Network Traffic Prediction in Network Security Based on EMD and LSTM

Wei Zhao, Huifeng Yang, Jingquan Li, Li Shang, Lizhang Hu, and Qiang Fu

Abstract With the rapid development of the Internet, the scale of the network continues to expand, and the situation of network security is getting more and more severe. Network security requires more reliable information to support, and the prediction of network traffic is an important part of network security. Network traffic prediction data can provide important data reference for network security, especially for reliable data transmission and network monitoring. In fact, network traffic data is affected by a variety of complex and random factors, so network traffic data is a nonlinear data sequence. This paper analyzes the characteristics of network traffic data and proposes an EMD-LSTM model for network traffic data prediction. Firstly, the complex and variable network data traffic is decomposed into several smooth data sequences, and then, the LSTM neural network model, which is suitable for data sequence prediction, is used to predict. The results of the comparison experiments show that the proposed network traffic prediction method reduces the prediction root mean square error in network traffic prediction.

Keywords Network traffic · Prediction · LSTM · EMD

1 Introduction

With the gradual development of the modern Internet, the network scale has been expanding, and the network data is becoming larger and larger. Network security has become a major issue in current network development, and network traffic provides data support for network security. Network traffic data plays a vital role in network monitoring, network attack and defense, etc. For example, network traffic prediction can help monitors quickly discover abnormal traffic data and prevent network attacks. How to more accurately predict future network traffic data becomes an important issue in network security.

W. Zhao (✉) · H. Yang · J. Li · L. Shang · L. Hu · Q. Fu
State Grid Hebei Electric Power Co., Ltd., Information and Telecommunication Company, Shijiazhuang, China
e-mail: 934501954@qq.com

Q. Liu et al. (eds.), *Proceedings of the 9th International Conference on Computer Engineering and Networks*, Advances in Intelligent Systems and Computing 1143, https://doi.org/10.1007/978-981-15-3753-0_50

Whether it is based on traditional mathematical models or network traffic prediction algorithms based on general neural network models, there are many limitations. The essence of network traffic data is a time-based data sequence, but this data sequence is affected by various uncertain factors, such as holidays and major emergencies, which are difficult to express in the network traffic data by mathematical models.

Aiming at the problems existing in the existing network traffic prediction model and the analysis of the nature of network traffic data, we propose an LSTM neural network model combined with EMD data processing to predict network traffic data. The EMD decomposition is used to decompose the complex and variable nonlinear network traffic data into a smoother IMF sequence, which effectively reduces the complexity of the data sequence and reduces the difficulty for subsequent prediction. Secondly, the LSTM neural network model is used to predict the decomposition.

Compared with other neural network models, LSTM neural network has a better expression of nonlinear sequences and has memory. It also has a good effect on time-based data sequence prediction and can solve the problem of the gradient disappears when inputting long time series. This paper also uses the Adam algorithm to update the gradient during training, which speeds up the convergence and is not trapped at the saddle point.

The remaining chapters of this paper are organized as follows: Chapter 2 introduces relevant domestic and foreign research work, Chap. 3 introduces the EMD-LSTM model, Chap. 4 carries out simulation experiments, and Chap. 5 summarizes the work of this paper.

2 Related Works

In recent years, network traffic prediction methods mainly include regression model, time series model, grey prediction method, neural network, fuzzy theory, mean-value method, wavelet theory, and statistical learning theory method [1].

According to the characteristics of short-term traffic data of LAN, Lin uses the ARMA method commonly used in time series analysis, establishes the time prediction model of network traffic based on ARMA model, and determines the prediction parameters of ARMA model based on short-term traffic prediction requirements [2]. This method changes the previous network management response method, making early warning of network overload and prediction of network traffic possible.

A large number of studies have found that the use of a linear method to predict nonlinear network traffic is theoretically inadequate, and its prediction accuracy is not high. Researchers have demonstrated that neural network systems with nonlinear structures can approximate arbitrary nonlinear functions [3]. However, the network traffic prediction algorithm based on the general neural network model such as the BP neural network model has a slow convergence rate and the over-fitting problems when there are more layers [4–6].

