

Research Article

LSTM with Wavelet Transform Based Data Preprocessing for Stock Price Prediction

Xiaodan Liang , Zhaodi Ge , Liling Sun , Maowei He , and Hanning Chen 

School of Computer Science and Technology, Tianjin Polytechnic University, Tianjin, 300387, China

Correspondence should be addressed to Liling Sun; sll198257@163.com

Received 2 May 2019; Accepted 19 June 2019; Published 16 July 2019

Academic Editor: Sylwester Samborski

Copyright © 2019 Xiaodan Liang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For profit maximization, the model-based stock price prediction can give valuable guidance to the investors. However, due to the existence of the high noise in financial data, it is inevitable that the deep neural networks trained by the original data fail to accurately predict the stock price. To address the problem, the wavelet threshold-denoising method, which has been widely applied in signal denoising, is adopted to preprocess the training data. The data preprocessing with the soft/hard threshold method can obviously restrain noise, and a new multi-optimal combination wavelet transform (MOCWT) method is proposed. In this method, a novel threshold-denoising function is presented to reduce the degree of distortion in signal reconstruction. The experimental results clearly showed that the proposed MOCWT outperforms the traditional methods in the term of prediction accuracy.

1. Introduction

Stock price prediction is a typical problem based on time series forecasting, and various stock forecasting methods emerge in an endless stream. Stock price prediction means to predict the stock price after a certain time, so as to help investors realize the maximum benefit. The methods proposed in the literatures can be roughly divided into two categories containing traditional mathematical methods and economic methods. Many previous works are based on traditional statistical methods [1]. Typically, Kalman filter and autoregressive model are very classic statistical methods, which are often used for financial sequence prediction. With the development of artificial intelligence, the deep learning methods are increasingly applied to the field of stock price prediction [2].

The deep learning method has a very superior performance compared with the traditional statistical method. One of the main reasons is that with the direct analysis, deep learning can map the original data to a nonlinear model thereby giving a better fitting effect through the multilayer neural network. In addition, deep learning has the advantage of self-selection in the application of the financial field. Most financial data is highly noisy and unstable. With deep learning, various events that have a significant impact on finance

can be expressed by knowledge maps. And then features are selected automatically through deep networks to adjust parameters and weights. The results obtained in this way may be more accurate and objective. Recurrent neural networks [3] (RNN) have been widely used in the field of natural language processing [4] and have achieved great success. With the Advantages of neural networks, it is possible to grasp the public sentiment more accurately. In this paper, the more powerful Long Short-Term Memory (LSTM) [5] neural network model based on the concept of RNN is used for stock prediction applications. It can save long-term memory more effectively. The “forget gate” and “input gate” are the core of its structure. With the “gate recurrent” structure [6], the model can remember the effective information while forgetting the useless information. In addition, due to the huge depth of the neural network in RNN, the gradient is calculated from the higher power of the matrix, sometimes accompanied by the problem of gradient explosion or gradient disappearance. However, the LSTM model with a special “gate recurrent” structure is very powerful to avoid this phenomenon.

In this paper, the wavelet denoising method is introduced into data preprocessing. The data with wavelet preprocessed were used as the training data. Our main contributions are as follows. First, we have improved the traditional wavelet denoising method and its performance is better than the

traditional wavelet denoising method. Secondly, we propose a new multi-optimal combination wavelet transform (MOCWT) method. Compared with the traditional wavelet method and the improved wavelet method, MOCWT has the best performance.

The remainder of this paper is organized as follows. Some related works are reviewed in Section 2. In Section 3, the LSTM model is utilized to handle the stock price forecasting task and a new method named multi-optimal combination wavelet transform (MOCWT) is proposed for the aim of data denoising. Section 4 demonstrates the performance of the proposed MOCWT with other canonical methods. Finally, the conclusion is given in Section 5.

2. Related Works

The issue of stock forecasting has been widely concerned by researchers. Various stock forecasting methods emerge in an endless stream. For example, some researchers use a news-oriented approach to predict stock trends.

Ziniu Hu et al. [7] proposed a Hybrid Attention Networks (HAN), which takes into account the use of analytical news factors to predict a stock trend in learning framework. Xi Zhang et al. [8] proposed that the investor social networks is also an emotion-oriented forecasting method, which utilizes emotional factors and stock correlation characteristics to model. In their further article [9], they considered the relationship between the single source and the multisource data and then employed the coupling matrix and tensor decomposition framework to study the impact of online news and user sentiment on stock price changes. In the research method of Huicheng Liu et al. [10], an Attention-Based RNN was used to accomplish this task and a Bidirectional-LSTM is used to capture the characteristics of the information in the news text. Based on factors such as online news and investor sentiment to prediction, Jieyun Huang et al. [11] designed tensors to capture the intrinsic connections of different sources of information and then used this method to solve the problem of data sparsity and finally proposed an improved submode coordinate algorithm (SMC) to match the use of tensors for improving the prediction accuracy. Marcelo Sardelich et al. [12] also studied the relationship between news and stock prices and then predicted the volatility of the day. Some researchers adopted mathematical methods to make stock market forecasts. In the literature of Mahsa Ghorbani et al. [13], the covariance information based on principal component analysis of dimensionality reduction method was designed for stock forecasting, and company stocks from different industries were used to illustrate this method.

