Long-Term Span Traffic Prediction Model Based on STL Decomposition and LSTM

1st Yonghua Huo
The 54th Research Institute of CETC
Shijiazhuang, China
tsdhyh2005@163.com

2nd Yu Yan
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China
2017459823@qq.com

3rd Dan Du
1st Military Representative Office of Military Equipment Shijiazhuang
Shijiazhuang, China
1437011418@qq.com

4th Zhihao Wang
The 54th Research Institute of CETC
Shijiazhuang, China
258214542@qq.com

5th Yixin Zhang
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China
953796789@qq.com

6th Yang Yang*
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China
yyang@bupt.edu.cn

***Abstract*—With the increasing complexity of the network, the current network traffic has strong nonlinearity and burstiness. Therefore, the traditional traffic prediction model is no longer applicable. The neural network model, especially the LSTM, can well fit the nonlinearity of time-series data and preserve the information memory of the past. However, as for the periodicity of long-term span network traffic data, the neural network model does not perform well. Based on this, this paper proposes LTS-TP (Long-Term Span Traffic Prediction model), a network traffic prediction model, to solve the problem. First, the model decomposes the collected network traffic data using the improved STL decomposition algorithm to preserve the seasonal component. Then, the trend component and the remainder component are input into the Seq2Seq model based on the LSTM added with the improved attention mechanism for prediction. Finally, the predicted value of the output is added to the seasonal component, and the final network traffic prediction value is obtained. In the simulation part, this paper uses the MAWI public data set to test the proposed network traffic prediction model and compared performance with other models. The results show that the network traffic prediction model proposed in this paper has a good predictive effect on long-term span network traffic data.**

Keywords—network traffic prediction, STL decomposition, LSTM, Seq2Seq, attention mechanism

I. INTRODUCTION

With the rapid development of information technology, the topology of the network has become increasingly complex, and the state of the network has become highly variable. Therefore, a more reasonable and effective traffic management strategy is formulated to better cope with various sudden problems such as network congestion and provide users with a high-quality network environment.

As a kind of time series data, network traffic has short-term characteristics and long-term characteristics. The traditional linear models such as Autoregressive Integrated Moving Average (ARIMA) model [1] are simple to implement and has good effects on short-term network traffic data prediction, but it is not suitable for frequently changing network structures. Therefore, this paper proposes a network traffic prediction model based on Seasonal-Trend decomposition procedure based on Loess (STL) and Long Short-Term Memory (LSTM). For the periodic characteristics of long-term span sequences, this paper uses STL decomposition algorithm and improves it to realize multi-periodic decomposition and extraction of network traffic data. In view of the long correlation of network traffic data, this paper improves the attention mechanism used in the neural network model, and adds information capture to important data points before long sequence spans. Thereby enhancing the feature fitting of the long-term span sequence and improving the accuracy of network traffic prediction.

The rest of this paper is organized as follows. It first introduces the related work, then describes the proposed Long-Term Span Traffic Prediction model (LTS-TP). This paper then introduced the data set used and compare the performance of proposed model with other sequence prediction model. Finally, this paper is concluded.

II. RELATED WORK

Commonly used artificial intelligence-based network traffic prediction methods include Elman neural network [2], BP neural network [3], Echo State Networks (ESN), and Extreme Learning Machine (ELM). When the neural network is trained by the gradient descent algorithm, it often has the defect that it is difficult to converge to the global optimal and easily fall into the local optimum. Many researches have improved this defect. The improved methods include Particle Swarm Optimization (PSO) [4] and Genetic Algorithm.

Ji Yimu et al. [5] proposed a new network traffic prediction method based on improved Echo State Networks to deal with the complex characters of network traffic, such as mutability, chaos, timeliness, and nonlinearity.

Jie Feng et al. [6] proposed Deep Traffic Predictor (DeepTP) to forecast traffic demands from spatial-dependent and long-period cellular traffic, which can be divided into two components: a general feature extractor for modeling spatial dependencies and encoding the external information, and a sequential module for modeling complicated temporal variations.

The family of Recurrent Neural Network (RNN) is often used for time series predictions and has a good performance. Vinayakumar, R. et al. [7] used various RNN networks to leverage the efficacy of RNN approaches towards traffic matrix estimation in large networks. The performance of various RNN networks is evaluated on the real data from GÉANT backbone networks. The results show that LSTM has a best performance in comparison to the other RNN and classical methods.