Aiming at the problems existing in the existing network traffic prediction model and the analysis of the nature of network traffic data, we propose an LSTM neural network model combined with EMD data processing to predict network traffic data.

3 Algorithm

The flow of the network traffic prediction model in this paper is designed as follows:

1. Collect and organize raw network traffic data.
2. Use the empirical mode to decompose the collated network traffic data so that the network traffic sequence is decomposed into 6–7 more stationary IMF sequences.
3. Convert the decomposed IMF sequence into a supervised learning column
4. Input the supervised learning sequence into the long and short memory neural network model and output the predicted IMF sequence value.
5. Calculate the predicted network traffic data by adding the values of the predicted IMF sequences.

This paper combines EMD and LSTM, proposes an EMD-LSTM hybrid model for network traffic prediction, and uses Adam optimization algorithm for gradient descent (Fig. 1).

3.1 Empirical Mode Decomposition

EMD is a new adaptive signal time–frequency processing method proposed by Huang et al. in 1998, which is very effective for analyzing nonlinear non-stationary signals.

EMD does not require a pre-set basis function for signal decomposition because it is based on the time-scale characteristics of the data itself. Any complex data set can be decomposed into a finite intrinsic mode function (IMF) by EMD. The intrinsic mode function is defined as a function that satisfies the following requirements:

(1) The number of extreme points of an IMF must be equal to the number of zero crossings, or the number of the two is only one difference.
(2) At all points in time, the average of the upper envelope defined by the local maximum and the lower envelope defined by the local minimum is zero.

The basic principle of empirical mode decomposition:

(3) Initialize the original time series: $r_0 = x(t), i = 1$
(4) Obtain the ith IMF
(5) Subtract the new arrival IMF component from the original sequence:

$$r_i(t) = r_{i-1}(t) - \mathrm{imf}_i(t) \tag{3.1}$$

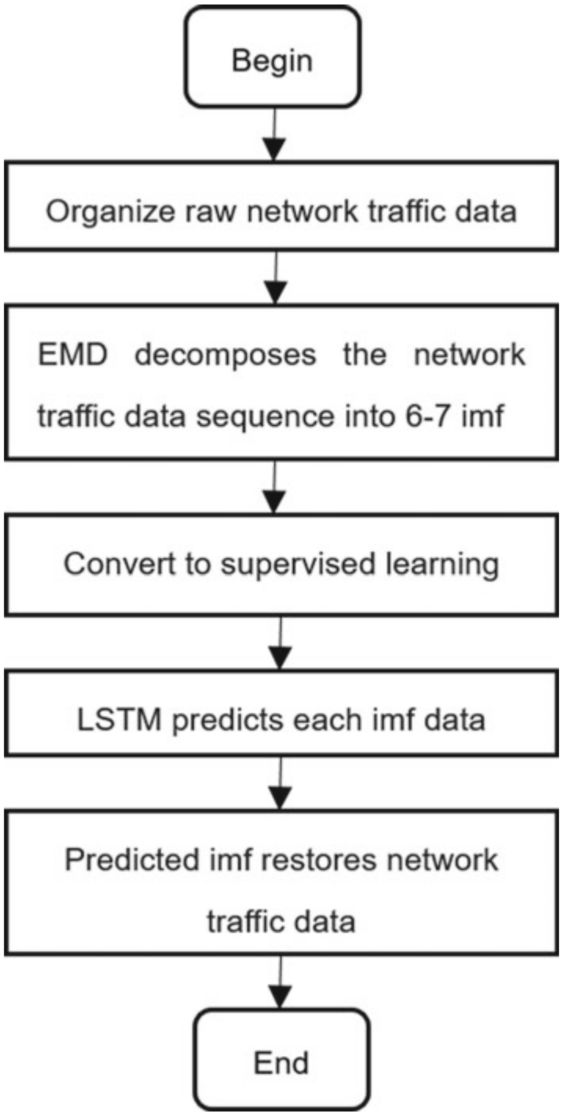


Fig. 1 Network traffic prediction model flowchart

(6) If the number of extreme points of $r_i(t)$ is still more than two, calculate $i = i+1$, and go to step 2; otherwise, the decomposition ends and $r_i(t)$ are the residual component.
(7) The algorithm is finally available:

$$x(t) = \sum_{i=1}^{n} \mathrm{imf}_i(t) + r_n(t) \tag{3.2}$$

Figure 2 is a graph of all the corresponding data. The signal column indicates the original data, and the IMF and res. are the decomposed data. Adding the IMF data of all the columns and the res. data can obtain the original data signal.