Currently, the deep learning methods to predict stock prices have become the most widely used method. Yue-Gang Song et al. [14] compared the predictive performance of five neural network models for stock price forecasting. Hyeon Kyu Choi [15] proposed the ARIMA-LSTM hybrid model to predict the stock price correlation coefficient. This method first employed the ARIMA model to filter out the linear trend in the data and then passed the residual value to the LSTM. FULI FENG et al. [16] proposed a deep learning program for Relational Stock Ranking (RSR). The Temporal

Graph Convolution method was brought in to simulate the temporal evolution of the stock and the relational network, and the solution was implemented using the LSTM network. In their other paper [17], a confrontational training method was proposed to simulate the randomness of stock price by increasing the perturbation. Its advantage is to prevent overfitting of data. Linyi Yang et al. [18] proposed a dual-layer attention-based neural network with GRU, which adopted an input attention mechanism to reduce the noisy news and an output attention mechanism to allocate different weights to different days. Experiments showed that their method is effective.

In the case of stock forecasting, unlike the deep learning methods that have been widely used, there are a few researchers using convolutional neural network (CNN). CNN is originally used in image processing and has excellent performance. Ehsan Hoseinzade et al. [19] used CNN to extract the correlation of multisource data to achieve stock market forecasting. Jinho Lee et al. [20] used stock chart images as input to the model and used the Deep Q-Network and convolutional neural networks for global stock market forecasting. There are also some researchers who use other methods to study the stock market. Arjun R et al. [21] proposed a triangulation qualitative reasoning to explore the strength of causal relationship between stock price search interest and real stock market outcomes on worldwide equity market indices.

3. Model and Data

Stock price data is a typical time series data; in this section, the LSTM model is utilized to handle the stock price forecasting task. First, different structures of LSTM are utilized. Then, a new method named multi-optimal combination wavelet transform (MOCWT) is proposed for the aim of data denoising.

3.1. Model Introduction. LSTM network is a special RNN. Due to its unique structure, LSTM is suitable for handling and predicting problems with long intervals and delays in time series. LSTM is commonly used in autonomous speech recognition [22, 23] and natural language processing. As a nonlinear model, LSTM regarded as a complex nonlinear unit is to construct larger deep neural networks.

The concept of gates is brought in LSTM. Through “gates,” the transmission of information can be controlled in LSTM, which results in enabling the activation of long-term information. The simplest way of controlling the transmission of information is to multiply the corresponding points of two matrices with exactly the same size. All the points of multiplied matrix is in the range of $[0, 1]$. 0 means suppressing the information release, and 1 means activating the information release.

Figure 1 shows the hidden layer cell structure in LSTM.

Cell information transmission is controlled by forgetting gate. The passed information can be determined by both of the output of the hidden layer h_{t-1} and the input x_t of the current layer, which can also determine whether the information is remembered or forgotten. The activation

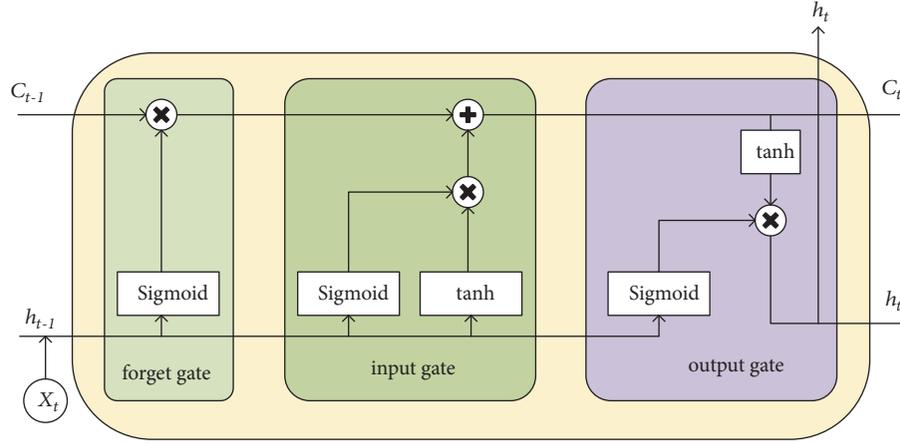


FIGURE 1: LSTM Neural Unit.

