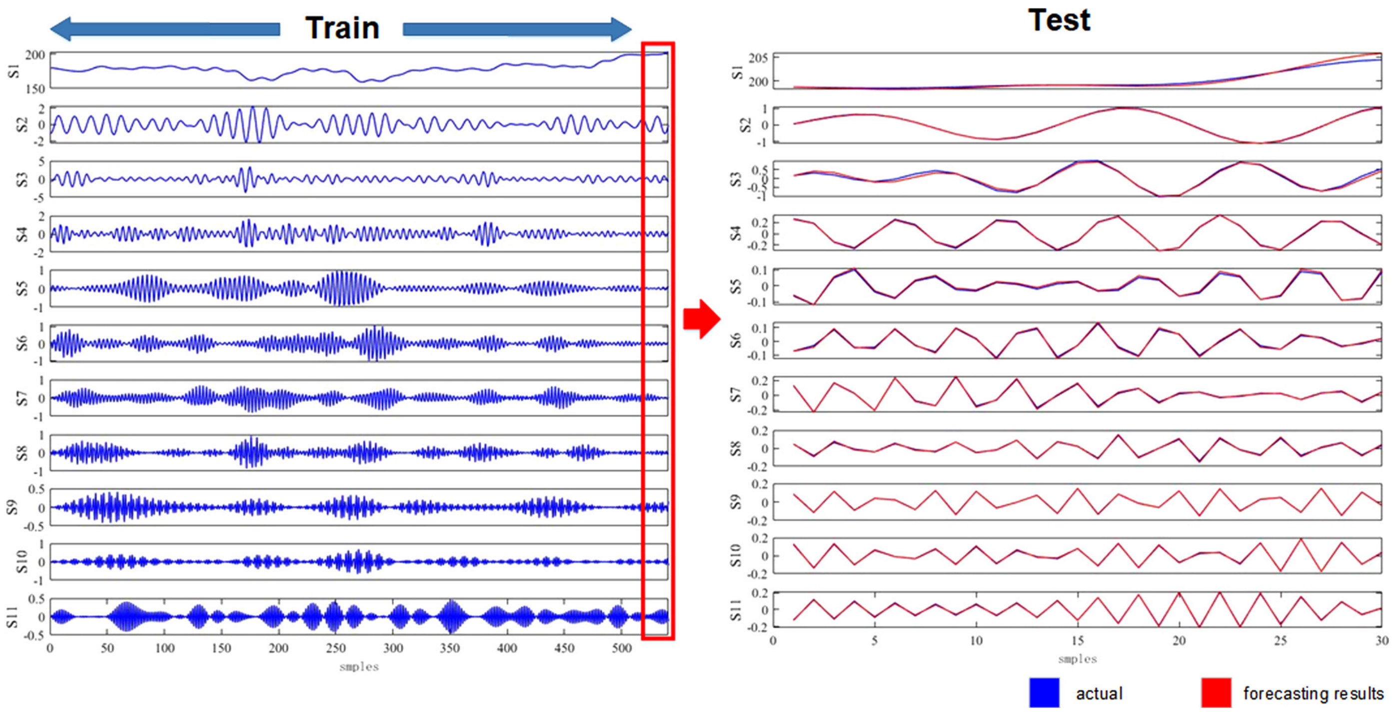


Fig. 10. Results of EWT decomposition and forecasting results for each sub-layer by EWT-dpLSTM-PSO for #3.

