Soft Computing
https://doi.org/10.1007/s00500-021-06122-4

DATA ANALYTICS AND MACHINE LEARNING

Research on a hybrid prediction model for stock price based on long short-term memory and variational mode decomposition

Yang Yujun[1,3,4] · Yang Yimei[1,3] · Zhou Wang[2]

Accepted: 5 August 2021

Abstract
The stock market plays a vital role in the economic and social organization of many countries. Since stock price time series are highly noisy, nonparametric, volatility, complexity, nonlinearity, dynamics, and chaos, the stock market prediction is an important issue for investors and professional analysts. In the financial field, stock market prediction is not only an important task but also an important research topic. For different problems, researchers have proposed many prediction methods. Many papers provide strong evidence that stock prices can be predicted from past price data. In this paper, we propose a hybrid prediction model for stock price based on long short-term memory (LSTM) and variational mode decomposition (VMD). We use the variational mode decomposition method to decompose the complex time series of stock prices into several relatively flat, regular, and stable subsequences. Then, we use each subsequence to train the long- and short-term memory neural network and predict each subsequence. Finally, we merge the predicted values of several subsequences to form the predicted results of the stock price complex original time series. To verify fully the method, we selected four experimental data for testing. Compared with the prediction results of various prediction methods, the prediction accuracy of our proposed model is higher. Especially in the R^2 index, the experimental effect is very good. The proposed method achieves good results of more than 0.991 on each data set. Therefore, our proposed hybrid prediction model is accurate and effective in forecasting stock prices and has practical significance and reference value.

Keywords LSTM · VMD · Stock price · Price prediction · Time series

1 Introduction

As of September 2020, the statistics of China Securities Registration and Clearing Corporation show that there are more than 164.2 million securities investors in China. Forecasting stock prices is an important issue for investors and professional analysts. In the financial field, the forecasting stock prices is not only an important challenge but also an important research topic. To effectively reduce investment risk and get a steady return on their investment, many researchers have proposed lots of prediction methods (Kaya and Karsligil 2010; Yang et al. 2017a; Yujun et al. 2016a; Yuan et al. 2020; Shi et al. 2019; Zhang et al. 2012). Many papers provide strong evidence that it is feasible to predict the stock price according to the past price data (Zhang et al. 2020; Zhou et al. 2018; Idrees et al. 2019). With the application of artificial intelligence in various fields more and more widely, especially the deep neural network is widely used in the field of finance and insurance. (Hasan and Akyokuş 2017; Wang et al. 2018; Yıldırım et al. 2019; Khare et al. 2017; Zheng and J. Zhu 2017; Xi 2018; Yu et al. 2017; Sohangir and Wang 2018; Lee et al. 2019), many investors are interested in these applications. The financial data mainly include high-frequency data and low-frequency data (Guo et al. 2018). At

✉ Yang Yimei
yym1630@163.com

1 School of Computer Science and Engineering, Huaihua University, Huaihua 418008, China

2 School of Computer and Software Engineering, Xihua University, Chengdu 610039, China

3 Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities, Huaihua 418000, China

4 Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, Huaihua 418000, China

Published online: 06 September 2021

present, the research work mainly focuses on data analysis and technical method analysis (Minh Dang et al. 2018).

In terms of technical methods, many researchers usually use mathematical statistics or fitting methods to analyze financial historical data. They use these historical data to predict their trend, such as forecasting or fitting the short- and medium-term stock prices. Over the past decade, artificial intelligence technology has developed rapidly and has been widely used in many fields. More and more researchers use various artificial intelligence methods to predict the stock price. These methods include neural network method (Gao and Chai 2018), multi-core learning method (Day and Lee 2016; Hassan 2017), stepwise regression method (Jeon et al. Mar. 2018; Cheng and Shiu 2014), deep machine learning method (Chong and Han Oct. 2017; Liu and Wang 2019; Sismanoglu et al. 2019; Alhazbi et al. 2020), wavelet analysis method (Tang et al. 2009), principal component analysis method (Waqar et al. 2017), and other methods (Liu and B. Song 2018; Lim and Yeo 2020; Kwon et al. 2011; Yujun et al. 2016b). Many methods have been widely used in some specific aspects because they have achieved good experimental results. However, there are many problems worth studying such as data selection, model assembly, and parameter configuration in the process of training model. How to solve these problems has always been a very popular research topic. When analyzing the company's fundamentals (Weng et al. 2017), some researchers mainly use natural language processing and social network technology (Mankar et al. 2018) to analyze the financial statements and news of company to the future trend of the company's stock price.

The LSTM is short for long short-term memory. It is a neural network model of deep learning and has strong nonlinear expression ability. Because of its memory characteristics, it is one of the good methods to deal with time series. Therefore, this paper applies the LSTM algorithm to the basic prediction model of stock price time series. The LSTM model improved by the RNN is short for the recurrent neural network (Hochreiter and Schmidhuber 1997). Therefore, the output data of the LSTM depend not only on its own weight and input data, but also on the input data of one or more previous neurons. As we all know, stock price is a typical time series data. Over the years, lots of researchers have used LSTM method (Ojo et al. 2019; Lai et al. 2019; Liu et al. 2018; Wei 2019; Chen et al. 2015; Yang and Yang 2020; Bukhari et al. 2020; Ma et al. 2019; Hu et al. 2019) to analyze the historical data of stocks to predict the future price or trend of stock price. Some research work has studied the correlation between several stock price data (Xiao et al. 2013; Yang et al. 2017b; Yujun et al. 2017; Gers and Schmidhuber 2000). The results of these studies show that the LSTM method is very suitable for processing stock price data. The authors of reference (Wei et al. 2017) have proved the advantage of the LSTM by using the result of predicting coding unit splitting of the LSTM. Stock price trend prediction research is also a new research direction of time series, so the LSTM is the inevitable choice of stock price prediction.

The VMD is short for variational mode decomposition. It is a widely used signal decomposition method. Unlike empirical mode decomposition (EMD), the VMD uses variational problem to solve that problem (Dragomiretskiy and Zosso 2014). The VMD is a decomposition algorithm with a solid mathematical foundation. It can establish constrained variational model expressions by adaptively decomposing the signal into several eigenmode components in a non-recursive manner. The VMD decomposition purpose is to find a specified number of intrinsic modal components. To solve this variational problem, researchers generally use the alternating direction multiplier method to solve the problem of the mode and corresponding center frequency. The VMD algorithm is usually suitable for processing non-stationary nonlinear signals.

For more than 20 years, many researchers pay more and more attention to the application of the VMD (Upadhyay and Pachori 2017; Wei and Wang 2019; Sun et al. 2019; Gendeel et al. 2018; Sun and Zhao 2020; Wang et al. 2019; Zhang et al. 2019; Xu et al. 2020; Zhou et al. 2020; Han et al. 2019). To suppress the white noise, as well as non-stationary acoustic noises, Upadhyay and Pachori (2017) proposed a novel and effective speech enhancement method that employs the combination of variational mode decomposition method and empirical mode decomposition method. In this method, first, the EMD is used to decompose the noisy speech signal into intrinsic mode functions (IMFs). Further, the VMD is applied to the summation of selected IMFs. This method is the selection of IMFs based on Hurst exponent and further applying steps of the VMD method and suitable to reduce the low-frequency noise, as well as high-frequency noise. To evaluate effectively the performance degradation degree of spindle, one of the core components of machine tool, Wei and Wang (2019) proposed a new degradation evaluation method based on variational mode decomposition and random forests (RFs). This method firstly uses the VMD technology to process the current signal to obtain several modal components. Then, this method calculates each mode component as eigenvalues and uses the RFs algorithm to classify the eigenvalues. The experimental results of this method show that VMD can decompose the signal better and avoid the phenomenon of the modal mixture. To predict effectively short-term wind power, Sun et al. (Sun et al. 2019) proposed a hybrid model that consists of the variational mode decomposition, the K-means clustering algorithm, and long short-term memory network. This method uses the VMD to decompose the raw wind power series into a certain number of sub-layers with different frequencies and uses K-means to execute for splitting the data into an ensemble of components with a similar fluctuant level of each sub-layer. Then, this method adopts the LSTM as the principal forecasting engine for capturing the unsteady characteristics of each component and generates the forecasting results by aggregating the predicted components. To obtain accurate wind speed forecasts, Gendeel et al. (2018) proposed an artificial neural networks (NNs) model with the variational mode decomposition for a short-term wind speed forecasting. This method uses the VMD to decompose the historical wind speed into different intrinsic mode functions (IMFs) and adopts the back-propagation NN with Levenberg–Marquardt to build sub-models according to the different characteristic of each IMF that was superposed to obtain wind speed-forecasting models. In order to detect and classify the power quality (PQ) disturbances for distribution networks with distributed generation, Xu et al. (2020) proposed a novel detection and classification method based on the variational mode decomposition and detrended fluctuation analysis (DFA). This method firstly builds a distribution system with photovoltaic and wind power generation as a test platform. It uses the VMD to decompose the data into nine types of power quality disturbances in the distribution network and filters the noise. Meanwhile, the mode functions containing characteristic information are extracted as input signals of detrended fluctuation analysis (DFA). Power quality disturbances are classified from the view of their distributed energy operational status, and three types of windows are set up to deal with different frequency disturbances. To efficiently manage the cloud resources, improve the quality of service and avoid the violations of Service-Level Agreement (SLA) agreements, Zhou et al. (2020) proposed an Ensemble Forecasting Approach (EFA) for highly dynamic cloud workload by applying variational mode decomposition (VMD) and R-Transformer to decrease the non-stationarity and high randomness of highly dynamic cloud workload sequences. This method uses the VMD that decomposes the workload into multiple Intrinsic Mode Functions (IMFs) which are then imported into our ensemble forecasting module based on R-Transformer and autoregressive model. To improve the accuracy of the multi-step wind power forecast, Han et al. (2019) proposed a variational mode decomposition—long short-term memory forecast method. This method adopts the variational mode decomposition method to decompose the wind power data into three constituent modes, named as the long-term component, the fluctuation component, and the random component. Then, it utilizes long short-term memory networks to learn deeply the characteristics of the three constituent modes.

These VMD-based methods improve the performance by using the strategy of decomposing time series. It is a difficult problem for researchers to decompose the stock price nonlinear complex data effectively. The time series of stock price has the characteristics of uncertainty and nonlinearity. The effect of using one method to predict stock price is generally not good. These methods mentioned above have greatly improved the prediction effect. Therefore, from the results of these literature works, we can speculate that the single specific method can get worse prediction results than the hybrid method. In this paper, combining the features of the LSTM and the VMD method that can divide and conquer complex problems, we propose a hybrid LSTM-VMD method to predict the price of several stock.

Other sections of this paper are designed as follows. The concepts and terms of the VMD and LSTM are introduced in Sect. 2. Section 3 mainly presents the structure and flowchart of the proposed hybrid LSTM-VMD predict method. We describe in detail the process of data collection, data preprocessing and model building in Sect. 4. In Sect. 5, we describe in detail the experimental results of several methods used in this paper and introduce the simulation experiments and experimental results analysis of the proposed hybrid prediction method. In Sect. 6, we introduce the conclusion and prospect of our work.