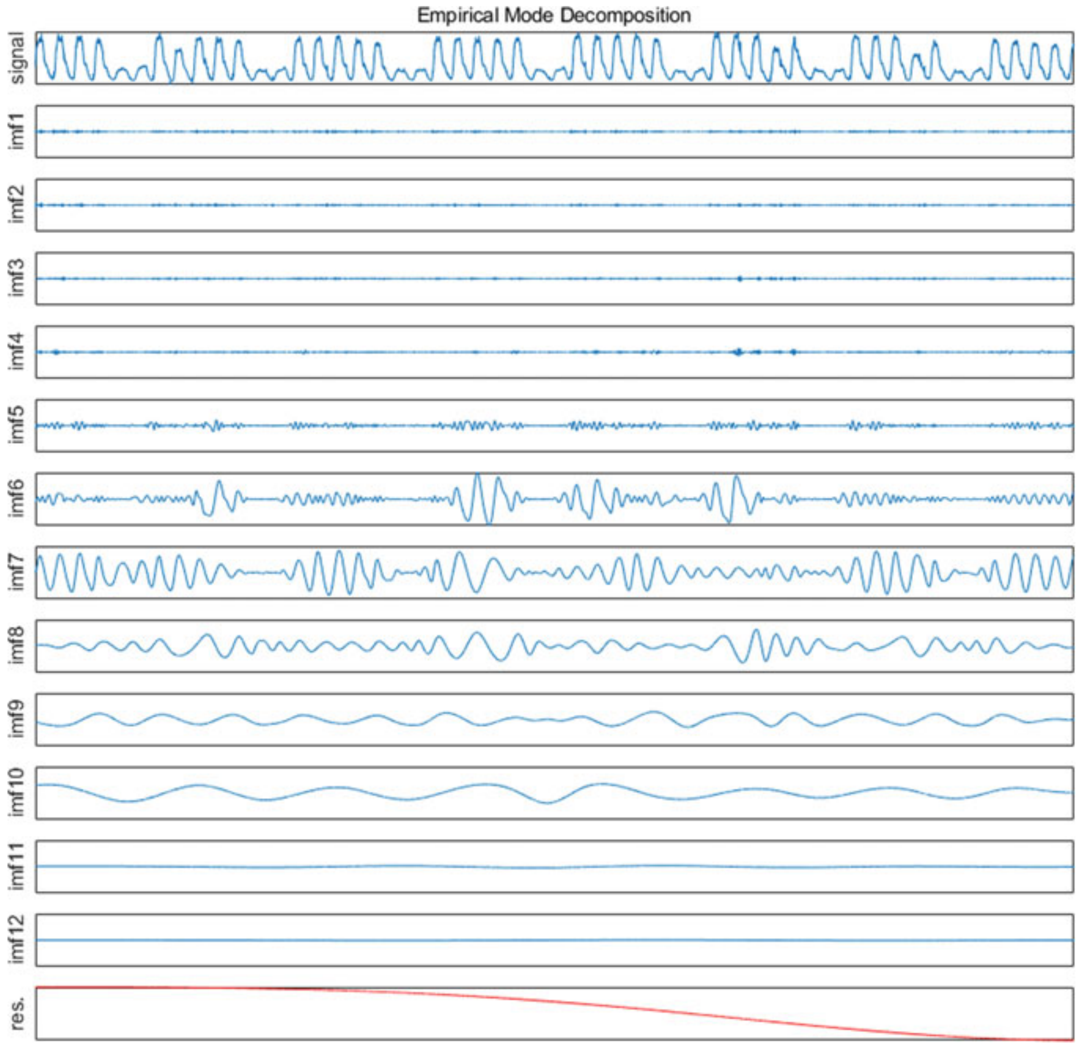


Fig. 2 EMD decomposition signal sequence

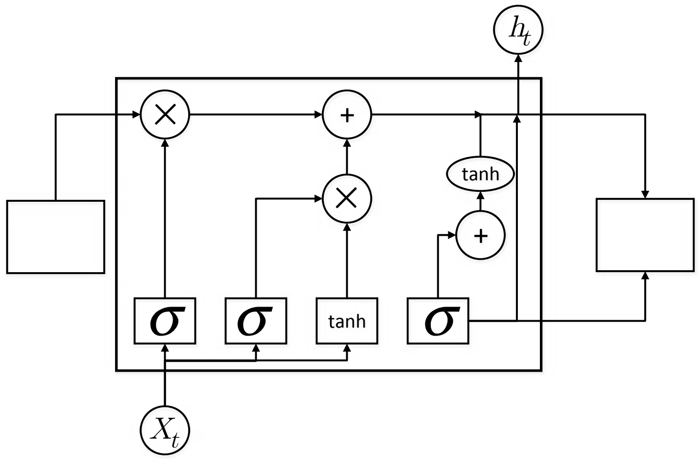
3.2 LSTM

Long short-term memory networks (LSTM) is a special kind of recurrent neural network (RNN). It was first proposed by Hochreiter & Schmidhuber in "LONG SHORT-TERM MEMORY" and improved by many people in the research and work. At present, LSTM has performed very well in natural language translation, computer vision, speech recognition, and trend prediction. It has been widely used in recent years.

The neuronal structure of LSTM includes input gates, output gates, forgetting gates, and cell units (Fig. 3).

The sigmoid activation function is used to control the information of the three gates, and the tanh activation function is used to update the candidate cell state and determine the output information. The sigmoid and tanh and their derivatives are as follows:

Fig. 3 Structure of LSTM neuron



$$\sigma z = y = \frac{1}{1 + \mathrm{e}^{-z}} \tag{3.3}$$

$$\sigma'(z) = y(1 - y) \tag{3.4}$$

$$\tanh(z) = y = \frac{\mathrm{e}^{z} - \mathrm{e}^{-z}}{\mathrm{e}^{z} + \mathrm{e}^{-z}} \tag{3.5}$$

$$\tanh'(z) = 1 - y^2 \tag{3.6}$$

The forgotten gate mainly determines the output value h_{t-1} at one moment and the input variable x_t at this moment. What information needs to be forgotten:

$$f_t = \mathrm{sigmoid}(W_{hf} * h_{t-1} + W_{xf} * x_t + b_f) \tag{3.7}$$

The input gate mainly determines the value to be updated:

