NTAM-LSTM models of network traffic prediction

Jihong Zhao[1,2], and *Xiaoyuan* He[1,*]

[1]School of Communication and Information Engineering, XI'an University of Post & Telecommunications, China
[2]School of Electronic and Information Engineering, Xi'an Jiaotong University, China

Abstract. Accurate prediction of network traffic is very important in allocating network resources. With the rapid development of network technology, network traffic becomes more complex and diverse. The traditional network traffic prediction model cannot accurately predict the current network traffic within the effective time. This paper proposes a Network Traffic Prediction Model----NTAM-LSTM, which based on Attention Mechanism with Long and Short Time Memory. Firstly, the model preprocesses the historical dataset of network traffic with multiple characteristics. Then the LSTM network is used to make initial prediction for the processed dataset. Finally, attention mechanism is introduced to get more accurate prediction results. Compared with other network traffic prediction models, NTAM-LSTM prediction model can achieve higher prediction accuracy and take shorter running time.

Keywords: Network traffic prediction, LSTM neural network, Attention mechanism

1 Introduction

With the development of communication technology, the number of network users and services increase rapidly, and network traffic becomes complex and diverse. Accurate prediction of network traffic is helpful to the actual analysis of network performance and the efficient allocation of network resources [1]. By optimizing the allocation of network resources, the network user's experience can be improved. Therefore, the prediction of network traffic [2] plays a very important role in the development planning of modern network.

At present, the prediction methods of network traffic are mainly divided into linear and nonlinear time series prediction. The traditional linear time series prediction models include Autoregressive Moving Average. (ARMA) model [3] and Autoregressive Integrated Moving Average (ARIMA) model [4], and such models will cause plenty prediction errors in the face of network traffic with sudden, nonlinear, multi-feature and other characteristics. However, the nonlinear time series prediction model has a strong function, and it also has a better performance than the linear prediction model, so it has become a research trend to

---

* Corresponding author: miayuanxiaohe@163.com

apply the neural network in the nonlinear prediction model to the network traffic prediction in recent years [5,6].

Ming Lei [7] et al. proposed a traffic prediction model based on wavelet neural network, and the prediction result of this model is stable. However, when the interval between the required traffic information and the current forecast state is too large, the problem of gradient disappearance will occur. Thus the model is only applicable to short-term sequence prediction. Shyam Nihale [8] et al. proposed a network traffic prediction algorithm based on the long-term and short-term memory neural network, which was used to predict the real network traffic trajectory. Vinayakumar R [9] et al used the Gate Recurrent Unit (GRU) model to predict network traffic. Jing Bi [10] et al. applied the Temporal Convolutional Network (TCN) to Network traffic prediction. These three prediction models can improve the prediction accuracy of network traffic to a certain extent, but they do not distinguish the influences of different characteristic values of network traffic, and the prediction accuracy of actual network traffic is insufficient.

In order to improve the accuracy of network traffic prediction and reduce the running time, this paper proposes a network traffic prediction method based on NTAM-LSTM. On the basis of the traditional LSTM prediction model [11], attention mechanism [12] is introduced for improvement, which solves the problem that the traditional LSTM prediction model sets up the same length for different characteristic values while ignoring their different influences on target values.

The organizational structure of the paper is as follows:

In the second section, the model architecture of network traffic prediction system is described, the LSTM model structure and attention mechanism principle are respectively introduced, and the NTAM-LSTM network traffic prediction algorithm is described in detail. In the third section, the proposed model is verified experimentally, and the experimental results are given and evaluated. The fourth section is the summary of the full text and the future research direction.

2 Network traffic prediction model

2.1 NTAM-LSTM system model architecture

Figure 1 shows the system model architecture of NTAM-LSTM. It consists of four layers: data processing layer, LSTM layer, attention mechanism layer and data output layer.
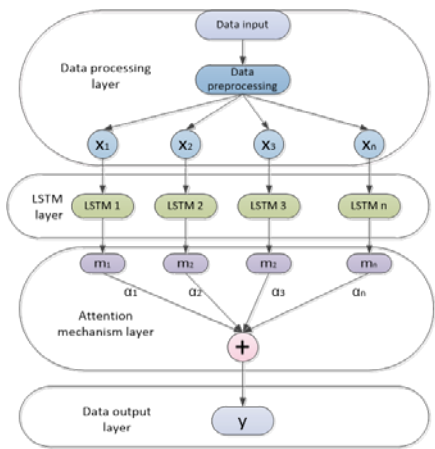

Fig. 1. Network traffic prediction system model architecture.