2 Related works

Over the years, there have been many studies on future stock price index predictions in the financial field. There are mainly the following three main research directions: (1) Forecasting methods based on time series analysis; (2) Forecasting methods based on machine learning; (3) Forecasting methods based on hybrid methods. After discussing related concepts and terms, we introduced LSTM and VMD related to this research.

2.1 Predicting stock price

The time series of stock price has the characteristics of large fluctuation and frequent fluctuation. The main reason is that many factors affect the volatility of stock prices. These factors include industry news, enterprise project contract, national policies, investor sentiment, interest rates, important meetings, economic policy, exchange rates, monetary policies, inflation, natural disaster, enterprise reorganization, temporary important events, and human intervention. The above factors will affect the stock price temporarily to varying degrees. In fact, these factors generally only change the stock price in the short term, not in the long term. To predict the stock price superficially, it is necessary to establish a relationship model between these factors and the stock price. People have always believed that stock price forecasts depend on historical prices, and regarded the forecasting problem as a time series fitting problem.

We believe that the effect of many studies is not satisfactory using a single analytical method to predict stock price based on historical data. If we consider multiple factors or use multiple techniques to build a mixed model, we may get better prediction results.

2.2 Long short-term memory

The long short-term memory (LSTM) neural network is a time recurrent neural network (RNN), suitable for processing and predicting events with relatively long intervals and delays in time series. Hochreiter and Schmidhuber (Wei 2019) proposed the LSTM in 1997. Alex Graves improved and promoted it recently. The LSTM has been widely used in many problems because that has achieved considerable success. The LSTM is designed deliberately to avoid long-term dependency problems. Remember long-term information is the default behavior of the LSTM in practice. The LSTM has many applications in the field of science and technology, such as learning translation language, controlling robots, image analysis, document summarization, speech recognition, image recognition, handwriting recognition, controlling chat robots, predicting diseases, click-through rates, and stocks, synthesizing music, and so on. We first introduce RNN and then LSTM.

Many of the problems we study have time attributes, such as natural language processing, speech recognition, time series data, machine translation, and so on. Generally, word processing needs to consider the relevance between text contexts. At the same time, it is necessary to consider the relationship between them, as well as the weather conditions for several consecutive days when predicting the weather. In response to these problems, the prediction effect of general neural network is poor, even very bad. The main goal of designing RNN is to solve the above kind of problem.

The RNN has a repeating neural network module in the form of a chain. In fact, the repetitive structure of the simplest RNN model only has the simple structure of a tanh function. Figure 1 shows the simplest structure of RNN.

The RNN is very suitable for dealing with nonlinear problems with time property. The data flow of general RNN generally only feedforward data, while bidirectional RNN allows forward and backward feedback data. The advantage of design feedback is that the algorithm can easily modify the weight and residual value of the previous layer. The above design is suitable for analyzing and predicting time series. It can predict the future value according

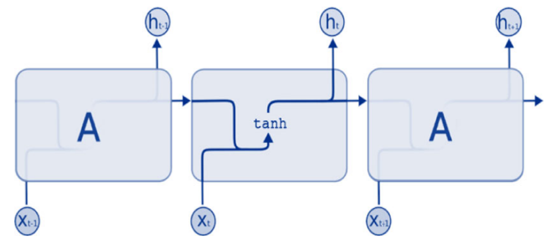

Fig. 1 The simple structure of the RNN

to the rules extracted from the historical data. Figure 1 shows the simplest RNN. The simplest structure of the RNN in Fig. 1 has three repeating units, each of which has only one *tanh* layer, two data input ports, and two data output ports. Figure 2 shows the basic structure of the RNN. The fold architecture of RNN is shown in the leftmost subgraph in Fig. 2. The right three subgraphs of Fig. 2 are the recursive loop expansion of the basic structure of RNN. In Fig. 2, y_t and x_t represent output and input data of the RNN module h at time t, respectively. Different from the traditional neural network, the RNN designs parameters W_{hx}, W_{hh}, W_{yh} to be shared in each network layer. On the surface, the parameters W_{hx}, W_{hh}, W_{yh} of each layer of the RNN are the same. It gives people the feeling that each layer of RNN has the same structure and is realizing the same function. In fact, the output data y_t and input data x_t of each layer of RNN are different. This greatly reduces the number of RNN parameters to learn. Through repeated expansion, RNN will be transformed into a multilayer neural network.

Each layer of RNN has output port and input port, but it does not require output data and input data. For example, in language translation, we only need to know the last output after the input of the last language symbol, but not the output after the input data of each language symbol. The core unit of RNN is self-expanding and has multiple hidden layers. It can capture short-term and long-term information in data.

In general, RNN is prone to gradient disappearance in the process of training. In training, once the gradient

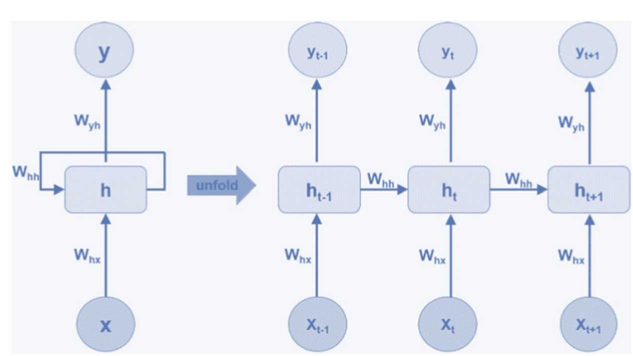

Fig. 2 The basic structure of the RNN

disappears, the RNN will become an infinite loop and the training of RNN does not converge. Therefore, the simple structure of the RNN is prone to gradient disappearance problems and is not suitable for predicting long-term events of time series.

The purpose of the LSTM is to solve gradient vanishing problem during network training. The LSTM network model adds the input gate, output gate, and forget gate based on the simplest RNN. In Fig. 3, three gate modules in accordance with σ can effectively avoid gradient elimination. Formula 1–formula 6 associated with the LSTM neural network are shown as follows. These six equations are for the output gate, memory cell, inverse of the memory cell, input gate, forget gate, and output data, respectively.

Here, x_t is input vector; i_t,o_t, and f_t represent the output results of input gate, output gate, and forget gate, respectively; c_t represents the activation status of each cell; h_t represents the output results of memory unit; *tanh* and σ represent the activation functions; b_i, b_c, b_f, and b_o represent the corresponding bias vectors; e_i, e_c, e_f and e_o represent the vectors of all one corresponding tob_i, b_c, b_f, and b_o, respectively; w_{xi}, w_{hc}, w_{hf}, w_{ho}, w_{hi}, w_{xc}, w_{xf}, and w_{xo} represent the corresponding weight vectors.

Therefore, the LSTM is prone to solve long-term dependence problems. The purpose of memory neuron design is to store some important long-term information of LSTM neural network for state information. In general, each gate module use an activation function to perform nonlinear conversion on the information passing through the gate module. The main function of forgetting gate f_t is to clear part of the state information of neurons.

$$\begin{aligned} f_t &= \sigma\left(W_f \cdot (h_{t-1} \parallel x_t) + b_f\right) \\ &= \sigma\left(\left[x_t, h_{t-1}, b_f\right]' \cdot \left[W_{xf}, W_{hf}, e_f\right]\right) \end{aligned} \tag{1}$$

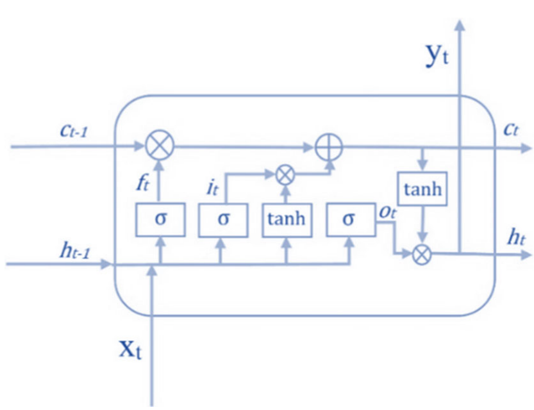

Fig. 3 The basic structure of the LSTM

$$\begin{aligned} i_t &= \sigma(W_i \cdot (h_{t-1} \parallel x_t) + b_i) \\ &= \sigma\left(\left[x_t, h_{t-1}, b_i\right]' \cdot \left[W_{xi}, W_{hi}, e_i\right]\right) \end{aligned} \tag{2}$$

$$\tilde{C}_t = \tanh(W_C \cdot (h_{t-1} \parallel x_t) + b_C) \tag{3}$$

$$\begin{aligned} C_t &= f_t \times C_{t-1} + i_t \times \tilde{C}_t \\ &= f_t \cdot C_{t-1} + i_t \cdot \tanh\left(\left[x_t, h_{t-1}, b_c\right]' \cdot \left[W_{xc}, W_{hc}, e_c\right]\right) \end{aligned} \tag{4}$$

$$\begin{aligned} o_t &= \sigma(W_o \cdot (h_{t-1} \parallel x_t) + b_o) \\ &= \sigma\left(\left[x_t, h_{t-1}, b_o\right]' \cdot \left[W_{xo}, W_{ho}, e_o\right]\right) \end{aligned} \tag{5}$$

$$h_t = y_t = o_t \ \times \tanh(C_t) \tag{6}$$

2.3 Variational mode decomposition

As a new signal decomposition method, variational mode decomposition (VMD) has been widely used in recent years. Unlike the recursive solution of empirical mode decomposition (EMD), VMD transforms the solution problem into a variational problem. The purpose of the VMD method is to find a specified number of intrinsic modal components. To solve this variational problem, researchers generally use the alternating direction multiplier method to solve the mode and the corresponding center frequency (Dragomiretskiy and Zosso 2014). The specific algorithm flow of the VMD method is as follows (Zhao et al. 2019).

(1) The constructive process.

To calculate the bandwidth of each model component, the analytic signal of each module component is obtained by Hilbert transform, and then the unilateral frequency spectrum is obtained by the formula (7).

$$u_k(t) \ \times \left[\sigma(t) + \frac{j}{\pi t}\right] \tag{7}$$

where intrinsic mode function (IMF) is defined as an amplitude modulated frequency modulated signal. $u_k(t)$ is the Dirac distribution defined as the formula (8).

$$u_k(t) \ = \sigma(t) \times A_k(t)\cos(\emptyset_k(t)) \tag{8}$$

By calculating the L2-norm of the derivative of the analytical signal and the bandwidth of each mode, the constrained variational problem is constructed as the formula (9).

$$\min_{u_k, w_k} \sum_k \left\| \partial_t \left[u_k(t) \times \left(\sigma(t) + \frac{j}{\pi t} \right) \right] \times \mathrm{e}^{-j w_k t} \right\|^2 \tag{9}$$

where $\sum_k u_k = s$ and s is an input time series data; u_k and w_k represent different modal components and corresponding frequency centers, respectively.