$$i_t = \mathrm{sigmoid}(W_{hi} * h_{t-1} + W_{xi} * x_t + b_i) \tag{3.8}$$

Update candidate cell unit status:

$$\tilde{C}_t = \tanh(W_{hc} * h_{t-1} + W_{xc} * x_t + b_c) \tag{3.9}$$

Update cell unit status:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3.10}$$

The output gate determines which part of the information can be outputted:

$$o_t = \mathrm{sigmoid}(W_{ho} * h_{t-1} + W_{xo} * x_t + b_o) \tag{3.11}$$

Final output value:

$$h_t = o_t * \tanh(C_t) \tag{3.12}$$

3.3 Adagrad Optimization Algorithm for Gradient Descent

A common method for updating the gradient parameters is to use the stochastic gradient descent algorithm SGD. But SGD has the problem of difficulty in selecting the learning rate and unbalanced learning rate, this paper adopts Adagrad to update the gradient. The algorithm has the following advantages:

(1) Good at dealing with problems involving higher noise and sparse gradients.
(2) Ability to adjust different adaptive learning rates for different parameters.

The Adagrad algorithm imposes a constraint on the learning rate. During each iteration, each parameter is optimized using a different learning rate.

$$n_t = n_{t-1} + g_t^2 \tag{3.13}$$

$$\Delta\theta_t = -\frac{\eta}{\sqrt{n_t + \epsilon}} * \varphi \tag{3.14}$$

g_t is the weight gradient, n_t is the second moment estimate of the weight gradient, $\Delta\theta_t$ is the amount of weight reduction, φ is the learning rate, and ϵ is used to ensure that the denominator is not 0.