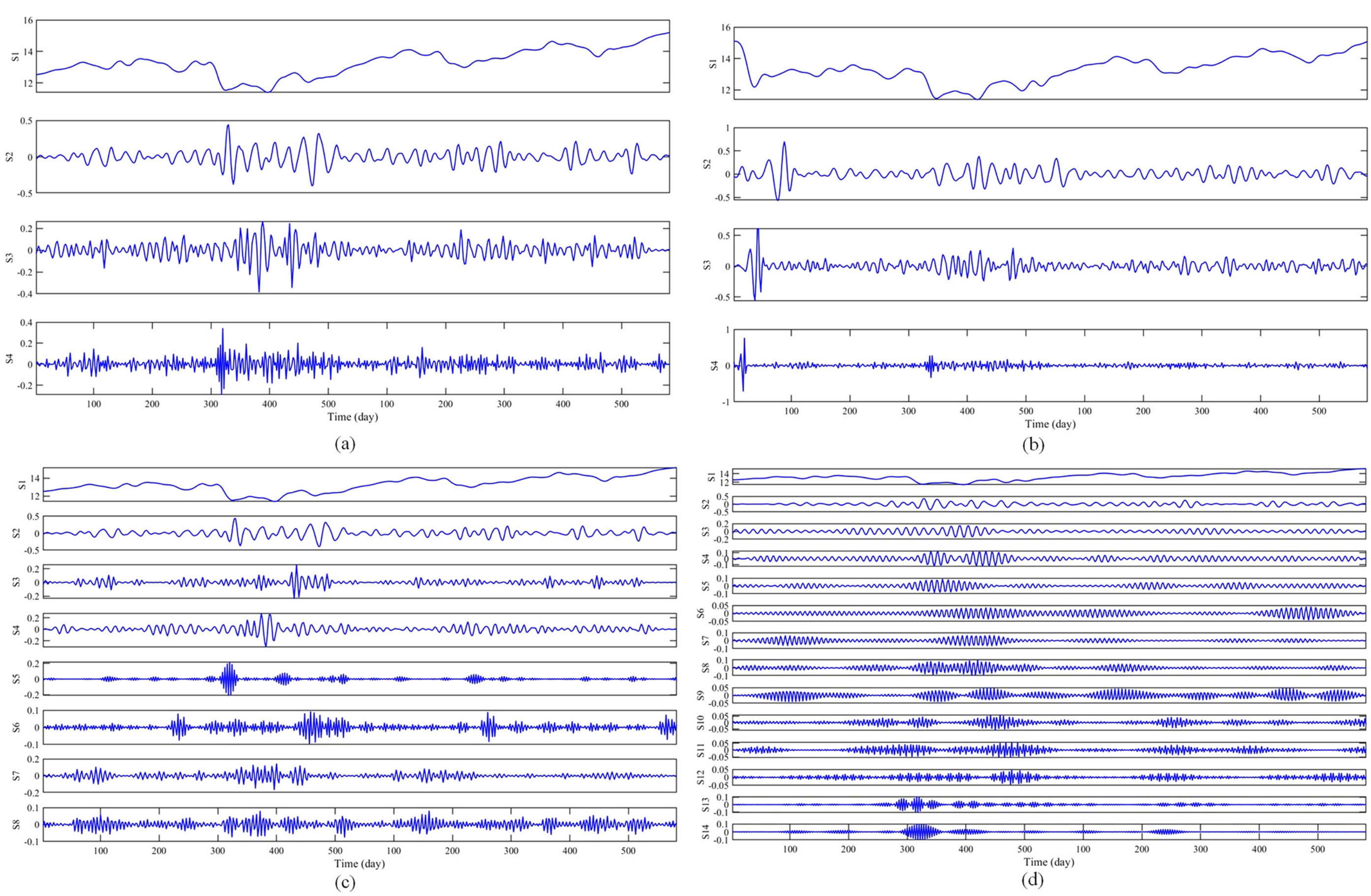


Fig. 11. (a) Decomposition results of WD; (b) Decomposition results of MODWT; (c) Decomposition results of WPD; (d) Decomposition results of EWT.

Table 8
MAPE values under different dropout rates and the number of neurons.

Dropout rates	The number of neurons	MAPE	The number of neurons	MAPE	The number of neurons	MAPE
0.2	30	0.3038	60	0.2891	100	0.7298
	150	0.2428	200	0.6624	300	1.4552
0.4	30	0.2690	60	0.2086	100	0.6846
	150	0.2529	200	0.5971	300	1.3907
0.6	30	0.3896	60	0.2849	100	0.4740
	150	0.3881	200	0.8049	300	1.4007
0.8	30	0.6310	60	0.3607	100	0.446
	150	0.2610	200	0.8573	300	2.3457

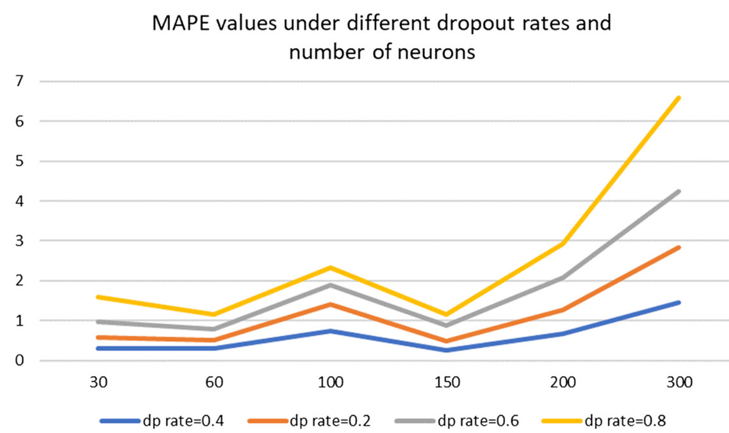


Fig. 12. MAPE values under different dropout rates and the number of neurons.