However, in the existing network traffic prediction research, it is impossible to fit the network traffic data with long-term span of periodicity and long correlation well. Aiming at this problem, this paper proposes an improved network traffic prediction algorithm: Long-Term Span Traffic Prediction model (LTS-TP), which combines STL decomposition algorithm and LSTM-based Seq2Seq model to predict network traffic, and improves the STL decomposition and attention mechanism.

III. Design of Network Traffic Prediction Model Based on STL-SEQ2SEQ Styling

In this paper, an improved network traffic prediction algorithm: Long-Term Span Traffic Prediction model (LTS-TP) is proposed for the periodicity and long correlation of long-term span network traffic data. The data is decomposed and preprocessed by the improved STL decomposition algorithm, and then the data is input into the Seq2Seq model with the improved attention mechanism for traffic prediction.

A. Improved STL Decomposition

Seasonal-Trend decomposition procedure based on Loess (STL) is a commonly used time series decomposition algorithm with strong robustness. The algorithm is based on locally weighted scatterplot smoothing, and the time series is decomposed into several subsequences by the additive model: trend component, seasonal component and remainder component.

Due to the strong periodic characteristics of long-term network traffic, this paper intends to decompose it periodically. The specific method is as follows: Firstly, the network traffic data is decomposed into seasonal component, trend component and remainder component by using STL decomposition algorithm. Subsequently, the seasonal component is retained and the trend component and the remainder component are input into the code-decoder model with the attention mechanism, thereby outputting the corresponding predicted value. Finally, the model output of the previous step is added to the retained seasonal component to obtain the final network traffic prediction value.

Considering that the long-term span network traffic data has multiple periodicities, the traditional STL decomposition algorithm can only extract a certain period when extracting seasonal components. Therefore, this paper improves the addition model of STL, and decomposes the time series into trend component, multiple seasonal component and remainder component. The specific expressions are as follows:

$$Y_v = T_v + S_{\varpi 1} + S_{\varpi 2} + \cdots + S_{\varpi i} + R_v \qquad (1)$$

T_v represents the trend component, $S_{\varpi i}$ represents the $i^{\text{-th}}$ seasonal component, R_v represents the remainder component, v=1, 2, ..., N represents a certain moment.

B. Seq2Seq network with improved attention mechanism

After the original network traffic data is periodically decomposed using STL decomposition algorithm, the obtained sequence needs to be input into the neural network for training. In this paper, the Seq2Seq model based on LSTM is selected in the choice of neural network model, and the attention mechanism is incorporated into the network. The model uses LSTM for prediction, which not only fits the nonlinear characteristics of network traffic data, but also avoids the gradient descent and gradient explosion problems that traditional RNN will face. The addition of improved attention mechanisms can also enable neural networks to learn more effective information, making predictions more accurate.

Since the traffic sequence to be predicted has a long correlation, when using the attention mechanism, the author finds that the time point which can have a greater impact on the current (that is the point at which attention should be focused) will have a long sequence span with the current time point. If the input sequence is set long, the attention mechanism needs to traverse the state of all hidden layers and determine its similarity to the state of the current hidden layer to give a score, but this will result in a large time overhead. However, if the input sequence is set to be short, the information for a long time will be lost, which will affect the accuracy of network traffic prediction.

Therefore, this paper improves the model by combining the data characteristics of network traffic when using the attention mechanism, and expects to enhance the accuracy of network traffic prediction without generating large time overhead.

The general attention mechanism scores all previous hidden layer states, and then assigns the corresponding weights, weights them, and passes them to the decoder. But when the time series is very long, there are a few important nodes that really have a big impact on the final output. Considering that network traffic has strong seasonality, there are two important time points in the long-term prediction, one year ago and one quarter ago.

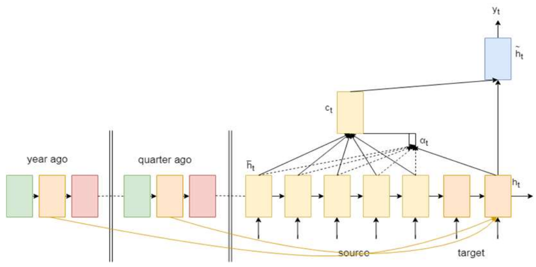

Fig. 1. the structure of the improved attention mechanism

Based on the general attention mechanism, this paper adds learning to these important time points. The specific processing method is:

- Keep the data of the two time points" one year ago" (current_day-365) and "one quarter ago" (current_day-90), and consider that due to the long-term series, year and season may have some offsets. Therefore, when acquiring data, it is necessary to consider the nearby nodes and weight them:

$$i_{365} = \alpha_1 x_{364} + \alpha_2 x_{365} + \alpha_3 x_{366} \qquad (2)$$

$$i_{90} = \alpha_4 x_{89} + \alpha_5 x_{90} + \alpha_6 x_{91} \qquad (3)$$

α is the weighted value of the encoder output of the nearby node. For the output of the encoder at the same time point, the sum of the items α is one.