4 Experiment

4.1 Simulation Implementation

The EMD-LSTM network traffic prediction model designed in this paper uses the EMD algorithm for data preprocessing, predicts through the LSTM-based neural network model, adds the perceptual layer and dropout layer to the neural network model, and uses Adagrad algorithm in gradient descent.

In this paper, the prediction effects of the four models are simulated. The first one is the model designed in this paper, and the other three are the models when a certain condition is missing

(1) Data preprocessing is performed using EMD, the LSTM with dense and dropout layer is used for prediction, and gradient degradation is performed using the Adagrad algorithm.

(2) Data preprocessing is performed using EMD, the LSTM with dense and dropout layer is used for prediction, and gradient reduction is performed using the SGD algorithm.
(3) Without decomposing the data, the LSTM with dense and dropout layer is used for prediction, and the Adagrad algorithm is used for gradient descent.
(4) Data preprocessing was performed using EMD decomposition, the only LSTM is used for prediction, and gradient degradation was performed using the Adagrad algorithm.

4.2 Performance Comparison

The data set simulated in this article is Internet traffic data (in bits) from a private ISP with centers in 11 European cities which downloaded from datamarket.com. In order to evaluate the performance of the network traffic prediction model, root mean squared error (RMSE) was selected as the evaluation index in this simulation experiment. The RMSE value reflects the extent to which the predicted data deviates from the true value. The smaller the value of RMSE, the smaller the deviation between the predicted value of a neural network and the true value, and the better the performance.

In the simulation, the first 12,000 of the 14,776 network traffic sequence data is used as the training set, and the last 2776 data is used as the test set.

Figure 4 shows the loss of the training set of the four models. Figure 5 shows the test set prediction values of the four models and the original values. From the comparison results in Fig. 4 and Fig. 5, the first model, which is the EMD-LSTM model designed in this paper, has the best prediction effect.

Table 1 shows the RMSE of the training set and test set prediction values of the four models. From the comparison results in Table 1, the first model, which is the EMD-LSTM model designed in this paper, has the best prediction effect.