The effects of different dropout rates and the number of neurons in hidden layers on the dropout optimization effect will be analyzed in this section. The dropout rates are selected as 0.2, 0.4, 0.6 and 0.8. The number of neurons in hidden layers is selected as 30, 60, 100, 150, 200 and 300. The improved percentage of MAPE for dpLSTM relative to the LSTM model is used to evaluate the dropout strategy.

Table 4 shows the specific experimental results. #1 is utilized for this experiment. Fig. 11 shows the experimental results more intuitively with a broken line diagram.

From Table 8 and Fig. 12, it can be concluded that:

(a) In the experimental range, there are two optimal values for the number of hidden layer neurons. On these two optimal values, a dropout strategy can achieve a better training effect. The two optimal values are 60 and 150, respectively.

(b) The optimization effect of dropout strategy will first decrease and then increase with the increase of dropout rate, and the best optimization effect can be achieved when the dropout rate is 0.4.

(c) The dropout strategy is mainly an optimization strategy oriented to the number of neurons, and the relationship with the number of iterations and the learning decline rate is relatively weak. Therefore, this experiment does not study the relationship between them. This can be left for later work. This paper is the first to study the relationship between dropout rate and parameters of deep learning network, which has some enlightenment.

4.3.2. Analysis of the PSO

In order to better illustrate the optimization effect of PSO on the training process of the LSTM model, the training process errors of the LSTM model and PSO-LSTM model are intuitively displayed. Fig. 13 shows different training errors of the LSTM model and PSO-LSTM model.

Table 9
The improved percentage of dpLSTM-PSO to dpLSTM.

dpLSTM-PSO v.s. dpLSTM				
datasets	$\mathbf{P}_{\text{MAPE}}$ (%)	$\mathbf{P}_{\text{MAE}}$ (%)	$\mathbf{P}_{\text{RMSE}}$ (%)	$\mathbf{P}_{\text{SDE}}$ (%)
#1	29.4954	29.8405	25.3448	35.5872
#2	26.5006	26.2008	23.2525	18.4168
#3	46.2549	46.0993	46.3040	52.1242

As can be seen from Fig. 13, the training error of the LSTM model selected according to experience increases all the time and finally converges to 0.05. After using the PSO algorithm to select hyperparameters, the training error of the LSTM model decreases after rising rapidly, and finally approaches to 0.0125. Thus it can be seen that the training accuracy of the LSTM model with parameters adjusted by the PSO algorithm can be improved significantly.

Table 9 shows the improved percentage of the four evaluation indicators in the three data sets before and after the PSO optimization algorithm is added.

As can be seen from Table 9, each indictor of LSTM after parameter optimization by the PSO algorithm has been reduced to different degrees, and the prediction accuracy has been greatly improved. This is consistent with the results shown in the Tables 5–7, which further proves the effectiveness of the PSO optimization algorithm.

After the PSO optimization process, the optimal parameter combination of the LSTM model is as follows: the number of neurons in the hidden layer is 45, the number of LSTM network iterations is 135 and the rate of learning decline is 0.0065. From the analysis here and Section 4.3.1, it can be seen that the number of neurons in the hidden layer is just around the optimal effect value of the