Data processing layer: This layer preprocesses the network traffic data input to the model, including deleting invalid data and normalizing the traffic data. Then, it converts the network traffic dataset into the suitable dataset $X = \{x_1, x_2, \cdots, x_n\}$ for neural network input.

LSTM layer: The result X of pretreated network traffic data is used as input for this layer. Initial prediction is made by using LSTM neural network's unique gate structure. Meanwhile, the Dropout method is used to prevent overfitting and improve model generalization.

Attention mechanism layer: Through the LSTM network layer, the status value of the network traffic at this moment can be obtained. The different characteristic status values of the network traffic at this time can be regarded as a set $M' = (m_1^t, m_2^t, \cdots, m_n^t)$. The network traffic set M' is taken as the input of the attention mechanism layer, and the corresponding weight α_i between the characteristic values of different traffic is calculated, and the weighted sum of them is carried out.

Data output layer: This layer outputs the final forecast results to complete the prediction of network traffic.

2.2 LSTM Network layer description

The LSTM network layer is used to make initial predictions of preprocessed network traffic dataset. LSTM neural network is improved through Recurrent neural network [13], which solves the problem of gradient disappearance in RNN model and has better applicability for predicting time series.

The LSTM network layer follows the basic standard structure of the traditional neural network, and its structural model is shown in Figure 2, including the input layer, the hidden layer and the output layer.
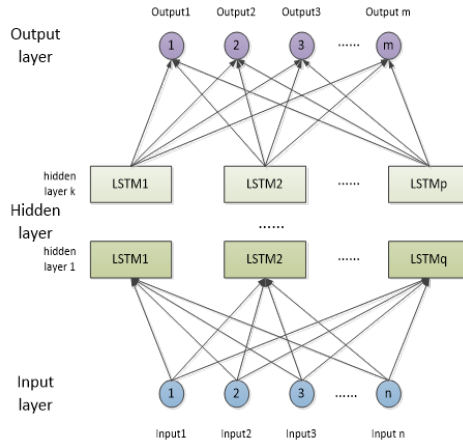

Fig. 2. LSTM network structure model.

The LSTM hidden layer contains three gate structures: Forget gate F, Input gate I and Output gate O. The schematic diagram of the internal department structure in LSTM is shown in Figure 3, where c_{t-1} and c_t represent the previous flow carrier and the current flow carrier of the LSTM unit, m_{t-1} and m_t represent the previous flow state and the current flow state, x_t and y_t represent the input and output of the current flow state. Its activation function adopts sigmoid function.
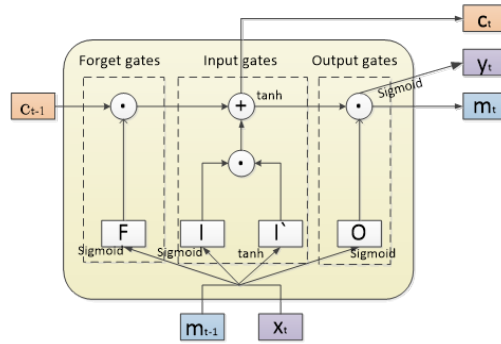
Fig. 3. Department structure within LSTM.

Forget gates F: Used to determine whether information needs to be saved or discarded from neurons. The process can be expressed as:

$$F = \sigma(W_F \bullet [m_{t-1}, x_t] + b_F) \tag{1}$$

where σ represents the sigmoid function, W_F represents the weight matrix of F, b_F represents the corresponding offset.

Input gates I: used to update the state of information and determine which new information needs to be stored in the state of the neuron. Two input steps are required to implement this process. The process can be expressed as:

$$I = \sigma(W_I \bullet [m_{t-1}, x_t] + b_I) \tag{2}$$

$$I' = \tanh(W_C \bullet [m_{t-1}, x_t] + b_C) \tag{3}$$

where W_I and W_C represent the weight matrix of I and I', b_I and b_C represent the corresponding bias.

After the information state to be updated is determined, the two steps are combined to update the state of the neuron with the information obtained in the previous two steps. The process can be expressed as:

$$c_t = F \bullet c_{t-1} + I \bullet I' \tag{4}$$

Output gates O: Used to update the neuron state of the system and determine the value of the output next hidden state. The process can be expressed as:

$$O = \sigma(W_O \bullet [m_{t-1}, x_t] + b_O) \tag{5}$$

$$m_t = O \bullet \tanh(c_t) \tag{6}$$

$$y_t = sigmoid(W_t \bullet m_t) \tag{7}$$

In the formula, W_O represents the weight matrix of O, b_O represents the offset of O, y_t represents the output of this time.