function is used to limit the output value in the range of $[0, 1]$, as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

The function of the input gate is to control which information among input information (h_{t-1}, x_t) should be added to the cell. The activation function used by the input information is to set the forgetting gate. Finally, multiplying both of functions is to achieve the purpose of retaining the corresponding information, and the formula is expressed as

$$i_t = \sigma(W_i [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

The output gate determines which information in the cell can be output. Similarly, the activation function used by the input information is to set the forgetting gate. And then the cell is activated by the tanh activation function. Finally, the dot multiplication determines which information should be output. The formula is expressed as

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = O_t \cdot \tanh(C_t) \quad (6)$$

In the RNN, the hidden layer unit usually only includes a single activation function. During the training process, the network is usually optimized by a backpropagation algorithm. Due to the multiplication mechanism of the reverse parameter optimization process, when the number of layers of the neural network is large, the gradient disappearance problems or the gradient explosion problems easily emerge. These problems can be well avoided by the structural mechanism of LSTM. In LSTM model, the structure of the hidden layer unit is more complicated, usually including multiple active layers, add operations, and multiplication operations.

In the experiment of this paper, LSTM with different number of layers were used to find the best training model. During the training process, data truncation length is set to 30, and the data distribution for training is shown in Figure 2.

3.2. Data Preprocessing. Wavelet transform is a time-frequency local analysis. The multiscale refinement of wavelet transform can be carried out by the stretching and translation of wavelet. When the frequency is high, the time is subdivided. When the frequency is low, the frequency is subdivided. Thereby the details of signal can be analyzed explicitly.

3.2.1. Introduction of Wavelet Transform. Let the signal length be n ; the noised signal be X is expressed:

$$X_n = f_n + e_n \quad (7)$$

where f is the real signal, and e is the noise signal.

Wavelet transform of a signal is its time domain and frequency domain transform. Useful information is extracted from the signal or noise information is removed; the basic wavelet is regarded as the simulation unit. To better preserve the original information, signals can be decomposed into a set of high frequency and low frequency by wavelet transform. The traditional high-pass or low-pass filter directly processes the original signal without decomposition, which may miss some usefulness of the signal information. Wavelet decomposes the original signal by operations such as stretching and translation of the basic wavelet. Then, a series of wavelet coefficients are obtained. Finally, through the low-pass or high-pass operation, the low-frequency information CA or the high-frequency information CD of the signal are obtained, respectively. Figure 3 shows the multilayer wavelet decomposition. The single layer decomposition was used in the subsequent experiments in this paper.

Since the effective signal generally is continuous in the time domain, the corresponding wavelet coefficients are generally large after the wavelet transform. While the noise signal generally is random and discontinuous in the time domain, accordingly, after the wavelet transform, the corresponding wavelet coefficients are relatively small. With a presetting threshold, the low-frequency wavelet coefficient and high-frequency wavelet coefficient are filtered. Then, the remaining part is inversely transformed by wavelet. Finally, the original

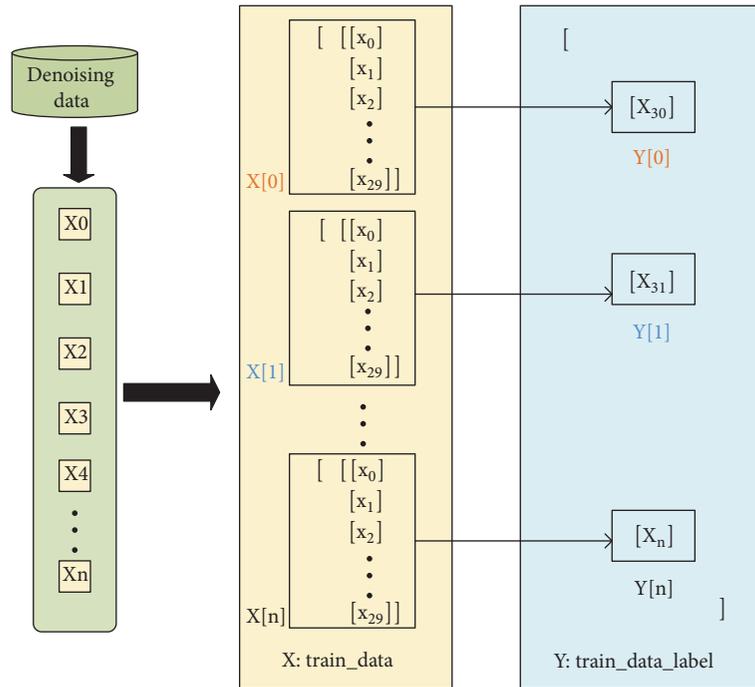


FIGURE 2: Model loading training data description.