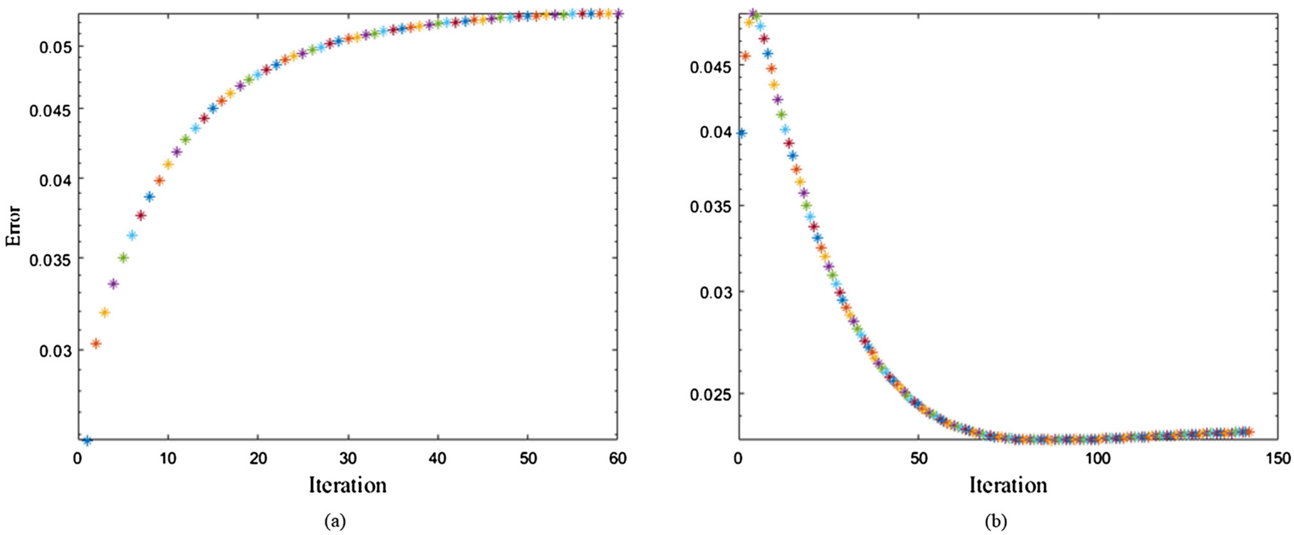


Fig. 13. (a) the training error of LSTM; (b) the training error of PSO-LSTM.

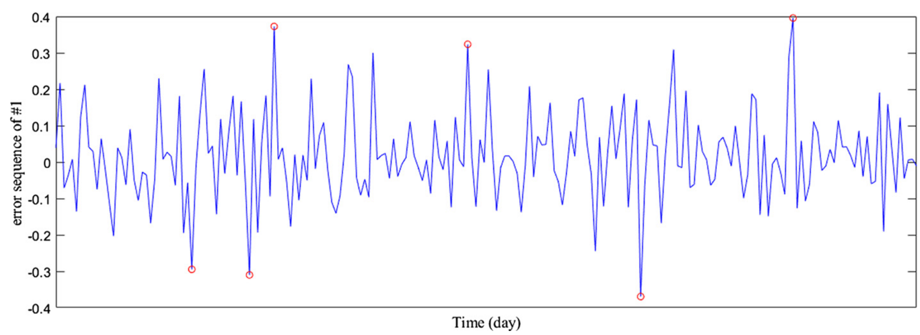


Fig. 14. Distribution of outliers in different error subsequences for #1. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

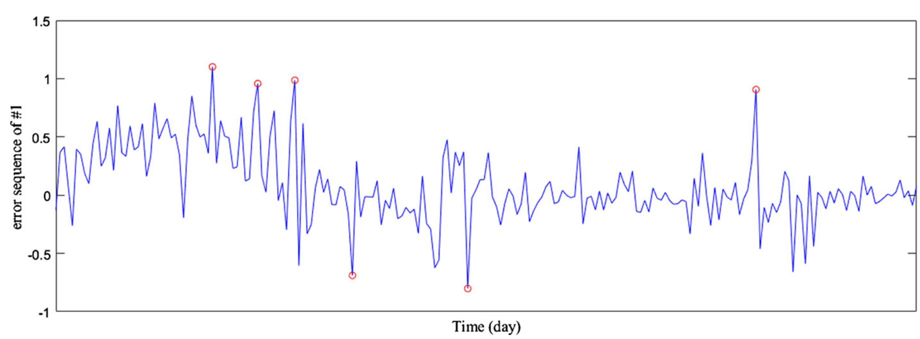


Fig. 15. Distribution of outliers in different error subsequences for #2.

dropout strategy, which further illustrates the superiority of model architecture.

4.4. Analysis of error correction by the ORELM

In this section, the error sequence obtained through the dpLSTM-PSO model is displayed. The Euclidean geometric distance is used to detect the outliers in the error sequence. Then, in the case of known outliers, the different impacts of various variants of the ELM model on the dpLSTM-PSO model will be compared and analyzed. And the effects of using the ORELM model for error modeling on the LSTM model alone and combined with the EWT algorithm will be analyzed.

Figs. 14–16 show the distribution of outliers in different error subsequences for three datasets. In Figs. 14–16, the point circled by a red circle is the detected outlier. It can be seen from Figs. 14–16 that there are a certain number of outliers in the error sequence obtained by the dpLSTM-PSO model. Tables 10–12 show the improved percentage of four evaluation indicators of relevant models. Figs. 17–19 show forecasting results of each sub-layer by the EWT-dpLSTM-PSO-ORELM model for three datasets.

From Tables 10–12, Figs. 8–10 and Figs. 17–19, it can be seen that:

(a) The error correction effect of the dpLSTM-PSO model directly using ORELM model is very poor, and even reduces the prediction accuracy of the original model. This may be due to the complexity and randomness of financial time series. In the case of complex original data, direct error correction will lead to the secondary accumulation of errors, thus reducing the prediction accuracy of the model.