- Input the weighted value into a fully connected layer for dimensionality reduction, which can be expressed as:

$$f = \beta_1 i_{365} + \beta_2 i_{90} \tag{4}$$

β_i is the weight of the fully connected layer

- The processed data f is input as a feature to the decoder for decoding, so that the decoder can better learn the information at important time points.

- The original attention mechanism is retained, and the state of the hidden layer of the encoder is scored and then weighted $\alpha_{t,i}$ using the score() function. The transfer vector c_t is calculated, and the transfer vector is synthesized to obtain the hidden layer state at time t, that is, the output.

$$\alpha_{t,i} = \frac{exp(score(h_t, \bar{h}_i))}{\sum_{i'} exp(score(h_t, \bar{h}_{i'}))} \tag{5}$$

$$c_t = \sum_i \alpha_{t,i} \bar{h}_i \tag{6}$$

$$\tilde{h}_t = tanh(W_c[c_t; h_t]) \tag{7}$$

C. Adaptive Moment Estimation

This paper uses the more effective Adaptive Moment Estimation (Adam) gradient descent algorithm when updating the gradient. The Adam algorithm adjusts the learning rate of the parameters according to the exponential decay values of the previous gradient and the previous square gradient. The specific steps are as follows:

- First, initialize the first and second order cumulants of the gradient.

$$v = 0, s = 0 \tag{8}$$

- In the t^{th} training, the parameters are updated according to the previous gradient cumulants, where β is the exponential decay parameter of the gradient accumulation, β_1 is often set to 0.9, and β_2 is often set to 0.9999.

$$v_t = \beta_1 v_{t-1} + (1 - \beta_1) dW \tag{9}$$

$$s_t = \beta_2 s_{t-1} + (1 - \beta_2) dW^2 \tag{10}$$

- Correct the offset of the above values.

$$v_t^c = \frac{v_t}{1 - \beta_1^t} \tag{11}$$

$$s_t^c = \frac{s_t}{1 - \beta_2^t} \tag{12}$$

- Update the parameters based on the corrected values obtained.

$$W = W - \alpha \frac{v_t^c}{\sqrt{s_t^c + \varepsilon}} \tag{13}$$

α is the learning rate, and ε is the smoothing term, which is often taken as 10^{-8}.

IV. SIMULATION RESULTS AND COMPARATIVE ANALYSIS

In this part, we will simulate the network traffic prediction model proposed in this paper, and compare the experimental results with the original model: classical LSTM, and Seq2Seq model based on LSTM with soft attention mechanism. All of these models use the Adam algorithm to update the parameters.

A. Data set

The data set used in this paper is the public network traffic dataset on the MAWI website. The data set is network traffic data recorded every two hours from March 7, 2013 to April 14, 2019, with a total of 27,065. We divide the original data set into 6 data sets according to the length of time, and further divide into training sets and test sets. The results are shown in the TABLE I:

TABLE I. THE PROCESSED DATA SET

	Time	Length	Train_data _length	Test_data _length
Datapart 1	2013.02.07-2 014.05.04	5413	4380	1033
Datapart 2	2013.02.07-2 015.07.29	10826	8760	2066
Datapart 3	2013.02.07-2 016.10.23	16239	13140	3099
Datapart 4	2013.02.07-2 018.01.17	21652	17520	4132
Datapart _all	2013.02.07-2 018.04.14	27064	24090	2974

B. Prediction error indicator

In the evaluation of the network traffic prediction model, this paper uses the Root mean squared error (RMSE) and the mean absolute error (MAE) to measure the accuracy of the network traffic model prediction.

RMSE is also called the standard deviation, which is the square root of the average of the difference between the predicted value and the actual value of the parameter.

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2} \tag{14}$$

m is the number of predicted values, y_i represents the actual value of the parameter, and $\hat{y}_i$ represents the predicted value obtained by the model.

MAE is the average of the absolute values of the deviations between the predicted values and the actual values of all parameters.

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |y_i - \hat{y}_i| \tag{15}$$

m is the number of predicted values, y_i represents the actual value of the parameter, and y_i represents the predicted value obtained by the model.

The smaller RMSE and MAE are, the more accurate predicted values are. This means that the trained neural network is more fitted to the actual model.