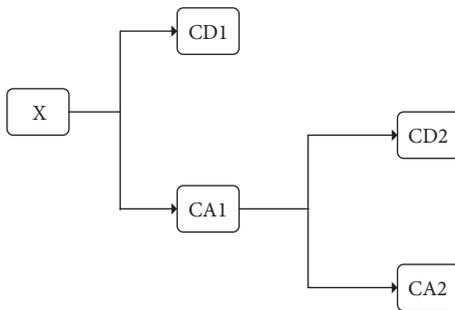


FIGURE 3: Multilayer wavelet decomposition.

signal is reconstructed. The wavelet noise reduction process is shown in Figure 4.

In the subsequent experiments, discrete wavelet transform DWT is used to discretize the power series of scale parameters, and then the time is discretized for analyzing signal.

3.2.2. Selection of Basic Wavelet. The basic wavelet in the subsequent experiments is Haar wavelet, which is a common basic wavelet. It has basic properties and meets the requirements of discrete wavelet transform. Its characteristics are as follows: Haar wavelet has the property of tight support; and discrete transformation can also be implemented; in addition, its support length is 1 and it is symmetric.

(1) One characteristics of the Haar wavelet is compacted support, with which the function has an outstanding sharp drop-off performance.

(2) Haar wavelet has a small support length, which conveniently shortens the computation time and obviously reduces the data processing time and the training time.

(3) The Haar wavelets are symmetric, which beneficially reduces the distortion rate during signal analysis and signal reconstruction. Therefore, the real price can be greatly restored after noise reduction.

3.2.3. Selection of Threshold. In order to reduce the fluctuations in stock prices, the threshold λ is indicated as follows:

$$\lambda = \varepsilon \cdot \sigma \tag{8}$$

where σ is the mean square error and ε is the control coefficient. σ is used as the threshold substrate for wavelet coefficient processing. The coefficients ε are introduced as the control coefficients for the threshold substrate σ . Control coefficient ε is controlled by loss generated after training; consequently, the threshold value λ is controlled globally. Finally, λ is used to improve the original wavelet transform method.

3.2.4. Selection of Threshold Functions. The threshold function mainly contains a soft threshold-denoising method and a hard threshold-denoising method. The soft threshold-denoising method is defined as follows: wavelet coefficient with its absolute less than the threshold is reset to zero; when the absolute value of the wavelet coefficient is greater than the threshold, the absolute value of the wavelet coefficient

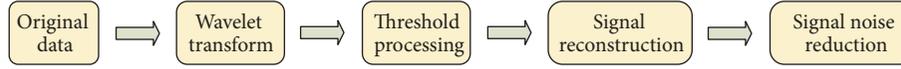


FIGURE 4: Wavelet noise reduction process.

is subtracted from the threshold value. Expressions are as follows:

$$\omega_\lambda = \begin{cases} \text{sign}(\omega) (|\omega| - \lambda) & |\omega| \geq \lambda \\ 0 & |\omega| < \lambda \end{cases} \quad (9)$$

The hard threshold-denoising method is defined as follows: wavelet coefficient with its absolute less than the threshold is reset to zero; inversely, wavelet coefficient retains. Expressions are as follows:

$$\omega_\lambda = \begin{cases} \omega & |\omega| \geq \lambda \\ 0 & |\omega| < \lambda \end{cases} \quad (10)$$

The soft threshold-denoising method is continuous. Although without additional oscillation, it fatally brings in a constant deviation. The hard threshold-denoising method with smaller mean square error inevitably introduces jump points and additional oscillation into the signal.

Both of the common methods above have their own drawbacks, respectively. In this paper, a new multi-optimal combination wavelet transform (MOCWT) method is proposed. Compared with the two traditional threshold functions, MOCWT has the advantage of continuity without jump phenomenon of hard threshold function, which can smooth the signal perfectly after noise reduction. Constant deviation will not be generated by MOCWT.

When the value of the signal is changing, especially, a produced difference can obviously reduce the degree of distortion in signal reconstruction. The expression is

$$\omega_\lambda = \begin{cases} \text{sign}(\omega) \left(|\omega| - \frac{\lambda}{e^{|\omega|/\lambda-1}} \right) & |\omega| \geq \lambda \\ 0 & |\omega| < \lambda \end{cases} \quad (11)$$

4. Results and Analysis

In this section, the experimental results comparing the proposed MOCWT with original wavelet method and improved wavelet method are given to demonstrate its outstanding performance.