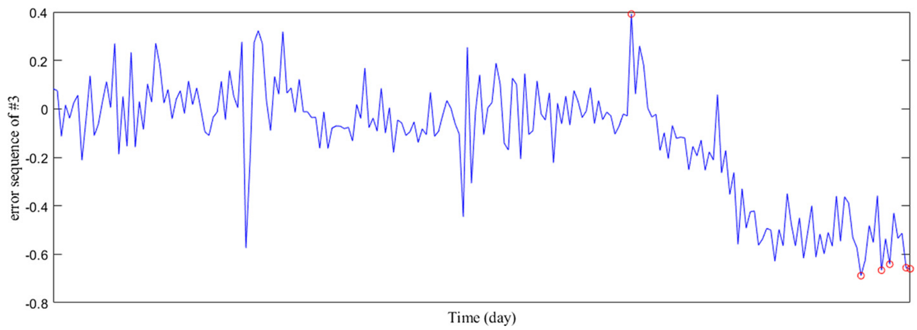


Fig. 16. Distribution of outliers in different error subsequences for #3.

Table 10
The improved percentage of four evaluation indictors of relevant models for #1.

Model	P_{MAPE} (%)	P_{MAE} (%)	P_{REMSE} (%)	P_{SDE} (%)
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO	−170.0302	−172.7511	−135.0922	−153.4933
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-ELM	64.6555	65.4235	74.6010	38.6801
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-RELM	16.7194	17.0136	13.4634	2.6540
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-WRELM	23.3231	23.5908	28.8061	12.8706
EWT-dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-ORELM	170.0302	172.7511	135.0922	153.4933

Table 11
The improved percentage of four evaluation indictors of relevant models for #2.

Model	P_{MAPE} (%)	P_{MAE} (%)	P_{REMSE} (%)	P_{SDE} (%)
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO	−226.3225	−221.5976	−182.1561	−200.5941
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-ELM	13.0715	12.8795	20.0264	16.2714
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-RELM	15.0311	14.9034	14.8880%	14.7563
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-WRELM	9.1918	9.1076	7.3781%	7.1146
EWT-dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-ORELM	87.7863	87.6725	88.4058	88.4058

Table 12
The improved percentage of four evaluation indictors of relevant models for #3.

Model	P_{MAPE} (%)	P_{MAE} (%)	P_{REMSE} (%)	P_{SDE} (%)
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO	−189.4980	−188.4868	−160.2597	−219.4539
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-ELM	1.6655	1.7104	5.3892	9.7222
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-RELM	13.1873	12.8848	21.5569	14.2094
dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-WRELM	8.0970	8.0281	18.9430	27.1854
EWT-dpLSTM-PSO-ORELM v.s. dpLSTM-PSO-ORELM	75.5647	75.4846	70.4591	73.0769

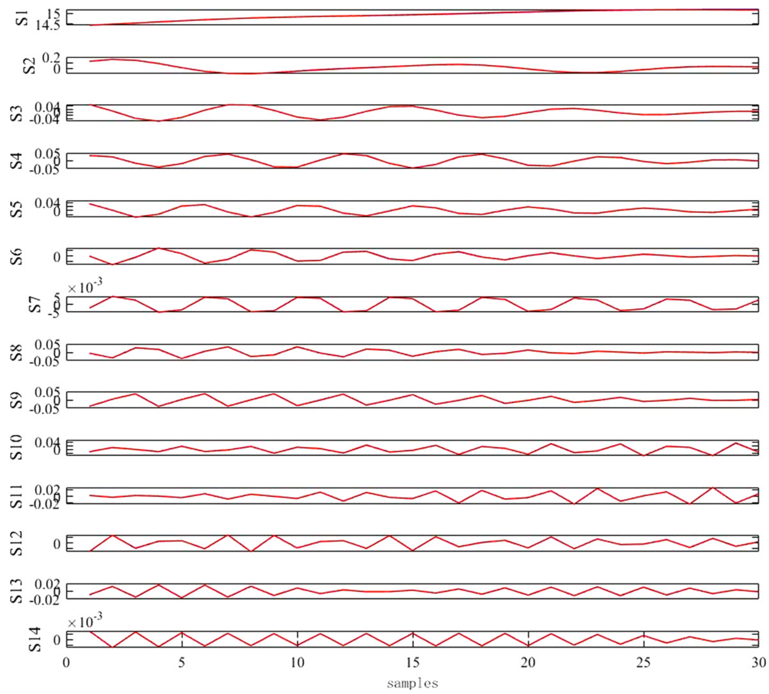


Fig. 17. Forecasting results of each sub-layer by EWT-dpLSTM-PSO-ORELM for #1.