(2) The solution process.

To solve this variational problem, the constrained variational problems of formula (9) are transformed into unconstrained variational problems by using the Lagrange multiplier method.

$$r(t) = \sum_k u_k(t) \tag{10}$$

$$\mathrm{lag}(t) = \left\| \partial_t \left[u_k(t) \times \left(\sigma(t) + \frac{j}{\pi t} \right) \right] \times \mathrm{e}^{-jw_k t} \right\|^2 \tag{11}$$

$$L(\{u_k\}, \{w_k\}, \lambda(t)) = \eta \sum_k lag(t) + \|s(t) - r(t)\|^2 + \langle \lambda(t), s(t) - r(t) \rangle \tag{12}$$

where λ(t) represents Lagrangian multipliers and η is a quadratic multiplication factor.

Researchers generally use the alternating direction multiplier method to solve the above variational problems. The alternating direction multiplier method finds the saddle point of the Lagrangian expression by alternately updating $\{u_k\}, \{w_k\}$, and $\lambda(t)$. Among them, $\{u_k\}$ can be updated using the formula (13).

$$\{u_k\} = \underset{u_k \in X}{\operatorname{argmin}} \left\{ \eta \sum_k \mathrm{lag}(t) + \left\| s(t) - r(t) + \frac{\lambda(t)}{2} \right\|^2 \right\} \tag{13}$$

When the number of iterations is n + 1, the formula (13) becomes the following formula (14).

$$u_k^{n+1} = \underset{u_k \in X}{\operatorname{argmin}} \left\{ \eta \left\| \partial_t \left[u_k(t) \times \left(\sigma(t) + \frac{j}{\pi t} \right) \right] \times \mathrm{e}^{-jw_k t} \right\|^2 + \left\| f(t) - \sum_i u_i(t) + \frac{\lambda(t)}{2} \right\|^2 \right\} \tag{14}$$

where the w_k and the $\sum_i u_i(t)$ are equivalent to w_k^{n+1} and $\sum_i u_i(t)^{n+1}$, respectively. $n + 1$ is the number of iterations.

Using Parseval or Plancherel–Fourier isometric transformation, formula (14) is transformed into frequency domain form and solved in the frequency domain.

$$\hat{u}_k^{n+1} = \underset{u_k \in X}{\operatorname{argmin}} \left\{ \eta \| jw[\hat{u}_k(w + w_k)(1 + \operatorname{sign}(w + w_k))] \|^2 + \left\| \hat{f}(w) - \sum_i \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2} \right\|^2 \right\} \tag{15}$$

where *sign* is sign function. $\widehat{u}, \widehat{f}, and\ \widehat{\lambda}$ are represented as the frequency form of the corresponding signal $u, w, and \lambda$, respectively.

When $w - w_k$ is used to update w, the formula (16) is obtained from formula (15).

$$\hat{u}_k^{n+1} = \underset{u_k \in X}{\operatorname{argmin}} \left\{ \eta \| j(w - w_k)[\hat{u}_k(w)(1 + \operatorname{sign}(w))] \|^2 + \left\| \hat{f}(w) - \sum_i \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2} \right\|^2 \right\} \tag{16}$$

The above problems can be converted into a nonnegative frequency interval integral form as the following formula (17).

$$\widehat{u}_k^{n+1} = \underset{u_k \in X}{\operatorname{argmin}} \left\{ \int_0^\infty 4\eta (w - w_k)^2 |\widehat{u}_k|^2 + 2 \left| \widehat{f}(w) - \sum\nolimits_k \widehat{u}_k(w) + \frac{\widehat{\lambda}(w)}{2} \right|^2 \mathrm{d}w \right\} \tag{17}$$

Finally, the solution of the quadratic optimization problem can be obtained as the following formula (18).

$$\widehat{u}_k^{n+1} = \frac{\frac{\widehat{\lambda(w)}}{2} + \widehat{f}(w) - \sum_{i \neq k} \widehat{u}_i(w)}{1 + 2\eta (w - w_k)^2} \tag{18}$$

where $\widehat{u}_k^{n+1}$ is regarded as the Wiener filtering which was used to minimize the mean square value of estimation error. Similarly, w_k^{n+1} and $\widehat{\lambda}(w)^{n+1}$ are updated by the following formula (19) and formula (20), respectively.

$$w_k^{n+1} = \int_0^\infty w |\hat{u}_k|^2 \mathrm{d}w \Big/ \int_0^\infty |\hat{u}_k|^2 \mathrm{d}w \tag{19}$$

$$\hat{\lambda}(w)^{n+1} = \hat{\lambda}(w)^n + \theta \left[\hat{f}(w) - \sum_k \hat{u}_k^{n+1}(w) \right] \tag{20}$$

where the w_k^{n+1} is the central frequencies of the corresponding modes and the power spectrum center of the k-th modal component at the $n + 1$ iteration, the θ is the update coefficient of the $\widehat{\lambda}(w)^{n+1}$.

(3) VMD algorithm flow.

Step 1 Initializing variables and parameters: $\widehat{u}_k^1$, $\widehat{w}_k^1$, $\widehat{\lambda}(w)^1$, $n = 0$.

Step 2 Increase the number of iterations by one: $n = n + 1$.

Step 3 Update $\{u_k\}$ and $\{w_k\}$ according to formulas (13–19).

Step 4 Increase the number of modal by one: $k = k + 1$, carry out the steps (1–2), repeat step (3) until k = K ends.

Step 5 Update $\left\{ \widehat{\lambda}(w) \right\}$ according to formulas (20).

Step 6 Repeat steps (2–5) until.

$\sum_k \|u_k^{n+1} - u_k^n\|^2 / \|u_k^n\|^2 < \varepsilon$, where ε is an accuracy error.

(4) VMD decomposition example.

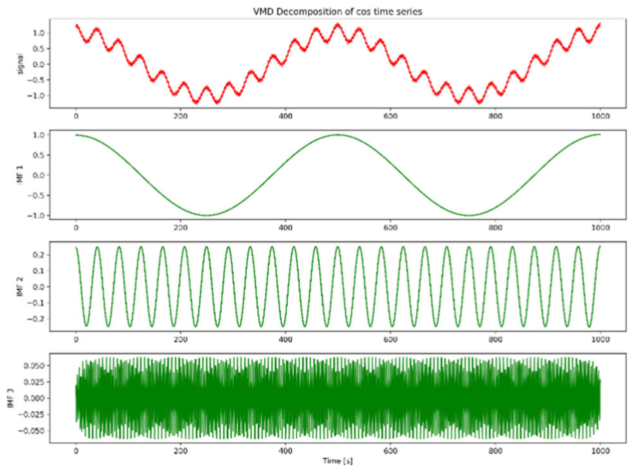

Fig. 4 The VMD decomposition process chart of cos sequences

To understand the VMD decomposition process more clearly, here is an example to illustrate. Figure 4 shows the VMD decomposition process of the time series represented by formula (21) when the VMD parameter K = 3.

$$x(t) = \cos(4\pi t) + \cos(48\pi t)/4 + \cos(576\pi t)/16 \quad (21)$$

where $t = f, 2f, \ldots, 1000f, f = 0.001$.

3 Methodology

Before introducing the structure and process of the LSTM-VMD hybrid stock price forecasting method based on long-term memory neural network method and integrated empirical mode decomposition method, this chapter first introduces the principle of the LSTM-VMD which is a hybrid stock price forecasting method based on the LSTM and the VMD. In this paper, the proposed LSTM-VMD hybrid forecasting method for stock price firstly decomposes the complex time series of stock price into several relatively simple subsequences by using the variational mode decomposition method. Then, the LSTM method is used to predict the value of each subsequence. Finally, the LSTM-VMD method gets the prediction results of the stock price complex original time series by fusing the LSTM prediction results of multiple subsequences.

The overall structure of our proposed hybrid LSTM-VMD stock price predict model is shown in Fig. 5. The three main processes of the proposed hybrid predict model for stock price include VMD decomposition for stock price original time series, LSTM prediction for each IMF of original time series, and fusion for all LSTM prediction values of all IMF. The proposed hybrid LSTM-VMD prediction model includes three stages and eight process steps. The flowchart of the proposed prediction method is shown in Fig. 6. The three stages of the proposed hybrid

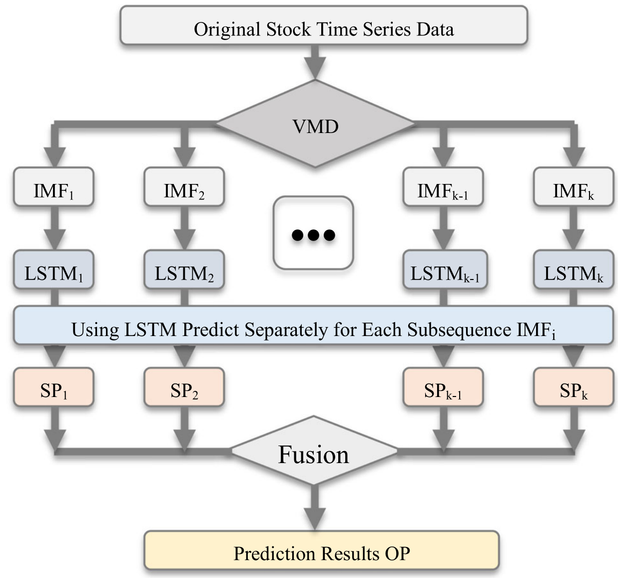

Fig. 5 The overall structure of the proposed hybrid LSTM-VMD predict model

LSTM-VMD prediction model include data input, model prediction, and model evaluation. The data input stage and the model evaluation stage each include four steps, and the model prediction stage includes three steps. The following is a brief introduction to the structure of our proposed hybrid LSTM-VMD predict model for a stock price.

(1) Generate simulation data of a complex cos function and collect stock price time series sequence data in the real financial market, preprocess and standardize the original stock index time series data, make the data format of stock index time series meet the format requirements of VMD method decomposition, and form the input data X of hybrid LSTM-VMD stock price predict model.

(2) Decompose input data X into K subsequences using the VMD method. The K is determined by the complexity of time series data. If the time series is more complex, the value of K will be larger. Otherwise, the value of K will be smaller. For example, when the cos function of the above example is decomposed, the value of K is 3. In general, the value range of K is 3 to 9. Therefore, K subsequences of VMD decomposition are K IMF subsequence which is, respectively, expressed as IMF_1, IMF_2, IMF_3, …, IMF_K.