The source data used in this experiment is the opening price of the S&P 500 index for nearly eighteen years. S&P 500 index is a comprehensive stock index that records 500 listed companies in the United States. Because it contains a lot of companies, the changes of the broader market can be greatly reflected. It also has the characteristics of wide sampling, strong representativeness, high precision, and good continuity.

In addition, the total error of training is employed as the evaluation criterion of experiment.

TABLE 1: Loss of different iterations.

Iterations	1	301	3301
Loss	0.010164	0.004343	0.000170

In Figure 5, the distribution of real and predicted values during training process is depicted. The smaller the error is, the closer the predicted curve is to the true curve. With the training process, the error gradually decreases. In Table 1, the static analysis of their loss is given. Final loss of the 3301 iteration less than 0.001 reveals the model is highly precision.

In the experiment, different modes of wavelet were used for data preprocessing. The preprocessed data was tested on LSTM models with 1, 2, 3, and 4 layers. With the increase of the number of layers in the model, its performance gradually deteriorated. The following five sets of experiment result show that the LSTM with two-layer structure has the best performance.

The experimental results of the original data without wavelet preprocessing, the data with original wavelet soft threshold preprocessing, and the data with original wavelet hard threshold preprocessing were compared. The experimental results are shown in Figure 6.

Figure 6 demonstrates the comparison results of the original data without wavelet preprocessing, the data with original wavelet soft threshold preprocessing, and the data with original wavelet hard threshold preprocessing.

The experimental results as shown in Figure 6 demonstrate that the data without wavelet processing is unsatisfactory. The performance of using the original wavelet hard threshold displays slightly better, and using the wavelet soft threshold occupies the most suitable performance.

The original wavelet threshold mode is improved, in which the threshold control function is introduced. The experimental result of the method without wavelet, original wavelet method, and improved wavelet method were compared and shown in Figures 7(a) and 7(b).

Experimental results reveal that the improved method is effective. The performance of the improved hard threshold processing method is improved obviously. The minimum loss reaches 2.453. And the improved soft threshold method performance is relatively weak. Its loss decreases by 5.163%.

The original wavelet soft threshold, the improved wavelet soft threshold, and the performance of MOCWT are compared. Figure 8(a) shows the experimental results.

The original wavelet hard threshold, the improved wavelet hard threshold, and the performance of MOCWT are also compared. The experimental results are shown in Figure 8(b).

The experimental results clearly show that the performance of MOCWT is the best. Compared with the original and improved wavelet method, MOCWT has an obviously improved performance. Although the processing loss of