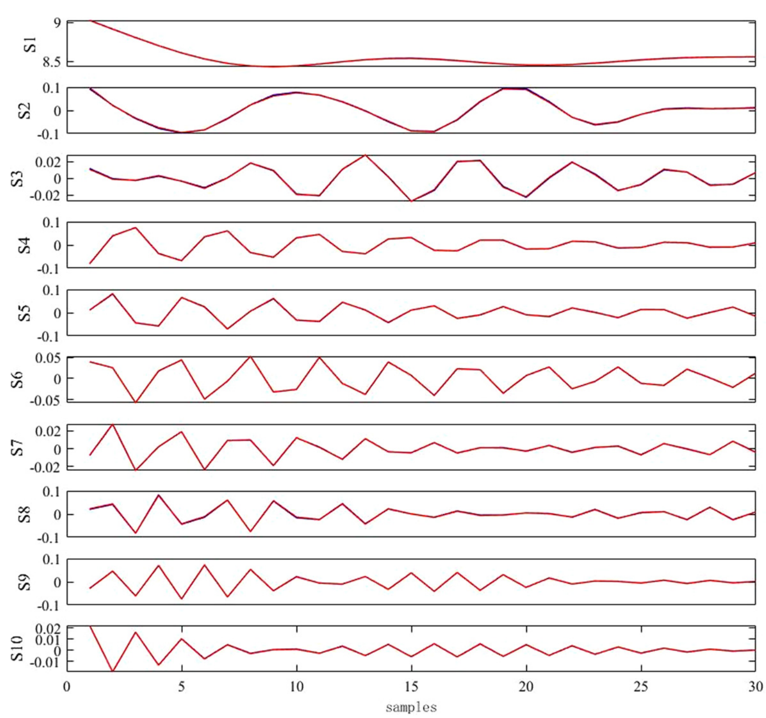


Fig. 18. Forecasting results of each sub-layer by EWT-dpLSTM-PSO-ORELM for #2.

(b) Among all the variants of ELM, the error correction by ORELM has the highest model accuracy. This illustrates the effectiveness of the ORELM model.

(c) The forecasting results and real values in each decomposition layer are almost identical in Figs. 12–14. It can be concluded that the combination of the EWT decomposition algorithm and error correction by ORELM can achieve excellent results. After error correction by ORELM, the prediction accuracy of main trend components is greatly improved, ensuring the effectiveness of the EWT decomposition algorithm. At the same time, the EWT decomposition algorithm solves the problem that error correction by the ORELM has a poor effect on random chaotic sequence. This point is also the subtlety of the model framework proposed in the study.

5. Conclusions

In the study, an improved hybrid forecasting framework is established to predict and analyze the daily stock closing price series. It based on EWT based decomposition, dropout strategy and PSO based LSTM deep network optimization and ORELM based error correction. The proposed hybrid model is compared with several

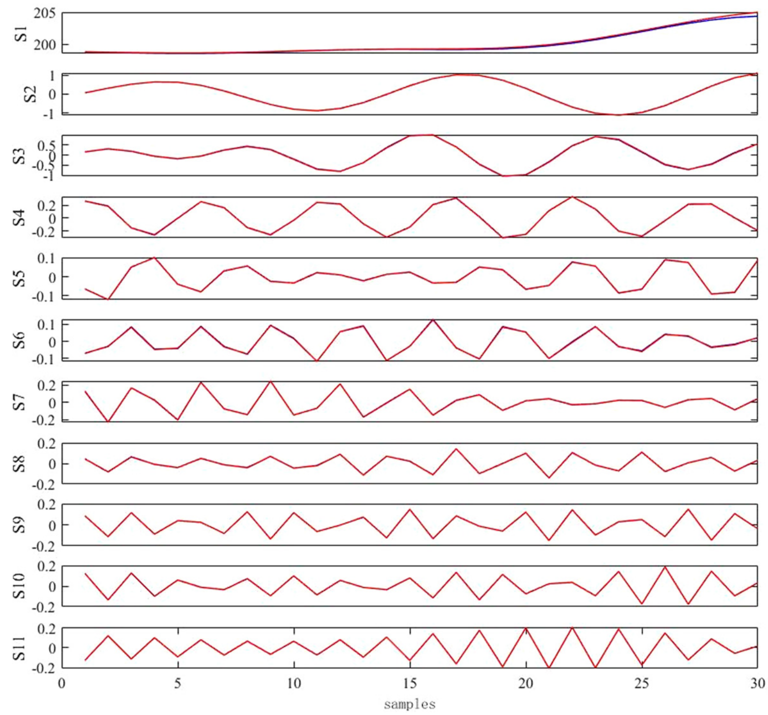


Fig. 19. Forecasting results of each sub-layer by EWT-dpLSTM-PSO-ORELM for #3.

benchmark models in the study to prove its effectiveness. In the study, the decomposition result of the EWT algorithm shows that the complex original stock closing price series is decomposed into several stable and regular sub-layers. In addition, ORELM based error correction and EWT decomposition method can be used together to learn from each other and greatly improve the prediction accuracy of stock closing price. The results of forecasting experiments show that: (a) the model integrated with the decomposition and error correction has higher prediction accuracy; (b) the dropout strategy and PSO algorithm can improve the forecasting accuracy of LSTM network; (c) the proposed hybrid framework has better prediction accuracy than other involved deep learning method or single models; (d) the proposed hybrid framework has achieved good experimental results by using stock closing price data from the United States and China over various time scales; and (e) the proposed hybrid forecasting framework can also be applied in other financial time series prediction. The forecasting results can be embedded in relevant stock market monitoring and financial data analysis.