(3) Build and train one LSTM neural network for each subsequence, and the processing of each subsequence is independent and does not affect each other, so K subsequences need to build K independent LSTM neural network. After the K LSTM neural network which is marked as LSTMk (k = 1, 2, 3, …, K − 1, K) is built, predict of stock index time series of K subsequences with K corresponding LSTM neural network, K prediction results SPk

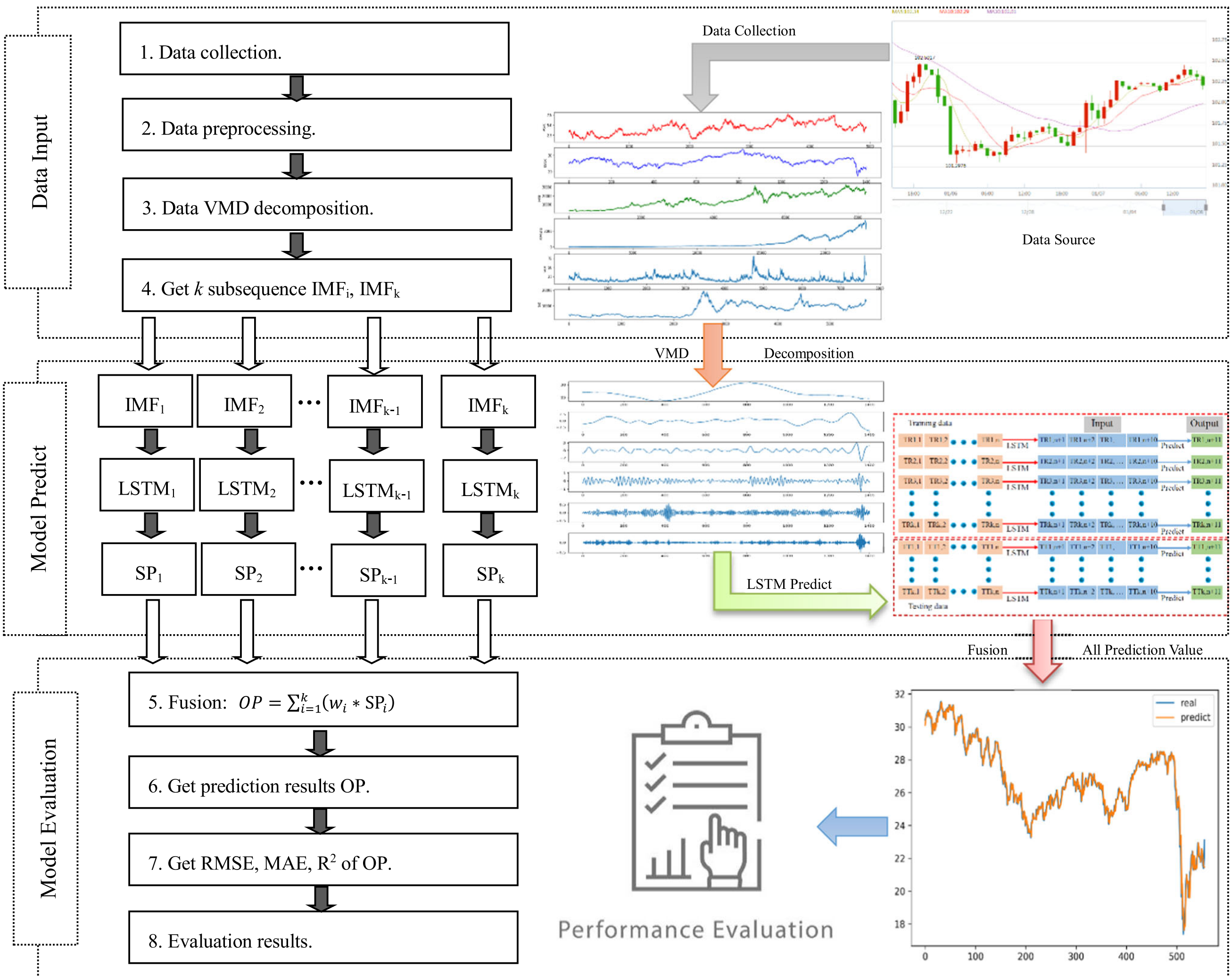

Fig. 6 The flowchart of the proposed prediction model

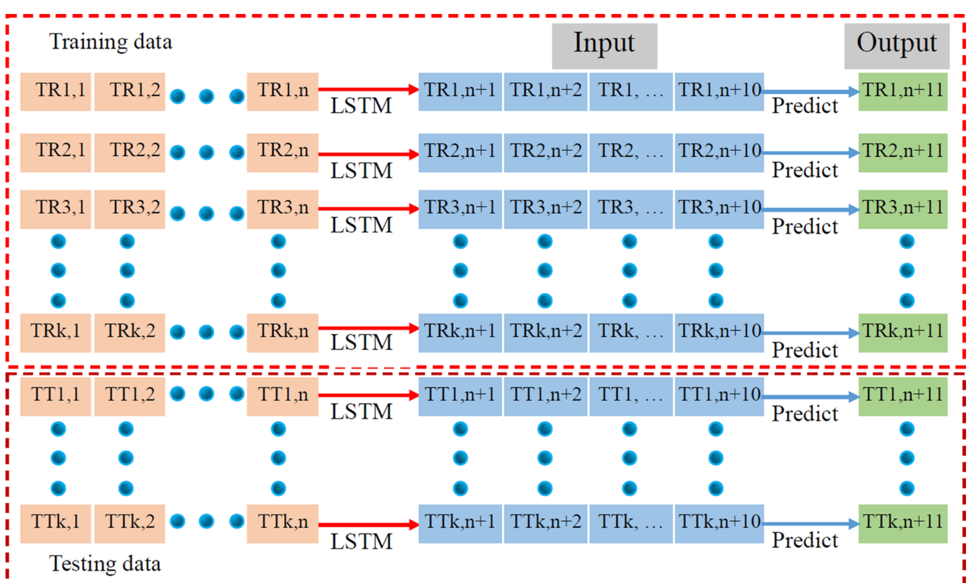

Fig. 7 The prediction process of the LSTM

(k = 1, 2, 3, …, K − 1, K) are obtained accordingly. The specific prediction process of the LSTM is shown in Fig. 7.

(4) Fuse the K LSTM prediction results of K subsequences. There are many fusion methods to fuse the prediction results of several subsequences to get the final prediction results of the stock index original time series sequence. The weighted addition is selected as the fusion method, which is to accumulate the weighted prediction results of all subsequence to form the final prediction results of the original stock index time series data.

(5) Finally, compare the prediction results with the actual stock index price. We evaluate the prediction performance of the hybrid LSTM-VMD stock price predict model by calculating RMSE, MAE and R^2 of the prediction results.

The flowchart of our proposed hybrid LSTM-VMD stock price predict model is shown in Fig. 6. The data flowchart and processing of the proposed model are divided into three stages: data input, model prediction, and model evaluation. The data input stage includes data collection, data preprocessing, data VMD decomposition, and generator K subsequence. In the prediction stage of the model, K subsequence data decomposed by VMD are, respectively, predicted by the LSTM model, and the prediction value of each subsequence data is obtained. The evaluation stage of the model includes the following steps. Firstly, the prediction value of K subsequence data is fused with weighted addition. In this paper, the fusion method of the weighted addition is used, that is, sum the weighted prediction value of K subsequence data and then get the prediction value of the stock index original time series data. The next operation is to calculate the RMSE, MAE, and R^2 values of the prediction values in combination with the original sequence data. Finally, we use these values of RMSE, MAE, and R^2 to evaluate the proposed hybrid prediction model.

4 Experiments data

In this section, we briefly introduce the experimental data used in our research work. In order to better test the prediction effect of our proposed model, we used two kinds of experimental data in the study. The first experimental data are the data of artificial simulation experiments which are the experimental data automatically generated by a computer algorithm. The purpose of using artificial simulation data is to verify the validity and correctness of the proposed model. Since the simulation data in many literature works are used to test the validity and correctness of one model, we also use the simulation data to test the validity and correctness of our proposed model in this study. The second experimental data are the real stock price data in the financial field. Only the real data in the real domain can be used to effectively test the proposed model so that the data have practical value and experimental purpose. It is only valuable to apply a model to a practical field.

4.1 Artificial simulation experimental data

In the experiment, we use the artificial simulation experimental data to verify the effectiveness of our proposed model. If the length of the artificial simulation data is long enough, better experimental results will be obtained. Therefore, the data length generated by a computer program is 10000 in the experiment. The artificial simulation experimental data are generated based on a complex cos combination function. A specific computer program automatically generates the artificial simulation experimental data according to Formula (21).

4.2 Real social experimental data

We collect stock index data from Yahoo Finance that is the real stock market in order to study the effect of the prediction model. To obtain more objective experimental results, we chose four stock indices from different countries. They are the most important stock indices of the financial time series in the world.

The four stock indices in this experiment are SP500, SZ, HSI, DAX, VIX, and ASX. In this paper, the SP500 is used as an abbreviation of the Standard & Poor’s 500th is that American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. SZ is the abbreviation of the Shenzhen Stock Exchange Component Index which is a component stock index compiled by the Shenzhen Stock Exchange. It takes stocks of 40 listed companies with market representativeness as the calculation object from all listed stocks and calculates the weighted stock price index with circulating shares as the weight, comprehensively reflecting the stock price trend of A and B shares listed on Shenzhen Stock Exchange Potential. The HSI the abbreviation of the HangSeng Index that is a free float-adjusted market capitalization-weighted stock market index in Hong Kong of Asia. The DAX is used as an abbreviation of the Deutscher Aktien index which is a blue-chip stock market index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange in Germany of Europe. The VIX is used as an abbreviation of the CBOE Volatility Index which is a weighted measure of the implied volatility for eight OEX put and call options. The VIX is a Chicago Options Delayed Price that represents the implied volatility for this hypothetical at-the-money OEX option in Chicago. The ASX is used as an abbreviation of the Australian Securities Exchange Ltd that is an

Australian public company that operates Australia's primary securities exchange in Australia.

The raw data used in the experiment can be obtained free of charge from Yahoo Finance (http://finance.yahoo.com). The data set of each index includes six daily attributes: trading volume, lowest price, highest price, closing price, adjusted closing price, and opening price from its opening date to May 18, 2020. In the experiment, we choose the adjusted closing price of each index as the experimental data. There are many literature works that choose the closing price as the experimental data. But we think the closing price itself may not fully reflect the value of the stock, such as cash dividends, stock dividends, stock splits. Comprehensive consideration, the adjusted closing price can better reflect the actual price of the stock. The following is a brief explanation of the impact of the three distributions on the stock price.

When distributions are made, the adjusted closing price calculations are quite simple. For cash dividends, the value of the dividend is deducted from the last closing sale price of the stock. For example, we assume that the closing price for one share of M company is \$40 on Monday. After the close on Monday, M company announces a dividend distribution of \$2 per share. The adjusted closing price for the stock would then be \$38 (\$40–\$2). For stock dividends, if M company announces a 2:1 stock dividend instead of a cash dividend, the adjusted closing price calculation will change. A 2:1 stock dividend means that for every share, an investor owns who will receive two more shares. In this case, the adjusted closing price calculation will be \$40*(1/(2 + 1)). This will give you a price of \$13.33, rounded to the nearest penny. For stock splits, if M company announces a 2:1 stock split, investors will receive an extra share for every share they already own. This time the calculation will be \$40*(1/(1*2)), resulting in an adjusted closing price of \$20.