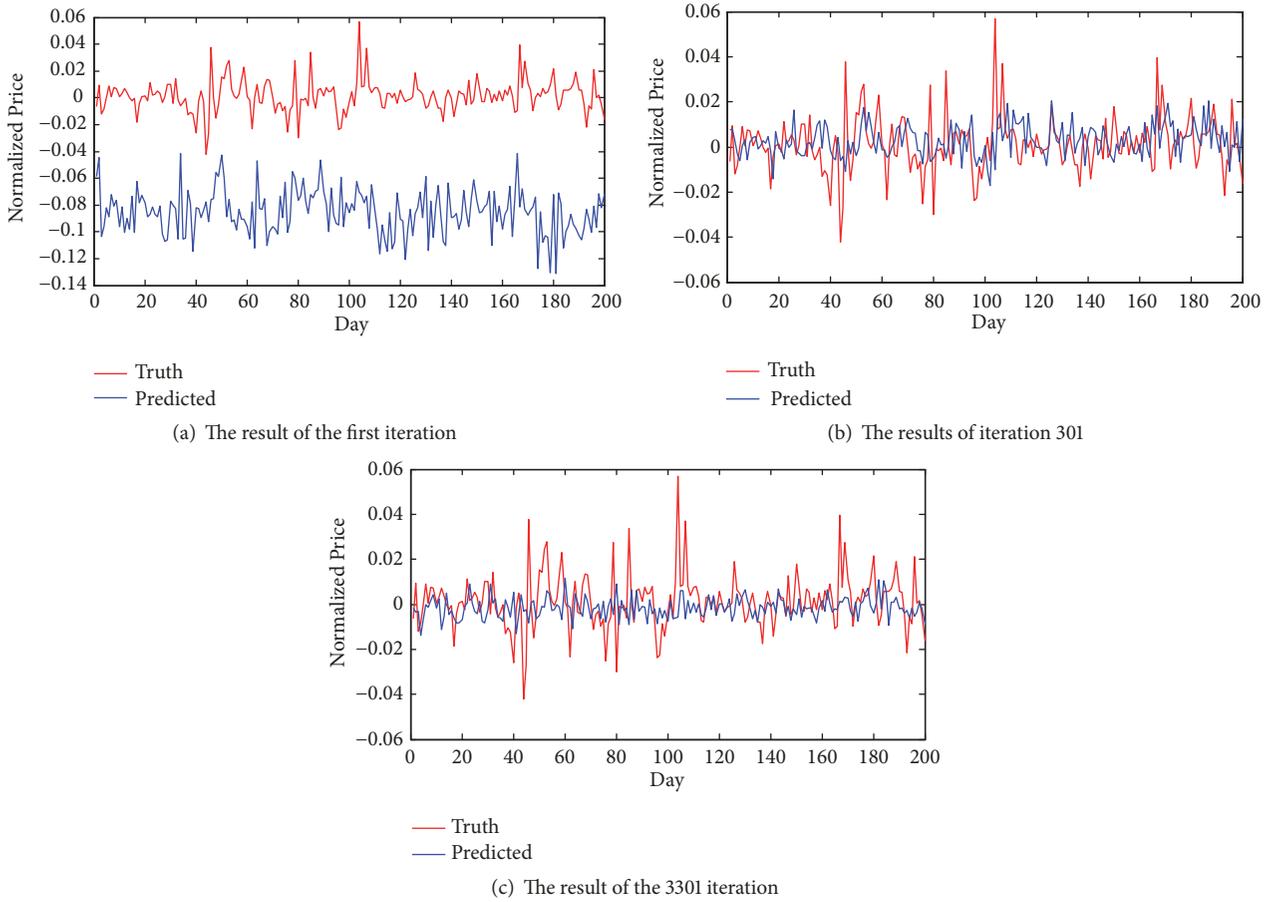


FIGURE 5: Distribution of real and predicted values.

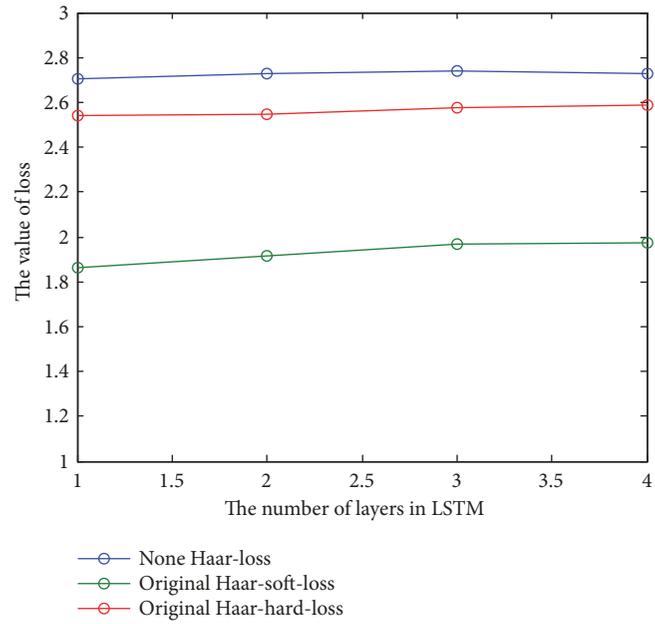
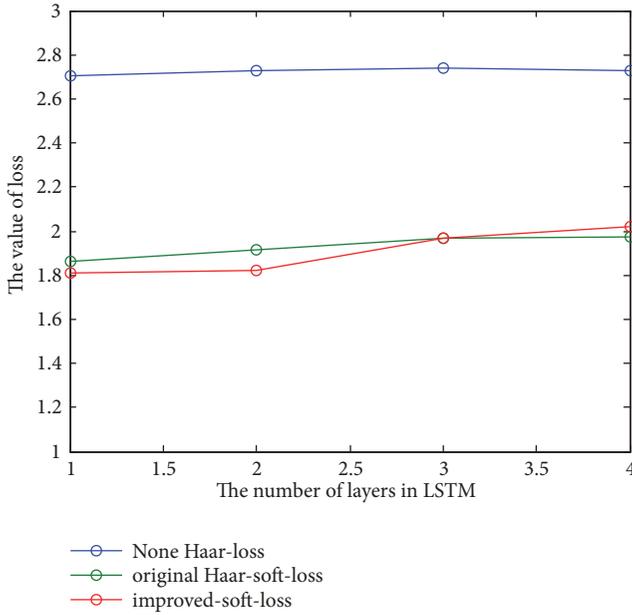
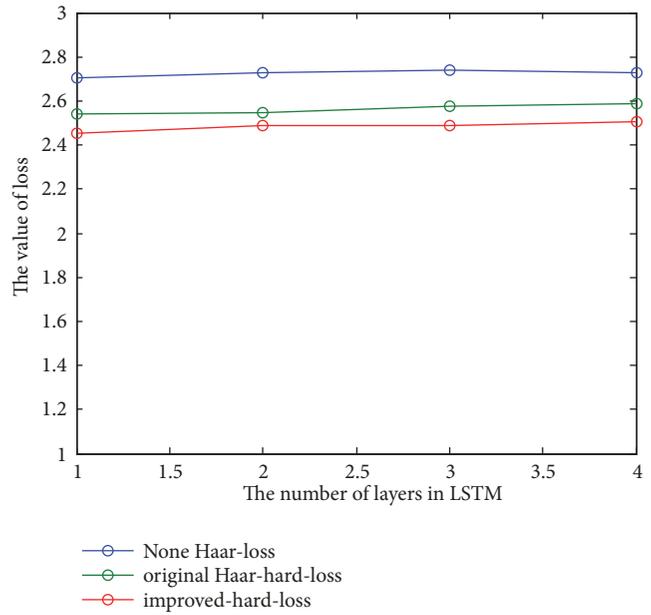