Declaration of competing interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgments

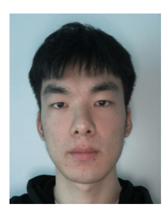
This study is fully supported by the National Natural Science Foundation of China (Grant No. 61873283), the Changsha Science & Technology Project (Grant No. KQ1707017), the Shenghua Yu-ying Talents Program of the Central South University and the innovation driven project of the Central South University (Project No. 2019CX005).

References

[1] N. Sim, H. Zhou, Oil prices, US stock return, and the dependence between their quantiles, J. Bank. Finance 55 (2015) 1–8.
[2] B. Blau, T.J. Brough, T. Griffith, Bank opacity and the efficiency of stock prices, J. Bank. Finance 76 (2017) 32–47.
[3] W. Yoon, K. Park, A study on the market instability index and risk warning levels in early warning system for economic crisis, Digit. Signal Process. 29 (2014) 35–44.
[4] N.D. Vanli, S. Tunc, M.A. Donmez, S.S. Kozat, Growth optimal investment in discrete-time markets with proportional transaction costs, Digit. Signal Process. 48 (2016) 226–238.
[5] F. Benedetto, G. Giunta, L. Mastroeni, A maximum entropy method to assess the predictability of financial and commodity prices, Digit. Signal Process. 46 (2015) 19–31.
[6] S. Basak, S. Kar, S. Saha, L. Khaidem, S.R. Dey, Predicting the direction of stock market prices using tree-based classifiers, N. Am. J. Econ. Finance 47 (2019) 552–567.
[7] E. Symitsi, L. Symeonidis, A. Kourtis, R. Markellos, Covariance forecasting in equity markets, J. Bank. Finance 96 (2018) 153–168.
[8] H. Nyberg, Predicting bear and bull stock markets with dynamic binary time series models, J. Bank. Finance 37 (9) (2013) 3351–3363.
[9] C.-H. Cheng, J.-H. Yang, Fuzzy time-series model based on rough set rule induction for forecasting stock price, Neurocomputing 302 (2018) 33–45.
[10] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, Y. Ma, PCANet: a simple deep learning baseline for image classification?, IEEE Trans. Image Process. 24 (12) (2015) 5017–5032.
[11] J. Xie, L. Xu, E. Chen, Image denoising and inpainting with deep neural networks, Adv. Neural Inf. Process. Syst. (2012) 341–349.
[12] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a Gaussian denoiser: residual learning of deep cnn for image denoising, IEEE Trans. Image Process. 26 (7) (2017) 3142–3155.
[13] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, B. Kingsbury, Deep neural networks for acoustic modeling in speech recognition, IEEE Signal Process. Mag. 29 (2012) 82–97.
[14] G.E. Dahl, D. Yu, L. Deng, A. Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, IEEE Trans. Audio Speech Lang. Process. 20 (1) (2012) 30–42.
[15] A. Graves, A.-r. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2013, pp. 6645–6649.
[16] F. Jia, Y. Lei, L. Guo, J. Lin, S. Xing, A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines, Neurocomputing 272 (2018) 619–628.