4.3 Data preprocessing

By removing the noise data from the experimental data, we use the experimental data as much as possible in order to obtain better prediction model experimental results. In the raw experimental data set of those six stock indexes, there are some abnormal data. For example, the records of two consecutive days have exactly the same data, the data are empty, the record of trading volume is equal to zero, and so on. In the experimental data set of the four major stock indexes, we should first eliminate all incorrect or noise data. After removing noise or false data, the overall running of the closing prices of the six stock market indexes is shown in Fig. 8.

In experiment, we need to standardize or preprocess the experimental data before conducting experiments. Let $X_i = \{x_i(t)\}$ be the ith index of the stock market at time t, where $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. The logarithmic daily return is defined by $G_i(t) = \log(x_i(t))—\log(x_i(t—1))$. The normalized daily return is defined as $R_i(t) = (G_i(t)—\bar{G}_{(t)}) / \delta$, where $\bar{G}_{(t)}$ is the mean values of the all elements of time series $G_i(t)$ and δ is the standard deviation of the time series $G_i(t)$ according to the following Formula 22.

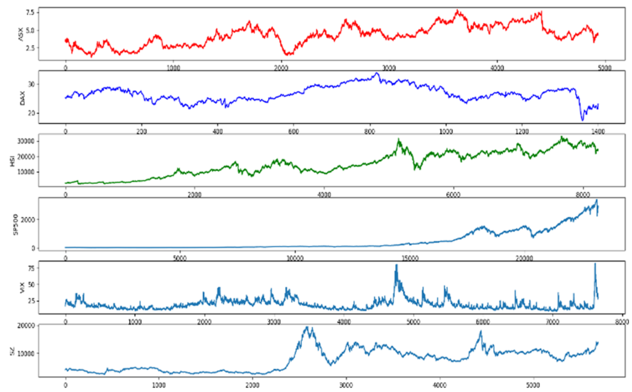

Fig. 8 The overall running of the six stock market indices

$$\delta = \sqrt{\frac{1}{n-1}\sum\nolimits_{i=1}^{n}(x_i - \bar{x})^2} \tag{22}$$

where $\bar{x} = \frac{1}{n}\sum\nolimits_{i=1}^{n} x_i$.

5 Experiments

We used two types of experimental data. One is the real data in the financial field, the other is the artificial simulation experiment data generated by a computer. To compare with other models, four traditional prediction models are selected to test the same set of data. In this section, we will explain and analyze the experimental results from three aspects: introduction and results analysis of other comparative methods, results analysis of artificial simulation data and experimental results analysis of real data in the financial field.

5.1 Analysis of experimental results of other methods

To compare the disadvantages, advantages, and effects of various prediction models, other four prediction methods are selected to carry out prediction experiments on the same set of experimental data. Table 4 shows the experimental results. We introduced the LSTM method briefly above. Here, we will introduce three other methods in detail.

5.1.1 SVR

The full name of SVR is support vector regression. It is a support vector regression method based on the support vector machine (SVM). The *libsvm* toolbox is the toolbox software of the support vector machine. According to the description of the *libsvm* toolbox, machine learning is controlled by the penalty function and loss function. Table 1 shows 10 parameters of the SVR. The support vector regression based on linear kernel uses *liblinear* instead of *libsvm* in the experiment.

5.1.2 BARDR

The BARDR is the abbreviation of Bayesian ARD regression. One of the main uses of BARDR is to fit the regression model weight. The premise of using this method is to assume that the regression model weight obeys Gaussian distribution. The BARDR has 12 parameters. The settings of these parameters are shown in Table 2. Assuming that the weight of the regression model is a Gaussian distribution, the BARDR uses parameters *lambda* and *alpha* to represent the accuracy of weight distribution and noise distribution, respectively.

5.1.3 RFR

The abbreviation of random forest regression is the RFR. Random tree is one of the widely used estimation methods. Many classification decision trees are constructed to fit the subsamples of the data set. The RFR controls over fitting and improves prediction accuracy by averaging the fitting values of these subsamples.

5.1.4 KNR

The KNR is the abbreviation of k-nearest neighbor regression. It uses local interpolation to predict the target associated with the nearest K-nearest neighbors in the training set. The setting of eight parameters of the KNR is shown in Table 3.

Table 4 shows the prediction results of six prediction methods for seven time series data. In order to distinguish and compare easily, the results of the two best prediction methods are shown in bold. We observe the six prediction methods (SVR, RFR, BARDR, KNR, LSTM, and LSTM-VMD) in Table 4. We find that the BARDR is the best method to predict the results of Cos artificial time series data, followed by the LSTM-VMD. When forecasting HSI, SZ, DAX, ASX, and VIX stock index prices, the LSTM-VMD and BARDR method have good results. The LSTM-VMD method has the best prediction effect in the prediction of HSI, SZ, DAX, ASX, and VIX stock index prices. However, the BARDR method is better than the LSTM-VMD method in predicting SP500 stock index price. Therefore, among the six prediction methods mentioned above, the LSTM-VMD and BARDR method are the best methods to predict the five above time series data, but the comprehensive prediction effect of the LSTM-VMD method is the best among the six above methods.

From Table 4, it is easy to find the prediction effect of cosine artificial simulated time series data. Except for the

Table 1 The value of eight parameters of the SVR

Name	Type	Value	Meaning
epsilon	float	0.0	Epsilon parameter in the epsilon-insensitive loss function
tol	float	1e-4	Tolerance for stopping criteria
C	float	1.0	Regularization parameter. The strength of the regularization is inversely proportional to C. Must be strictly positive
loss	'epsilon' or 'squared_epsilon'	epsilon	Specifies the loss function. The epsilon loss (standard SVR) is the L1 loss, while the squared epsilon loss ('squared_epsilon') is the L2 loss
fit_intercept	bool	True	Whether to calculate the intercept for this model. If set to false, no intercept will be used in calculations
intercept_scaling	float	1.0	When fit_intercept is True, a "synthetic" feature with constant value equals to intercept_scaling is appended to the instance vector
dual	bool	True	Select the algorithm to either solve the dual or primal optimization problem. Prefer dual = False when n_samples > n_features
verbose	int	1	Enable verbose output. Note that this setting takes advantage of a per-process runtime setting in liblinear that, if enabled, may not work properly in a multithreaded context
random_state	int	0	Controls the pseudo-random number generation for shuffling the data. Pass an int for reproducible output across multiple function calls
max_iter	int	1000	The maximum number of iterations to be run

Table 2 The value of eight parameters of the BARDR

Name	Type	Value	Meaning
n_iter	int	300	Maximum number of iterations
tol	float	1e-3	Stop the algorithm if w has converged
alpha_1	float	1e-6	Hyper-parameter: shape parameter for the Gamma distribution prior over the alpha parameter
alpha_2	float	1e-6	Hyper-parameter: inverse scale parameter (rate parameter) for the Gamma distribution prior over the alpha parameter
lambda_1	float	1e-6	Hyper-parameter: shape parameter for the Gamma distribution prior over the lambda parameter
lambda_2	float	1e-6	Hyper-parameter: inverse scale parameter (rate parameter) for the Gamma distribution prior over the lambda parameter
compute_score	bool	False	If True, compute the objective function at each step of the model
threshold_lambda	float	10,000	The threshold for removing (pruning) weights with high precision from the computation
fit_intercept	bool	True	whether to calculate the intercept for this model. If set to false, no intercept will be used in calculations
normalize	bool	False	This parameter is ignored when fit_intercept is set to False. If True, the regressors X will be normalized before regression by subtracting the mean and dividing by the l2-norm
copy_X	bool	True	If True, X will be copied; else, it may be overwritten
verbose	bool	False	Verbose mode when fitting the model

Table 3 The value of eight parameters of the KNR

Name	Type	Value	Meaning
n_neighbors	int	5	Number of neighbors to use by default for k-neighbors queries
weights	'uniform', or 'distance'	uniform	weight function used in the prediction
algorithm	{'auto', 'ball_tree', 'kd_tree', 'brute'}	auto	The algorithm used to compute the nearest neighbors
leaf_size	int	30	Leaf size passed to BallTree or KDTree
p	int	2	Power parameter for the Minkowski metric
metric	string or callable	minkowski	the distance metric to use for the tree
metric_params	dict	None	Additional keyword arguments for the metric function
n_jobs	int	1	The number of parallel jobs to run for neighbors search

SVR and the LSTM, the prediction results of these two methods are slightly worse, and the other four methods have very good prediction results. The R^2 evaluation indexes of the RFR, BARDR, KNR, and LSTM-VMD were all greater than 0.995. From the R^2 evaluation index, its value is greater than 0.999999, so the BARDR obtained the best prediction effect. However, from the RMSE evaluation index, the RFR, BARDR, and LSTM-VMD prediction methods have the best experimental results, and their RMSE evaluation values are all less than 0.02. Among them, the experimental results of the BARDR are the smallest, and the RMSE evaluation value is 0.000542. From the three evaluation indices, the experimental value of the BARDR prediction method is the best. For the time series with regularity and stability, these results show that the above six methods have a better prediction effect, especially the BARDR method. The prediction results of the RFR prediction method, one of the six methods, are shown in the figure, as shown in Fig. 9. For clarity and readability, the result graph in Fig. 9 shows only the last 400 prediction results among the prediction results.