To sum up, the LSTM layer obtains the output result $M' = (m_1^t, m_2^t, \cdots, m_3^t)$ by passing the input historical traffic dataset X.

2.3 Attention mechanism layer description

When the input sequence of the LSTM network model is a multi-dimension time series with different characteristic values, the influence of different characteristic values on the target value cannot be distinguished, which will cause certain prediction errors in the training process. Therefore, in this paper, attention mechanism is introduced into the traditional LSTM network model, which can determine the effective information in more accurately characteristic value of network traffic, and allocate the corresponding weight ratio to it, there by optimizing the LSTM model. The calculation process of weight distribution value of attention mechanism layer is shown in Figure 4.
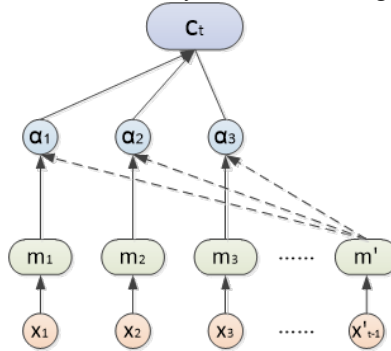

Fig. 4. Structure of attention mechanism.

The purpose of using the attention mechanism is to calculate the probability distribution of the input attention assignment to the target value. With the training network, the hidden layer node state m' randomly generated at a certain moment is compared one by one with LSTM hidden layer node state m_i corresponding to each input value x_i, and the similarity calculation function M is used to determine which hidden node state is most similar to m_i. At the same time, soft Max normalization is performed. M can be expressed as:

$$M = \tanh(\omega_1 m' + \omega_2 m_i) \qquad (8)$$

When the similarity coefficient corresponding to other network traffic characteristic values is obtained, the probability distribution of attention distribution is obtained, attention weight α_i:

$$\alpha = soft\max(\omega M) = \frac{\exp(M)}{\sum_i \exp(M)} \qquad (9)$$

The output adds up all eigenvalue states by α_i weighted sum, combined with m' to get the final result.

$$c_t = \sum_i \alpha_i m_i \qquad (10)$$

Through the introduction of attention mechanism, the attention weight α_i is obtained by allocating different weight proportions to different eigenvalues of network traffic. Finally, the forecast target value is output by multiplying α_i with the corresponding eigenvalues.

2.4 Algorithm description

In this paper, the NTAM-LSTM network traffic prediction model is proposed to train the acquired historical network traffic data and accurately predict the network traffic status at the next moment based on the relationship between different network traffic characteristic values. The specific algorithm steps and procedures are shown in Table 1:

Table 1. NTAM-LSTM prediction algorithm steps.

Input: Historical network traffic dataset X

Output: The predicted value of the next moment y

---

(1) The original traffic dataset is preprocessed to obtain X'

(2) $Input = X'$

(3) For the traffic value x_i in X'

(4) $dataset = \{x_1, x_2, \cdots, x_n\}$

(5) Segmentation $dataset\{train, test\}$

(6) Input the train set in model

(7) LSTM layer is used for initial prediction to solve the target value

(8) Compared the result with the test data

(9) Repeat step (6)(7) to reach the number of iterations

(10) Input M' into attention layer to obtain $\alpha = \dfrac{\exp(M)}{\sum_i \exp(M)}$

(11) x_i plus α to get the result y

3 Performance evaluation and result analysis

3.1 Experimental dataset

Experiments in this paper used the network traffic data collected from the university of new Brunswick (UNB), network security institute and Canada (CIC) public release of traffic dataset, which covers all data from March 7 to 11, 2017 in the network traffic. The test used the data from 9:00 to 9:15 on March 7 which consists of the 10000 groups of traffic data generated, and the optimal network traffic characteristics could be obtained by reporting [14], including: B.packet Len Min, Subflow F. bytes, Total Len F.packets, F. packet Len Mean. The first 9000 sets of data were selected for the training of NTAM-LSTM model, and the last 1000 were used to test the performance of the system model.

In the simulation experiment, the network traffic data was preprocessed. Because there were missing and invalid values in the original network traffic dataset, the dataset needed to cleaned up. Then the invalid values were deleted, and the missing values were interpolated. The magnitudes of the data in the traffic dataset were different, in order to eliminate the effects, the dataset was normalized. The results not only accelerated the convergence rate of the model, but also did not affect the characteristics of network traffic. In this paper, we

used Min-max Normalization of network traffic dataset, and its transformation function is as follows:

$$x' = \frac{x - \min}{\max - \min}$$

(11)

After the normalized processing, the original dataset is scaled in an equal scale, and the final range is between 0 and 1, ensuring the reliability of the results.