[17] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mech. Syst. Signal Process. 72 (2016) 303–315.
[18] M. He, D. He, Deep learning based approach for bearing fault diagnosis, IEEE Trans. Ind. Appl. 53 (3) (2017) 3057–3065.
[19] H.-z. Wang, G.-q. Li, G.-b. Wang, J.-c. Peng, H. Jiang, Y.-t. Liu, Deep learning based ensemble approach for probabilistic wind power forecasting, Appl. Energy 188 (2017) 56–70.
[20] A.S. Qureshi, A. Khan, A. Zameer, A. Usman, Wind power prediction using deep neural network based meta regression and transfer learning, Appl. Soft Comput. 58 (2017) 742–755.
[21] S. Selvin, R. Vinayakumar, E.A. Gopalakrishnan, V.K. Menon, K.P. Soman, Stock price prediction using LSTM, RNN and CNN-sliding window model, in: International Conference on Advances in Computing, IEEE, 2017, pp. 1643–1647.
[22] R. Singh, S. Srivastava, Stock prediction using deep learning, Multimed. Tools Appl. 76 (18) (2017) 18569–18584.
[23] A. Yoshihara, K. Fujikawa, K. Seki, K. Uehara, Predicting stock market trends by recurrent deep neural networks, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2014, pp. 759–769.
[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[25] X. Qiu, Y. Ren, P.N. Suganthan, G.A. Amaratunga, Empirical mode decomposition based ensemble deep learning for load demand time series forecasting, Appl. Soft Comput. 54 (2017) 246–255.
[26] H. Liu, X. Mi, Y. Li, Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM, Energy Convers. Manag. 159 (2018) 54–64.
[27] J. Hu, Y. Li, W. Xie, Hyperspectral image super-resolution by spectral difference learning and spatial error correction, IEEE Geosci. Remote Sens. Lett. 14 (10) (2017) 1825–1829.
[28] F.S. Duarte, R.A. Rios, E.R. Hruschka, R.F. de Mello, Decomposing time series into deterministic and stochastic influences: a survey, Digit. Signal Process. (2019) 102582.
[29] S. Lahmiri, Intraday stock price forecasting based on variational mode decomposition, J. Comput. Sci. 12 (2016) 23–27.
[30] C.-Y. Kuo, Does the vector error correction model perform better than others in forecasting stock price? An application of residual income valuation theory, Econ. Model. 52 (2016) 772–789.
[31] K. Gangopadhyay, A. Jangir, R. Sensarma, Forecasting the price of gold: an error correction approach, IIMB Manag. Rev. 28 (1) (2016) 6–12.
[32] R. Bondia, S. Ghosh, K. Kanjilal, International crude oil prices and the stock prices of clean energy and technology companies: evidence from non-linear cointegration tests with unknown structural breaks, Energy 101 (2016) 558–565.
[33] J. Gilles, Empirical wavelet transform, IEEE Trans. Signal Process. 61 (16) (2013) 3999–4010.
[34] S. Hochreiter, J.J.N.c. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[35] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, Neural Netw. 2 (2004) 985–990.
[36] W.-Y. Deng, Q.-H. Zheng, L. Chen, X.-B. Xu, Research on extreme learning of neural networks, Chinese J. Comput. 33 (2) (2010) 279–287.
[37] W. Zong, G.-B. Huang, Y. Chen, Weighted extreme learning machine for imbalance learning, Neurocomputing 101 (2013) 229–242.
[38] P. Horata, S. Chiewchanwattana, K.J.N. Sunat, Robust extreme learning machine, Neurocomputing 102 (2013) 31–44.
[39] Y. Sun, Y. Gao, An improved hybrid algorithm based on PSO and BP for stock price forecasting, Open Cybern. Syst. J. 9 (1) (2015) 2565–2568.
[40] S. Basak, S. Kar, S. Saha, L. Khaidem, S.R. Dey, Predicting the direction of stock market prices using tree-based classifiers, N. Am. J. Econ. Finance 47 (2019) 552–567.
[41] M. Roondiwala, H. Patel, S. Varma, Predicting stock prices using LSTM, Int. J. Sci. Res. 6 (4) (2017) 1754–1756.
[42] Y. Seo, S. Kim, O. Kisi, V.P. Singh, Daily water level forecasting using wavelet decomposition and artificial intelligence techniques, J. Hydrol. 520 (2015) 224–243.
[43] S. Ghimire, R.C. Deo, N. Raj, J. Mi, Wavelet-based 3-phase hybrid SVR model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction, Renew. Sustain. Energy Rev. 113 (2019) 109247.
[44] A. Meng, J. Ge, H. Yin, S. Chen, Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm, Energy Convers. Manag. 114 (2016) 75–88.
[45] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, Z. He, Wavelet transform based on inner product in fault diagnosis of rotating machinery: a review, Mech. Syst. Signal Process. 70 (2016) 1–35.

Hui Liu received the B.Sc. (Honors) and M.Sc. degrees from Central South University, China, in 2004 and 2008, respectively, the Ph.D. from Central South University, China, in 2011, the Dr.-Ing. degree from University of Rostock, Germany, in 2013 and the habil. degree from University of Rostock, Germany in 2015. Currently, he is a full professor of artificial intelligence & smart cities at Central South University, China. He is deputy dean of the faculty of traffic and transportation engineering at Central South University, China. He previously served as the BMBF junior group leader appointed by the Ministry of Education & Research of Germany at University of Rostock, Germany. His current research areas of interest include time series forecasting, big data analysis, smart devices and systems, etc.



Zhihao Long received the B.Sc. (Honors) degree from Central South University, China, in 2018. Currently, he is working towards the M.Sc. degree at Central South University, China. His current research focuses on artificial intelligence.