By observing the experimental results of the SVR prediction method in Table 4, it is found that the prediction effect of the SVR prediction method on SZ, DAX, VIX, ASX, and Cos time series data is better than that of HSI and SP500. Figure 10 shows the SVR prediction results for the SZ stock index time series data. The SVR prediction method has a good prediction effect on SZ, DAX, ASX, and VIX stock time series data. But the prediction effect of the SVR prediction method for SP500 and HSI stock time series data is very poor. This conclusion can be drawn according to the R^2 index of the experimental results of HSI and SP500 stock index by SVR. The SVR prediction method has the best experimental results for Cos artificial

Table 4 The prediction results of the time series

Models	Cos	SP500	HSI	SZ	DAX	ASX	VIX
SVR							
RMSE	0.107773	1219.904783	1334.991215	4.094287	0.653453	0.206921	3.430912
MAE	0.089628	990.639020	1282.862498	2.757474	0.415796	0.146022	1.790241
R^2	0.975407	−1.665447	−1.088731	0.942958	0.938262	0.959223	0.885500
RFR							
RMSE	0.022769	11.972911	1.882800	2.212179	0.636910	0.257781	4.137058
MAE	0.017229	9.358394	1.548856	1.612024	0.364778	0.149846	1.717665
R^2	0.998902	−1.567542	−1.330839	0.983348	0.941348	0.936714	0.833518
BARDR							
RMSE	**0.000542**	**114.140660**	**326.487107**	**1.759013**	**0.387025**	**0.110975**	**2.048969**
MAE	**0.000480**	**110.146472**	**233.641980**	**1.171820**	**0.256125**	**0.075355**	**1.165715**
R^2	**0.999999**	**0.998275**	**0.992213**	**0.989471**	**0.978343**	**0.988271**	**0.959163**
KNR							
RMSE	0.046355	1198.484251	553.641943	3.313639	0.740763	0.303352	4.607431
MAE	0.035261	937.309021	452.300293	2.421727	0.445212	0.186848	2.006647
R^2	0.995450	−1.572662	−1.239516	0.962636	0.920661	0.912361	0.793508
LSTM RMSE	0.052190	**30.261563**	306.481549	1.963855	0.654662	0.135236	2.414541
MAE	0.046363	**19.164419**	224.250031	1.500719	0.504645	0.106639	1.216333
R^2	0.948028	**0.997128**	0.980270	0.980562	0.962750	0.931165	0.940744
Proposed LSTM-VMD							
RMSE	**0.018238**	0.393651	**0.043551**	**0.527323**	**0.166337**	**0.011807**	**0.765925**
MAE	**0.016470**	0.296235	**0.037125**	**0.456034**	**0.125283**	**0.008971**	**0.458708**
R^2	**0.997846**	0.995140	**0.998346**	**0.998284**	**0.996376**	**0.999877**	**0.991299**

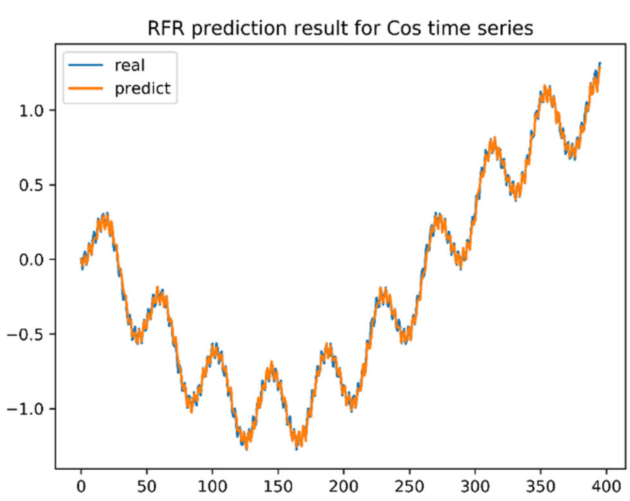

Fig. 9 Prediction result of the RFR method for the Cos time series

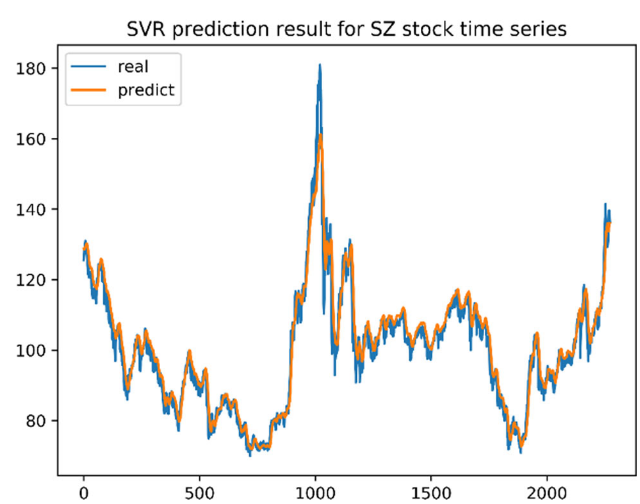

Fig. 10 Prediction results of the SVR method for the SZ stock time series

time series data. The possible reason is that the change of the cos time series is regularity and stability, which indicates that the SVR is suitable for the prediction of time series data with stable and good regularity. It can be inferred that SZ, ASX, VIX, and DAX are regularity and stability. The experimental results of the SVR prediction method for SP500 and HSI time series data are very poor. The SVR prediction method is not suitable to predict HSI and SP500. This shows that the regularity and stability of the SP500 and HSI time series data are poor.

By observing the experimental results of the KNR and the RFR prediction methods in Table 4, it is easy to find that these two methods have similar experimental results and prediction performance as that of the SVR method. The prediction results of DAX, SZ, ASX, VIX, and Cos time series data are better than those of HSI and SP500. Especially in the prediction of Cos time series data, the experimental results are very good. Similarly, the KNR and the RFR prediction methods have poor performance in predicting SP500 and HSI time series data. As we all know, the stock index prices are affected by human factors. The change of stock index prices is complex, unstable, and irregular. We can draw a conclusion that the KNR and the RFR are not suitable for forecasting stock index prices from the experimental results. Therefore, the KNR and the RFR prediction methods are not suitable for the prediction of time series with unstable and irregular changes, but they are suitable for the prediction of data with regular and stable changes similar to the Cos time series.

Observing the results of the BARDR and the LSTM in Table 4, we found that the experimental results of the BARDR and the LSTM prediction methods are better. The experimental values of R^2 are all greater than 0.93, especially the experimental values of R^2 evaluation indexes of the BARDR prediction method are greater than 0.95. This shows that the BARDR and the LSTM prediction methods not only have a better prediction effect on SZ, VIX, ASX, DAX, and Cos but also have a more significant prediction effect on HSI and SP500 time series data than other methods. Especially, the prediction results of SP500 and HSI time series data are much better than the KNR, RFR, and SVR. Although the change of stock index price is complex, irregular and unstable, the prediction results of the LSTM and the BARDR are still very good. We can draw a conclusion that the LSTM and BARDR methods can be used to predict the stock index time series. Therefore, the LSTM and BARDR are suitable for forecasting time series with irregular and unstable changes. The result of the BARDR prediction method for DAX is shown in Fig. 11. Observing Fig. 11, we can find that the BARDR prediction method has a good prediction effect on DAX.

Observing the experimental results of the BARDR and the LSTM prediction methods in Table 4, the experimental values of R^2 are all greater than 0.93, it is easy to find that the two methods have a good prediction effect for the seven different time series data in the experiment. By observing and comparing the experimental values of R^2 evaluation indexes of the BARDR and the LSTM prediction methods, we found that the prediction effect of the BARDR method is better than that of the LSTM method. Except in the prediction effect of the SP500 stock index time series, the

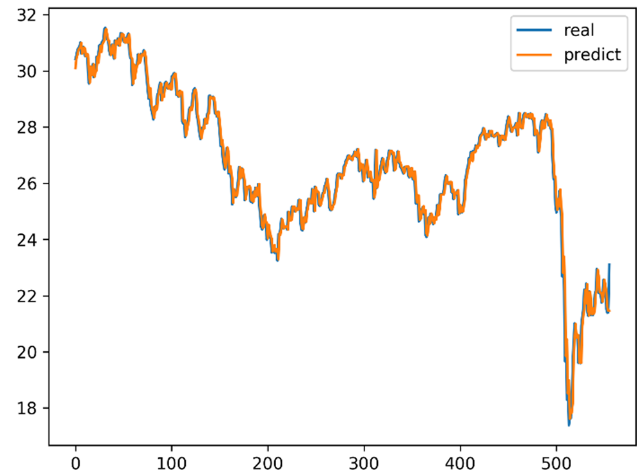

Fig. 11 Prediction results of BARDR method for DAX time series

LSTM method is better than the BARDR method in the prediction effect of the other six time series. The prediction performance of the BARDR is better than that of the LSTM, but the experimental training time of the BARDR is several times longer than that of the LSTM, therefore, the BARDR method is not suitable for real-time prediction. In addition, although the experimental results of the LSTM method are worse than those of the BARDR method in our experiment, the training parameters are fewer and the experimental time is shorter than the BARDR. The reason for the above experimental results is that the LSTM model we used in the experiment is relatively simple, without setting more neurons and parameters, and the number of iterations is also very small. Figure 12 shows the LSTM model we used in our experiment, which uses 50 neurons and sets the iteration number to 20. In the future work, we can improve the prediction performance by increasing the number of neurons, optimizing training parameters, and training iterations of the LSTM. We think that the prediction effect of the LSTM method will exceed that of the

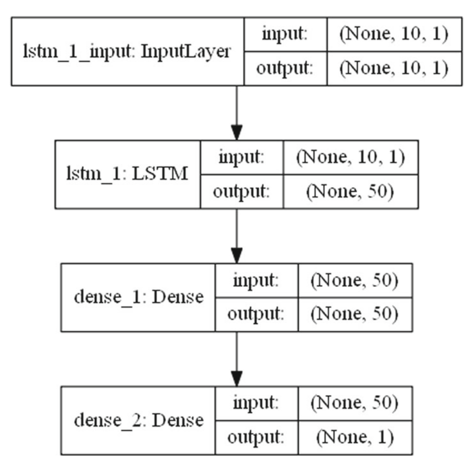

Fig. 12 Visualization flowchart of the LSTM model in the experiment

BARDR method by training iterations reasonably, optimizing training parameters, and increasing the number of neurons based on the principle analysis of the LSTM.

5.2 Analysis of experimental results based on artificial simulation data

In order to test the effectiveness and correctness of our proposed method, we first use the artificial simulation experimental data to verify it. In order to obtain sufficient and effective experimental results, the artificially generated experimental data should be moderate. When the data are too small, the experiment cannot effectively verify the proposed method. When the data are too long, the experimental time will increase significantly. Therefore, we choose the length of artificial simulation experiment data as 10,000, as shown in Table 5.

Before discussing the experimental results of the proposed method applied to the actual stock index time series data, we first use the proposed LSTM-VMD method to predict the artificial simulation time series data. The simulation experiment can verify the correctness and effectiveness of our proposed method.

By carefully observing the Cos data column in Table 4 and comparing the three indicators of the LSTM method and the LSTM-VMD method, it is easy to find that under the same number of neurons and iteration times, the experimental result of the R^2 evaluation index of the LSTM method is 0.9480, while that of the LSTM method is 0.9978. There is little difference between the LSTM and the LSTM-VMD. But our proposed LSTM-VMD method is better than the traditional LSTM method in three experimental evaluation indexes. These three pairs of results show that the two methods have a similar prediction effect on Cos simulation time series data. Among them, the LSTM-VMD method has the best prediction effect, which shows that our proposed method is effective and can improve the prediction effect of time series. Observing the Cos data column in Table 4, the experimental results of the LSTM method are worse than those of the LSTM-VMD method, but the difference is very small. After in-depth analysis, although the Cos time series is a mixture of multiple cos functions, in fact, the Cos time series data itself still have certain regularity. The traditional LSTM method can find the internal rules of the Cos time series data, while our proposed LSTM-VMD method uses the VMD algorithm to decompose the Cos time series data into several sub-time series more regularly, but it has little effect on the prediction effect. To sum up, the comprehensive effect of our proposed LSTM-VMD prediction method is better than the traditional LSTM method, which shows that our proposed LSTM-VMD prediction method is correct and effective.