FIGURE 6: Comparison of wavelet transforms and unused wavelet transforms.

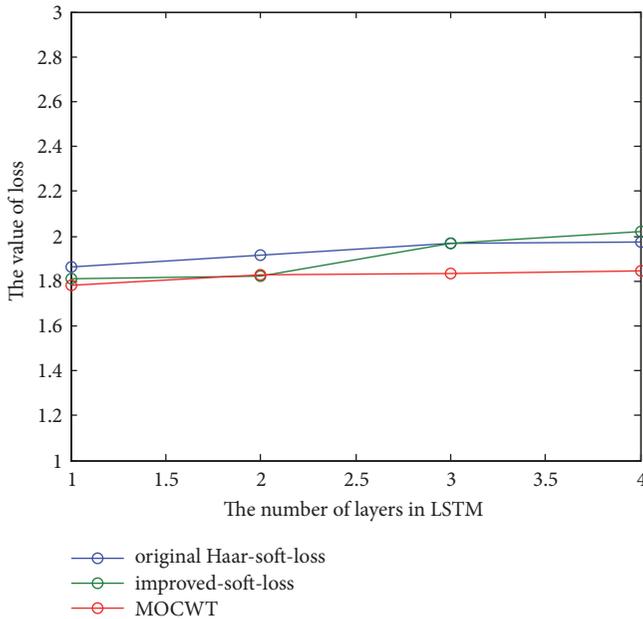


(a) Comparison of unused wavelets, original wavelet soft thresholds, and improved wavelet soft threshold results

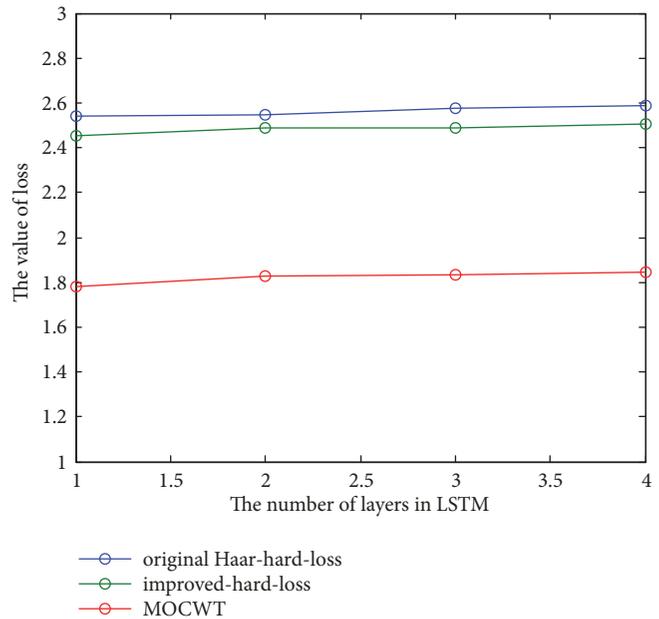


(b) Comparison of unused wavelets, original wavelet hard thresholds, and improved wavelet hard threshold results

FIGURE 7



(a) Comparison of original wavelet soft threshold, improved wavelet soft threshold, and MOCWT results



(b) Original wavelet hard threshold, improved wavelet hard threshold, MOCWT results comparison

FIGURE 8

MOCWT is nearly the same as that of the improved soft threshold when the number of LSTM layers is 2. Generally, the performance of MOCWT is outstanding compared with the traditional methods and the improved methods.

From above experimental results, we have successfully demonstrated that the MOCWT models are able to effectively improve the prediction accuracy.

5. Conclusion and Future Work

In this paper, we improved the original wavelet denoising method. The performance of improved method is better than traditional methods.