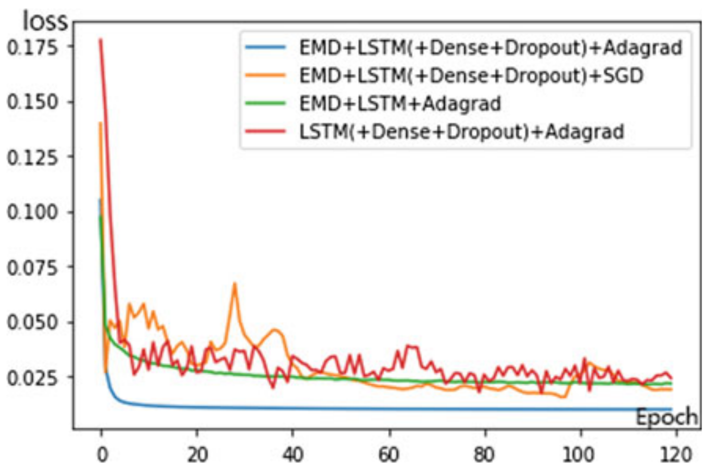


Fig. 4 Loss of each model

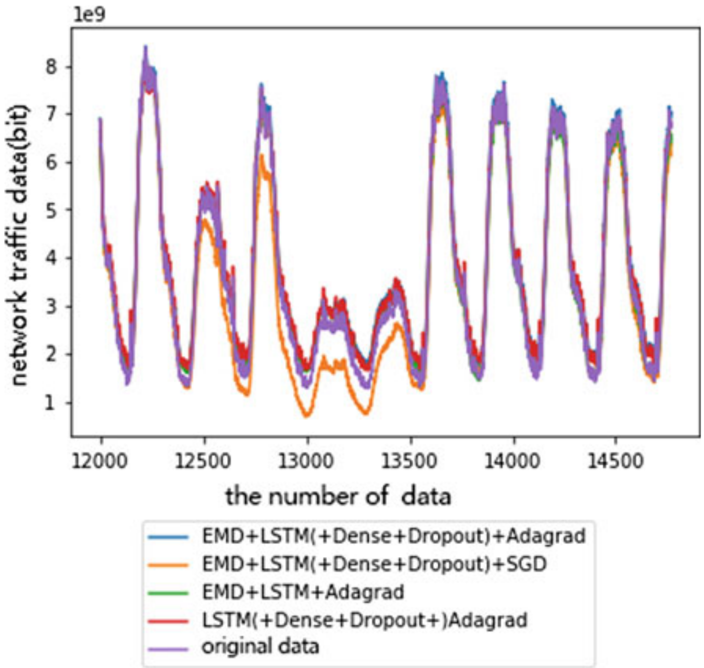


Fig. 5 Prediction data and the origin values

Table 1 Performance comparison of network traffic prediction

	Train RMSE	Test RMSE:
(1) EMD + LSTM (+Dense + Dropout) + Adagrad	112,290,084.077	103,795,012.873
(2) EMD + LSTM (+Dense + Dropout) + SGD	264,396,199.224	261,092,042.262
(3) EMD + LSTM + Adagrad	127,929,176.098	125,715,469.326
(4) LSTM (+Dense + Dropout) + Adagrad	145,828,276.612	132,289,805.802

Compared with the model using the SGD algorithm, the RMSE of the training set is reduced by 57.53%, and the root mean square error of the test set is reduced by 60.24%. The RMSE on the training set was reduced by 23.00% compared to the model without EMD and was reduced by 21.54% on the test set. The LSTM model in this paper adds the perceptual layer and the dropout layer. The RMSE of the training set is reduced by 12.22%, and the RMSE of the test set is reduced by 17.44%. It can be seen that the Adagrad gradient descent algorithm is more suitable for the data set of this paper. The data preprocessing using EMD effectively reduces the prediction error, and the addition of the dense layer and the dropout layer also achieve certain optimization effects. Analysis of performance comparison data can lead to the following conclusions

(1) The Adagrad algorithm is more suitable for training the EMD-LSTM model of this paper than the SGD algorithm.
(2) Using the EMD, preprocess data effectively reduces the prediction error.
(3) Adding the dense layer and the dropout layer to the LSTM model is more effective than using a single LSTM model directly, but the improvement effect is not obvious, which may be caused by fewer data in this paper.

5 Conclusions

In order to reduce the prediction error of the network traffic prediction model in current network security, this paper proposes a method to predict network traffic by combining EMD and LSTM neural network. We use the LSTM model which is good at processing time series because the time series of network traffic data is nonlinear. Considering that the change of network traffic data contains many influencing factors that are difficult to collect and express, we add the EMD decomposition to decompose a complex single network data sequence into multiple smoother sequence data as input in the data preprocessing part. In the simulation experiment, we use the Adagrad algorithm to carry out the gradient descent and achieve good results. The experimental result shows that the EMD-LSTM prediction method reduces the prediction error and can effectively predict the network traffic in the network traffic prediction task.

References

1. Wang, X.S.: Network traffic predicting model based on improved support vector machine. Comput. Syst. Appl. **26**(3), 230–233 (2017) (In Chinese)
2. Lin, Y.Y., Ju, F.C., Lu, Y.: Research on application of ARMA model in short in short-term traffic prediction in LAN. Comput. Eng. Appl. **53**(S2), 88–91 (2017) (In Chinese)
3. Hornik, K., Stinchcombe, M., White, H.: Multilayer feed forward networks are universal approximators. Neural Netw. **2**(5), 359–366 (1989)
4. Chen, Y., Yang, B., Meng, Q.: Small-time scale network traffic prediction based on flexible neural tree. Appl. Soft Comput. **12**(1), 274–279 (2012)
5. Miguel, M.L.F., Penna, M.C., Nievola, J.C., et al.: New models for long-term internet traffic forecasting using artificial neural networks and flow based information. In: Network Operations and Management Symposium, pp. 1082–1088 (2012)
6. Park, D.C.: Structure optimization of BiLinear recurrent neural networks and its application to ethernet network traffic prediction. Inf. Sci. **237**(13), 18–28 (2013)