3.2 Experimental evaluation index

In this paper, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used in the performance evaluation of NTAM-LSTM network traffic prediction model. The calculation formula of the two is:

$$MAE = \frac{\sum_{i=1}^{N} y_i^* - y_i}{N}$$

(12)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} y_i^* - y_i}{N}}$$

(13)

In the formula, N represents the number of data samples, y_i represents the true value of the network traffic in the samples, and y_i^* represents the predicted value.

3.3 Analysis of simulation results

In order to verify the overall performance of the NTAM-LSTM network traffic prediction model, this paper adopts the traditional LSTM prediction model, GRU prediction model [15] and TCN prediction model [16] as the comparison model, and verifies the performance of the NTAM-LSTM model by predicting and analyzing the network traffic in the same time period. The experimental results of the test are shown in Figure 5 (a-d), and the comparison results of the error values of different algorithms are shown in Table 2.

Table 2. Model error results of different prediction algorithms.

	MAE	RMSE	Elapsed time/s
LSTM	3.215	4.690	1.115
GRU	1.776	3.426	1.100
TCN	2.501	4.093	0.503
NTAM-LSTM	0.401	0.637	0.045

According to the observation and comparison of the experimental results, the three comparison models can all predict the network traffic to a certain extent, but their prediction effects are different.

The traditional LSTM contains a special internal gate structure, which is effective in predicting network traffic. However, the observation results show that the fitting degree between the predicted value and the real value of the traditional LSTM is poor, MAE and

RMSE are much greater than that of NTAM-LSTM, and the model also takes a long time to run. The reason is that the length of the traditional LSTM is fixed when processing sequence eigenvalues, and the final result cannot be determined according to the different influence factors of different eigenvalues, so the prediction target value obtained is inaccurate. The experimental results of GRU are better than the traditional LSTM, and the error values of MAE and RMSE are reduced. Because GRU uses update gate and reset gate, the complexity is reduced, so the error and the training time are fewer. However, compared with NTAM-LSTM, the error value is slightly larger and the running time of the model is slightly longer. The reason is that GRU does not distinguish the difference of influence factors between different eigenvalues. TCN is improved based on the convolutional neural network CNN. The gate structure is removed and the residual structure is introduced. Compared with NTAM-LSTM, the fitting degree between the predicted value and the true value of the flow in TCN prediction results is poor, MAE and RMSE are greater than that of NTAM-LSTM. The reason is that TCN will be limited when using convolution, and it is difficult to grasp the correlation of different traffic characteristic values.
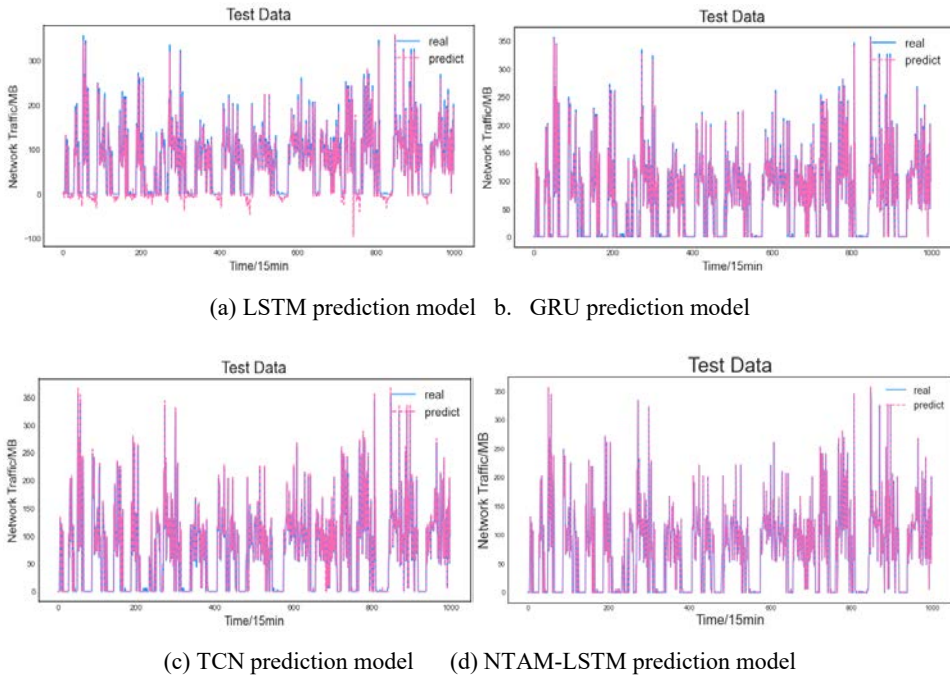

(a) LSTM prediction model   b.   GRU prediction model

(c) TCN prediction model   (d) NTAM-LSTM prediction model

Fig. 5. Comparison of results of different prediction models.