5.3 Analysis of experiment results based on real data

To verify the practicability and effectiveness of the LSTM-VMD prediction method, we use the LSTM-VMD prediction method to predict the actual stock index time series of six different regions. The six stock index are SZ, HSI, SP500, VIX, ASX, and DAX. As we all know, the changes of stock prices are closely related to policy and social events, so the time series of the stock index is seriously affected by human factors, which reflects the complexity and instability of stock index time series. These six time series are very representative. They are all the main stock indexes of the financial market in the world. At the same time, they come from six different regions or countries in the world, which can reflect the economic changes of different countries.

By comparing and analyzing the three evaluation values of R^2, MAE, RMSE and of LSTM-VMD and LSTM prediction methods in Table 4, it is easy to find that except in the prediction effect of the SP500 stock index time series, the LSTM method is worse than the LSTM-VMD method in the prediction effect of the other five time series. The prediction experimental results of the LSTM-VMD in five time series data are better than those of the traditional LSTM. The experimental results show that our proposed LSTM-VMD is better than the traditional LSTM. The prediction results of the LSTM-VMD method for the last 200 data of the DAX time series are shown in Fig. 13.

Through careful observation of Table 4, we can find that on the five time series data of SZ, HSI, VIX, DAX, and ASX, the experimental results of the proposed LSTM-VMD are much better than those of the traditional LSTM. There are two main reasons for this result. For one thing, the proposed method decomposes SZ, HSI, VIX, DAX, and ASX time series by the VMD, thus decomposing the

Table 5 The length of the data in the experiment

Name of data	Cos	SP500	HSI	SZ	DAX	ASX	VIX
All data length	10,000	23,204	8237	5113	1401	4937	7654
Train data length	9000	20,884	7413	4601	1261	4443	6495
Validate data length	900	2088	741	460	126	444	650
Test data length	1000	2320	824	512	140	494	1159

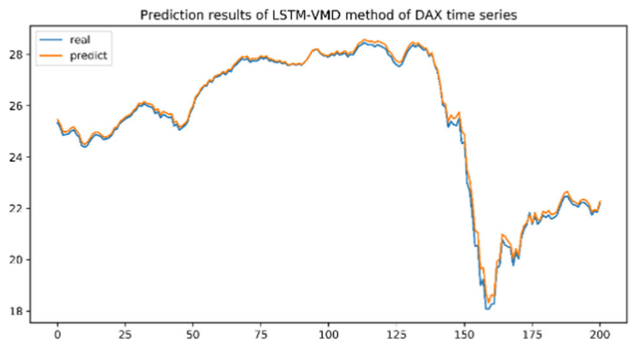

Fig. 13 Prediction results of the LSTM-VMD method of DAX time series

complex original SZ, ASX, VIX, HSI, and DAX into more stable and regular sub-time series. Compared with the complex original time series, because the changes of multiple sub-time series are more stable and regular, we think it is more convenient to use the method to predict these sub-time series. Therefore, the experimental results of a few sub-time series predictions are better and more accurate than those of direct prediction of complex original time series. For another, although the time series of SZ, DAX, VIX, ASX, and HSI are complex, every subsequence of them decomposed by VMD is stable and regular. To understand clearly the decomposition of SZ, DAX, VIX, ASX, and HSI and the VMD decomposition, Fig. 14 shows the change of six sub time series that is each sub-time series of the VMD decomposition of the DAX time series. The original change of the DAX is the second subgraph from the top in Fig. 8. All subgraphs in Fig. 14 are six subgraphs of the VMD decomposition of DAX. The VMD decomposition subgraph has the characteristics of smoother, more stable, more predictable, and more regular than the original time series, and to get good prediction experimental results.

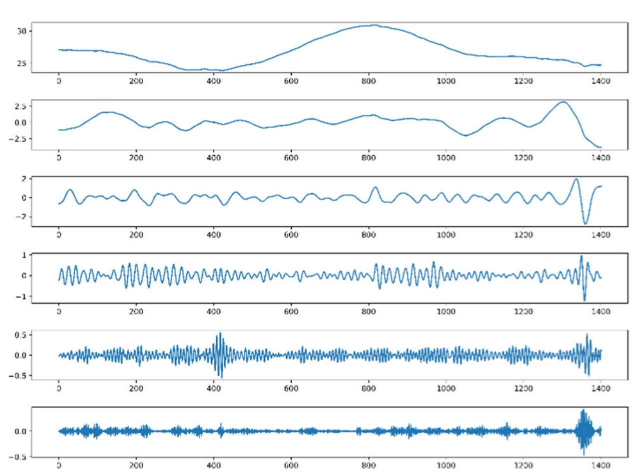

Fig. 14 The VMD decomposition graphs of the DAX time series

Carefully observe the prediction experimental results of the LSTM and LSTM-VMD prediction method in Table 4 of ASX, DAX, SZ, VIX, and HSI, we can find that our proposed LSTM-VMD prediction method has a better experimental results than the traditional LSTM. It can be inferred that the LSTM-VMD prediction method proposed by us has better prediction effect and prediction performance than the traditional LSTM method. The experimental results of the MAE, R^2, and RMSE used in the experiment confirmed the above relationship. According to the principle of the above three evaluation indices, if the MAE or RMSE experimental value of a prediction method is smaller, the prediction effect of the method is better. However, if the R^2 value of a prediction method is larger, the prediction effect is better.

Besides, this hybrid prediction method is not sensitive to the value of its initial parameters. Different parameters have different training speed and training efficiency of the model. However, in the case of sufficient training, the initial values of the parameters have little difference to the results of the final model in our method. The sensitivity of the model to parameters is much lower than that of data. In this paper, we are concerned that this method has a better predictive effect than other methods, rather than pursuing the best effect of this method. In actual application, the users need to adjust the parameters reasonably to make the method have the best predictive effect. If we need to adjust the parameters, we can use some method based on machine learning for dynamic adjustment. Due to the relatively large variation range of the parameters, the parameters of the adjustment method cannot be adjusted manually, but the adjustment or training parameters of the machine learning method are used.

A hybrid prediction method based on the LSTM and the VMD is proposed to predict the stock index. The experimental results show that the prediction results of the hybrid prediction method are much better than those of the traditional method in most cases. In this paper, the prediction results of our hybrid method based on LSTM and VMD are compared with the SVR, RFR, BARDR, KNR, and LSTM. Although the prediction effect of the LSTM-VMD hybrid prediction method in the five stock index time series is better than the LSTM method, the prediction results of the LSTM-VMD hybrid method on the SP500 time series are worse than the LSTM. Different time series have different characteristics, and different prediction methods are suitable for different time series. The experimental results show that the SVR, RFR, BARDR, KNR, LSTM, and LSTM-VMD reflect the advantages and disadvantages of prediction experiments on different time series. In practical application, people can choose different methods according to the actual situation and apply them to specific practical fields. The experimental results provide a valuable reference for people to use in specific practical scenes.

6 Conclusion and future work

In this paper, we propose a hybrid method LSTM-VMD for stock index short-term time series prediction based on the LSTM neural network and the VMD decomposition method. Our proposed method LSTM-VMD is based on the idea of dividing and solving complex problems. The proposed prediction method combines the advantages of the VMD and the LSTM. The hybrid method LSTM-VMD uses the VMD method to decompose complex time series into several relatively smooth, regular, and stable sub-time series. The LSTM neural network is used to train each time series in the LSTM-VMD. The prediction process of the proposed LSTM-VMD is very simple which has only two steps. First, the LSTM-VMD uses the LSTM to predict the value of each sub-time series. Then, the LSTM-VMD fuses the prediction results of a few sub-time series to form prediction results of complex original time series. To verify fully the effectiveness and practicability of the LSTM-VMD method, we select six representative time series data to test. The experimental results show that the proposed method LSTM-VMD has a good comprehensive experimental effect and application value. However, this method also has some shortcomings. The prediction results of the proposed method for time series with stable and regular changes are not particularly good, which is a little unexpected.

The research of time series analysis and prediction method has a rapid and long history development. The prediction effect of traditional methods cannot meet the higher and higher requirements of practical application in some fields. The main goal of scholars is to improve the prediction effect. Based on the research work of this paper, we still have a lot of work to do in the future. In the future, we will combine the VMD with many other methods. We can also combine LSTM with another decomposition method such as wavelet decomposition. We will study how to determine the number of VMD components in the future. Is it possible for different time series to decompose with different number of components to achieve the best effect? In addition, we intend to combine the optimization algorithm (Abualigah and Diabat 2021; Abualigah, et al. 2021; Mhha 2021; Elaziz et al. 2021; Altabeeb, et al. 2021) or text clustering (Abualiga and Qasim 2018) to improve the prediction effect of the proposed method.

Acknowledgements This work was supported in part by the Scientific Research Fund of Hunan Provincial Education under Grants 20C1487 and 19C1472, Research results of Educational Science Planning in Hunan Province under Grant XJ212259, the Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province under Grants ZNKZD2020-1 and ZNKZ2018-5, Key scientific research projects of Huaihua University under Grant HHUY2019-08, Project of Huaihua City Social Science Achievement Review Committee under Grant HSP2021YB101 and HSP2021YB102, Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities.

Authors' Contributions Yang Yujun contributed to all aspects of this work. Yang Yimei and Zhou Wang conducted the experiment and analyzed the data. All authors reviewed the manuscript.

Data Availability Data are fully available without restriction. The original experimental data can be downloaded from Yahoo Finance for free(http://finance.yahoo.com).

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

References

Abualiga H, Qasim LM (2018) Feature selection and enhanced krill herd algorithm for text document clustering. Stud Comput Intell (2018).