In addition, we propose a new multi-optimal combination wavelet transform (MOCWT) method, the experimental

results of which show that its performance is the best compared with the traditional wavelet method. The proposed MOCWT is obviously superior to the traditional method and the improved method. For the original data method without wavelet processing, the prediction results will always have a large oscillation, and the fitting effect of the real data is poor. The overall performance of the model is affected. The experimental results illustrate that the data characteristics are of great significance to the performance of the whole model.

In this work, there still are some key experimental points worth exploring in the future. For example, the optimization method and structure of the neural network and the loss function, as well as the parameter variables in the experiment, are all worth further optimizing.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the manuscript.

Acknowledgments

This work is supported by National key Research and Development Program of China under Grants nos. 2017YFB1103603 and 2017YFB1103003, National Natural Science Foundation of China under Grants nos. 61602343, 51607122, 61772365, 41772123, 61802280, 61806143, and 61502318, Tianjin Province Science and Technology Projects under Grants nos. 17JCYBJC15100 and 17JCQNJC04500, and Basic Scientific Research Business Funded Projects of Tianjin (2017KJ093, 2017KJ094).

References

- [1] F. Jassim and F. Altaani, "Statistical approach for predicting factors of mood method for object oriented," <https://arxiv.org/abs/1302.5454>.
- [2] N. Milosevic, "Equity forecast: predicting long term stock price movement using machine learning," <https://arxiv.org/abs/1603.00751>.
- [3] T. Mikolov, M. Karafiát, L. Burget, C. Jan, and S. Khudanpur, "Recurrent neural network based language model," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association: Spoken Language Processing for All, INTERSPEECH 2010*, pp. 1045–1048, Japan, September 2010.
- [4] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," <https://arxiv.org/abs/1708.05148>.
- [5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [6] G. Shen, Q. Tan, H. Zhang, P. Zeng, and J. Xu, "Deep learning with gated recurrent unit networks for financial sequence predictions," *Procedia Computer Science*, pp. 131895–131903, 2018.
- [7] Z. Hu, W. Liu, and J. Bian, "Listening to chaotic whispers: a deep learning framework for news-oriented stock trend prediction," <https://arxiv.org/abs/1712.02136>.
- [8] X. Zhang, J. Shi, D. Wang, and B. Fang, "Exploiting investors social network for stock prediction in China's market," *Journal of Computational Science*, vol. 28, pp. 294–303, 2018.
- [9] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, and P. S. Yu, "Improving stock market prediction via heterogeneous information fusion," *Knowledge-Based Systems*, vol. 143, pp. 236–247, 2018.
- [10] H. Liu, "Leveraging financial news for stock trend prediction with attention-based recurrent neural network," <https://arxiv.org/abs/1811.06173>.
- [11] J. Huang, Y. Zhang, J. Zhang, and X. Zhang, "A tensor-based sub-mode coordinate algorithm for stock prediction," <https://arxiv.org/abs/1805.07979>.
- [12] M. Sardelicha and S. Manandhar, "Multimodal deep learning for short-term stock volatility prediction," <https://arxiv.org/abs/1812.10479>.
- [13] M. Ghorbania and E. K. P. Chong, "Stock price prediction using principle components," <https://arxiv.org/abs/1803.05075>.
- [14] Y.-G. Songa, Y.-L. Zhou, and R.-J. Han, "Neural networks for stock price prediction," <https://arxiv.org/abs/1805.11317>.
- [15] H. K. Choi, "Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model," <https://arxiv.org/abs/1808.01560>.
- [16] F. Feng, X. He, X. Wang, C. Luo, Y. Liu, and T.-S. Chua, "Temporal relational ranking for stock prediction," <https://arxiv.org/abs/1809.09441>.
- [17] F. Feng, H. Chen, X. He, J. Ding, M. Sun, and T.-S. Chua, "Improving stock movement prediction with adversarial training," <https://arxiv.org/abs/1810.09936>.
- [18] L. Yang, Z. Zhang, S. Xiong et al., "Explainable text-driven neural network for stock prediction," <https://arxiv.org/abs/1902.04994>.
- [19] E. Hoseinzade and S. Haratizadeh, "CNNPred: CNN-based stock market prediction using several data sources," <https://arxiv.org/abs/1810.08923>.
- [20] J. Lee, R. Kim, Y. Koh, and J. Kang, "Global stock market prediction based on stock chart images using deep Q-network," <https://arxiv.org/abs/1902.10948>.
- [21] R. Arjun and K. R. Suprabha, "Predictive modeling of stock indices closing from web search trends," <https://arxiv.org/abs/1804.01676>.
- [22] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *Proceedings of the 31st International Conference on International (ICML '14)*, vol. 14, pp. 1764–1772, Springer, 2014.
- [23] A. Hannun, C. Case, J. Casper et al., "Deep speech: Scaling up end-to-end speech recognition," <https://arxiv.org/abs/1412.5567>.