According to the comparison of traffic prediction errors and running time of these algorithms, it can be concluded that NTAM-LSTM model introduces attention mechanism to improve the traditional LSTM, and distinguishes different traffic characteristic values. In the prediction experimental results, the coincidence degree between the predicted value and the real value is the highest, and MAE and RMSE values are the lowest. The model takes the least time to run.

4 Conclusion

After the accurate prediction of network traffic, better analysis and decision-making can be achieved in the aspects of network resource allocation. In this paper, NTAM-LSTM model

is proposed. Based on the traditional LSTM prediction model, attention mechanism is introduced to improve the model, which solves the problem of inaccurate prediction results caused by the traditional LSTM model setting up the same length for different characteristic values. The simulation experiment shows that compared with the traditional LSTM model, GRU model and TCN model, the NTAM-LSTM model has higher prediction accuracy and takes shorter running time, and it can grasp the changes of network traffic better, which is helpful for the subsequent analysis and decision making of network resource allocation. In the future work, the predicted network traffic will be applied to the decision making of network resource allocation to improve the utilization of network resources.

References

1. Li Junyi, Guo Fangce, Sivakumar Aruna et al. Transferability improvement in short-term traffic prediction using stacked LSTM network[J]. Transportation Research Part C, 2021,124: 102977-102995

2. Yonghua Huo, Chunxiao Song, Sheng Gao et al. Network Traffic Prediction Method Based on Time Series Characteristics[C]//Proceedings of the 10th International Conference on Computer Engineering and Networks, 2020: 1547-1555.

3. Hui Xia, Chenhao Zhao, Zhiguo Tang et al. Application of ARMA Model in Prediction of Development Trend of Partial Discharge[C]//Poceedings of 5th International Conference on Vehicle, Mechanical and Electrical Engineering. 2019: 395-402.

4. Jian Zhou, Qidong Yang, Xiaofei Zhang et al. Traffic Prediction Method for GEO Satellites Combining ARIMA Model and Grey Model[J]. Journal of Shanghai Jiaotong University,2020,25(8):65-69.

5. Makuvaza Auther, Jat Dharm Singh, Gamundani Attlee M. Deep Neural Network (DNN) Solution for Real-time Detection of Distributed Denial of Service (DDoS) Attacks in Software Defined Networks (SDNs)[J]. SN Computer Science,2021,2(2):107-117.

6. Zhongda Tian. Network traffic prediction method based on wavelet transform and multiple models fusion[J]. International Journal of Communication Systems,2020, 33(11):4415-4440.

7. Lei Ming, Qin Rui, Mao Wentao, Lu Hongxia. Traffic data prediction of mobile communication base station based on wavelet neural network[J]. Journal of Physics: Conference Series,2021,1883(1):12065-12071.

8. S. Sharma, L. Parashar and U. Singh. Network Traffic Prediction Using Long Short-Term Memory[C]//2020 International Conference on Electronics and Sustainable Communication Systems. 2020:338-343.

9. R. Vinayakumar, K. P. Soman and P. Poornachandran. Applying deep learning approaches for network traffic prediction. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017: 2353-2358.

10. Bai S , Kolter J Z , Koltun V . Trellis Networks for Sequence Modeling[J]. 2018.

11. Luming Bai, Hui Wang, Zhibo Zhai. Research on Network Traffic Forecast Based on Improved LSTM Neural Network[J]. International Core Journal of Engineering,2020,6(6):225-234.

12. Hongzhi Liu, Yingpeng Du, Zhonghai Wu. AEM: Attentional Ensemble Model for personalized classifier weight learning[J]. Pattern Recognition,2019,96:106976-106984.

13. Song Shuang, Li Shugang, Zhang Tianjun et al. Research on a Multi-Parameter Fusion Prediction Model of Pressure Relief Gas Concentration Based on RNN[J]. Energies,2021,14(5):1384-1384.

14. Sharafaldin, I., Lashkari, A. and Ghorbani, A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffific Characterization[C]//Proceedings of the 4th International Conference on Information Systems Security and Privacy,2018:108-116.

15. Gao Shuai, Huang Yuefei, Zhang Shuo et al. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation[J]. Journal of Hydrology,2020, 589:125188-125197.

16. L. Zhang et al. LNTP: An End-to-End Online Prediction Model for Network Traffic[J]. IEEE Network, 2021: 226-233.