Abualigah L et al. (2021) Matlab code of aquila optimizer: a novel meta-heuristic optimization algorithm. Comput Ind Eng

Abualigah L, Diabat A (2021) Advances in sine cosine algorithm: a comprehensive survey. Artif Intell Rev 3

Alhazbi S, Said AB, Al-Maadid A (2020) Using deep learning to predict stock movements direction in emerging markets: the case of Qatar stock exchange. 2000 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, pp 440–444

Altabeeb AM et al (2021) Solving capacitated vehicle routing problem using cooperative firefly algorithm. Appl Soft Comput 1:107403

Bukhari H, Raja MAZ, Sulaiman M, Islam S, Shoaib M, Kumam P (2020) Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2985763

Chen K, Zhou Y, Dai F (2015) A LSTM-based method for stock returns prediction: a case study of China stock market. 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, 2015, pp 2823–2824. https://doi.org/10.1109/BigData.2015.7364089

Cheng, Shiu H (2014) A novel GA-SVR time series model based on selected indicators method for forecasting stock price. 2014 international conference on information science, electronics and electrical engineering, Sapporo, 2014, pp 395–399

Chong C, Han FC (2017) Park, Deep learning networks for stock market analysis and prediction: methodology data representations and case studies. Expert Syst Appl 83:187–205

Day M, Lee C (2016) Deep learning for financial sentiment analysis on finance news providers. 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), San Francisco, CA, pp 1127–1134

Dragomiretskiy K, Zosso D (2014) Variational mode decomposition. IEEE Trans Signal Process 62(3):531–534

Elaziz MA, Abualigah L, Attiya I (2021) Advanced optimization technique for scheduling IoT tasks in cloud-fog computing environments. Future Gener Comput Syst 124.9
Gao T, Chai Y (2018) Improving stock closing price prediction using recurrent neural network and technical indicators. Neural Comput 30(10):2833–2854
Gendeel M, Yuxian Z, Aoqi H (2018) Performance comparison of ANNs model with VMD for short-term wind speed forecasting. IET Renew Power Gener 12(12):1424–1430
Gers FA, Schmidhuber J (2000) Recurrent nets that time and count. Proceedings of IEEE-INNS-ENNS International Joint Conference Neural Network (IJCNN) Neural Comput. New Challenges Perspect. New Millennium, vol 3, pp 189–194, 2000
Guo Y, Han S, Shen C, Li Y, Yin X, Bai Y (2018) An Adaptive SVR for High-Frequency Stock Price Forecasting. IEEE Access 6:11397–11404
Han L, Zhang R, Wang X, Bao A, Jing H (2019) Multi-step wind power forecast based on VMD-LSTM. IET Renew Power Gener 13(10):1690–1700
Hasan OK, Akyokuş S (2017) Predicting financial market in big data: Deep learning. 2017 International conference on computer science and engineering (UBMK), Antalya, pp 510–515.
Hassan (2017) Exploiting noisy data normalization for stock market prediction. J Eng Appl Sci 12(1):69–77
Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780
Hu Y, Sun X, Nie X, Li Y, Liu L (2019) An enhanced LSTM for trend following of time series. IEEE Access 7:34020–34030. https://doi.org/10.1109/ACCESS.2019.2896621
Idrees SM, Alam MA, Agarwal P (2019) A prediction approach for stock market volatility based on time series data. IEEE Access 7:17287–17298
Jeon S, Hong B, Chang V (2018) Pattern graph tracking-based stock price prediction using big data. Future Gener Comput Syst 80:171–187
Kaya MY, Karsligil ME (2010) Stock price prediction using financial news articles. 2010 2nd IEEE international conference on information and financial engineering, Chongqing, pp 478–482.
Khare K, Darekar O, Gupta P, Attar VZ (2017) Short term stock price prediction using deep learning. 2017 2nd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), Bangalore, pp 482–486
Kwon Y, Sun H-D (2011) A hybrid system integrating a piecewise linear representation and a neural network for stock prediction. Proceedings of 2011 6th International Forum on Strategic Technology, Harbin, Heilongjiang, pp 796–799
Lai CY, Chen R, Caraka RE (2019) Prediction stock price based on different index factors using LSTM. International conference on machine learning and cybernetics (ICMLC), Kobe, Japan, 2019, pp 1–6
Lee J, Kim R, Koh Y, Kang J (2019) Global stock market prediction based on stock chart images using deep Q-network. IEEE Access 7:167260–167277. https://doi.org/10.1109/ACCESS.2019.2953542
Lim M, Yeo CK (2020) Harvesting social media sentiments for stock index prediction. IEEE 17th annual consumer communications and networking conference (CCNC), Las Vegas, NV, USA, pp 1–4
Liu F, Li X, Wang L (2019) Exploring cluster stocks based on deep learning for stock prediction. 2019 12th international symposium on computational intelligence and design (ISCID), Hangzhou, China, pp 107–110
Liu S, Liao G, Ding Y (2018) Stock transaction prediction modeling and analysis based on LSTM. 2018 13th IEEE conference on industrial electronics and applications (ICIEA), Wuhan, 2018, pp 2787–2790. https://doi.org/10.1109/ICIEA.2018.8398183
Liu H, Song B (2018) Stock price trend prediction model based on deep residual network and stock price graph. 11th international symposium on computational intelligence and design (ISCID), Hangzhou, China, pp 328–331.
Ma S, Gao L, Liu X, Lin J (2019) Deep learning for track quality evaluation of high-speed railway based on vehicle-body vibration prediction. IEEE Access 7:185099–185107. https://doi.org/10.1109/ACCESS.2019.2960537
Mankar T, Hotchandani T, Madhwani M, Chidrawar A, Lifna CS (2018) Stock market prediction based on social sentiments using machine learning. 2018 International conference on smart city and emerging technology (ICSCET), Mumbai, pp 1–3
Mhha B et al. (2021) Development and application of slime mould algorithm for optimal economic emission dispatch. Expert Syst Appl
Minh Dang L, Sadeghi-Niaraki A, Huy HD, Min K, Moon H (2018) Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. IEEE Access 6:55392–55404. https://doi.org/10.1109/ACCESS.2018.2868970
Ojo SO, Owolawi PA, Mphahlele M, Adisa JA (2019) Stock market behaviour prediction using stacked LSTM networks. International multidisciplinary information technology and engineering conference (IMITEC), Vanderbijlpark, South Africa, 2019, pp 1–5. https://doi.org/10.1109/IMITEC45504.2019.9015840
Shi L, Teng Z, Wang L, Zhang Y, Binder A (2019) DeepClue: visual interpretation of text-based deep stock prediction. IEEE Trans Knowl Data Eng 31(6):1094–1108
Sismanoglu MA, Onde FK, Sahingoz OK (2019) Deep learning based forecasting in stock market with big data analytics. Scientific meeting on electrical-electronics and biomedical engineering and computer science (EBBT), Istanbul, Turkey, pp 1–4
Sohangir S, Wang D (2018) Finding expert authors in financial forum using deep learning methods. 2018 Second IEEE international conference on robotic computing (IRC), Laguna Hills, CA, pp 399–402
Sun Z, Zhao M (2020) Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis. IEEE Access 8:134422–134434
Sun Z, Zhao S, Zhang J (2019) Short-term wind power forecasting on multiple scales using VMD Decomposition, K-Means clustering and LSTM principal computing. IEEE Access 7:166917–166929
Tang L, Sheng H, Tang L (2009) Stock returns prediction using manifold wavelet Kernel. International conference on electronic commerce and business intelligence, Beijing, pp 306–309
Upadhyay A, Pachori RB (2017) Speech enhancement based on mEMD-VMD method. Electron Lett 53(7):502–504
Wang J, Sun T, Liu B, Cao Y, Wang D (2018) Financial markets prediction with deep learning. 17th IEEE international conference on machine learning and applications (ICMLA), Orlando, FL, pp 97–104
Wang C, Li H, Huang G, Ou J (2019) Early fault diagnosis for planetary gearbox based on adaptive parameter optimized VMD and singular kurtosis difference spectrum. IEEE Access 7:31501–31516
Waqar M, Dawood H, Guo P, Shahnawaz MB, Ghazanfar MA (2017) Prediction of stock market by principal component analysis. 2017 13th 2017 International conference on computational intelligence and security (CIS), Hong Kong, pp 599–602
Wei D (2019) Prediction of stock price based on LSTM neural network. International conference on artificial intelligence and advanced manufacturing (AIAM), Dublin, Ireland, 2019, pp 544–547. https://doi.org/10.1109/AIAM48774.2019.00113
Wei P, Wang H (2019) Evaluation method of spindle performance degradation based on VMD and random forests. J Eng 23:8862–8866

Wei Y, Wang Z, Xu M, Qiao S (2017) An LSTM method for predicting CU splitting in H. 264 to HEVC transcoding. Proceedings of IEEE visual communication image process (VCIP), pp 1–4

Weng B, Ahmed MA, Megahed FM (2017) Stock market one-day ahead movement prediction using disparate data sources. Expert Syst Appl 79:153–163

Xi G (2018) A novel stock price forecasting method using the dynamic neural network. 2018 International conference on robots and intelligent system (ICRIS), Changsha, pp 242–245. https://doi.org/10.1109/ICRIS.2018.00069

Xiao Y, Che W, Wang Z, Yang C (2013) The research of morphological characteristics in time series of stock prices based on CBR. 2013 Third international conference on intelligent system design and engineering applications, Hong Kong, pp 1518–1521.

Xu Y, Gao Y, Li Z, Lu M (2020) Detection and classification of power quality disturbances in distribution networks based on VMD and DFA. CSEE J Power Energy Syst 6(1):122–130

Yang YJ, Yang YM (2020) Hybrid method for short-term time series forecasting based on EEMD. IEEE Access 8:61915–61928

Yang Y, Yang Y, Li J (2017) Role of mean in the multifractal analysis of financial time series. 2017 14th International computer conference on wavelet active media technology and information processing (ICCWAMTIP), Chengdu, pp 74–78

Yang Y, Li J, Yang Y (2017b) The cross-correlation analysis of multi property of stock markets based on MM-DFA. Phys A Stat Mech Appl 481:23–33

Yıldırım S, Jothimani D, Kavaklioğlu C, Başar A (2019) Deep learning approaches for sentiment analysis on financial microblog dataset. IEEE international conference on big data (Big Data), Los Angeles, CA, USA, pp 5581–5584.

Yu Y, Wang S, Zhang L (2017) Stock price forecasting based on BP neural network model of network public opinion. 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, pp 1058–1062. https://doi.org/10.1109/ICIVC.2017.7984716

Yuan X, Yuan J, Jiang T, Ain QU (2020) Integrated long-term stock selection models based on feature selection and machine learning algorithms for China stock market. IEEE Access 8:22672–22685

Yujun Y, Yimei Y, Jianping L (2016) Research on financial time series forecasting based on SVM. 2016 13th International computer conference on wavelet active media technology and information processing (ICCWAMTIP), Chengdu, pp 346–349.

Yujun Y, Jianping L, Yimei Y (2016b) An efficient stock recommendation model based on big order net inflow. Math Probl Eng 5725143:1–15

Yujun Y, Jianping L, Yimei Y (2017) Multiscale multifractal multiproperty analysis of financial time series based on Rényi entropy. Int J Mod Phys C 28(02)

Zhang L, Liu N, Yu P (2012) A novel instantaneous frequency algorithm and its application in stock index movement prediction. IEEE J Select Topics Signal Process 6(4):311–318

Zhang Y, Zhao Y, Gao S (2019) A novel hybrid model for wind speed prediction based on VMD and neural network considering atmospheric uncertainties. IEEE Access 7:60322–60332

Zhang J et al (2020) Can the exchange rate be used to predict the shanghai composite index? IEEE Access 8:2188–2199

Zhao Q, Bao K, Wang J, Han Y, Wang J (2019) An online hybrid model for temperature prediction of wind turbine gearbox components. Energies 12(20):3920

Zheng C, Zhu J (2017) Research on stock price forecast based on gray relational analysis and ARMAX model. 2017 International Conference on Grey Systems and Intelligent Services (GSIS), Stockholm, pp 145–148. https://doi.org/10.1109/GSIS.2017.8077689

Zhou P, Chan KCC, Ou CX (2018) Corporate communication network and stock price movements: insights from data mining. IEEE Trans Comput Soc Syst 5(2):391–402

Zhou S, Li J, Zhang K, Wen M, Guan Q (2020) An accurate ensemble forecasting approach for highly dynamic cloud workload with VMD and R-transformer. IEEE Access 8:115992–116003