C. Performance comparison analysis

In this paper, the data set is processed and input into the designed network traffic prediction model for prediction. In order to better compare the prediction results of each model, this paper selects the data segments with faster changes for display. Fig. 2. shows the prediction results of each model on the overall data set:

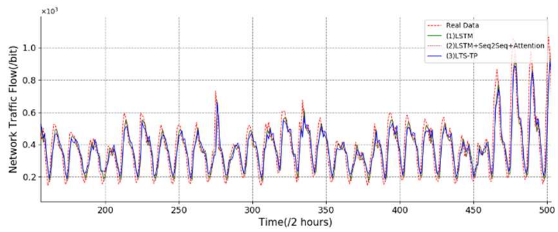

Fig. 2. prediction results of the datapart_all

In order to better demonstrate the prediction effects of each model, this paper calculated the evaluation indicators of each model on the six data sets, which are shown in TABLE II.

TABLE II. THE RMSE AND MAE OF EACH ALGORITHM

	LSTM		LSTM+Seq2Seq+Attention		LTS-TP	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Datapart 1	0.08058701	0.055201931	0.081003884	0.05525528	0.079264825	0.053979541
Datapart 2	0.058579279	0.044725155	0.059927939	0.045160914	0.05686983	0.04305841
Datapart 3	0.02746894	0.01944536	0.027315593	0.019582542	0.025713635	0.01802445
Datapart 4	0.029253097	0.021953303	0.029314188	0.021912744	0.026829304	0.019785842
Datapart _all	0.045999928	0.033767929	0.046031982	0.033739659	0.039746352	0.029878396

The RMSE in all dataset partitions of Long-term TP is 6.57% better than the LSTM model and 7.05% better than the LSTM+Seq2Seq+Attention model. The MAE in all dataset partitions of Long-term TP is 6.93% better than the LSTM model and 7.21% better than the LSTM+Seq2Seq+Attention model. On the total dataset, the RMSE of Long-term TP is 13.59% better than the LSTM model and 13.65% better than the LSTM+Seq2Seq+Attention model. On the total dataset, the MAE of Long-term TP is 11.52% better than the LSTM model and 11.44% better than the LSTM+Seq2Seq+Attention model. Experiments show that the Long-term TP model proposed in this paper has a good predictive effect on long-term span network traffic data.

The following is a visual representation of the RMSE and MAE percentage reduction of the LTS-TP model relative to the two comparison models in different data sets, as shown in Fig. 3. and Fig. 4.

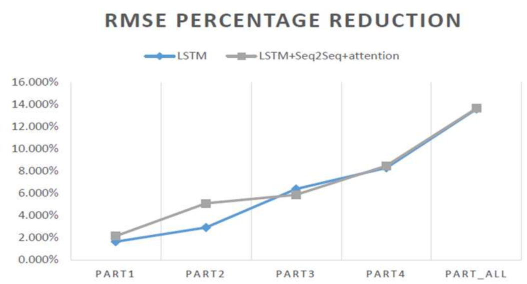

Fig. 3. The RMSE percentage reduction of each data set

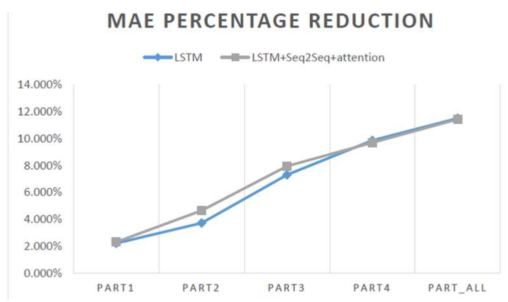

Fig. 4. The MAE percentage reduction of each data set

As can be seen from the figures, as the size of the data set increases, RMSE and MAE percentage reductions of the LTS-TP model relative to the other two comparison models are increasing. When the data time span is small, the performance optimization effect is not obvious enough. This reflects that the LTS-TP model is more suitable for fitting long-term span sequences. It can effectively learn the periodicity and long correlation of long-term span sequences, thus achieving more accurate network traffic prediction.

Overall, the LTS-TP network traffic prediction model proposed in this paper has a great improvement in prediction accuracy, especially for long-term span sequences.

V. CONCLUSION

Aiming at the Periodicity and long correlation of long-term span network traffic, this paper proposes an improved network traffic prediction algorithm: Long-Term Span Traffic Prediction model (LTS-TP). In the improvement of the STL decomposition, this paper firstly uses Fourier analysis to obtain its periodicity, and then extracts and stores its corresponding seasonal components for each period, finally retains the trend component and remainder component obtained in the last period extraction. In the improvement of attention mechanism, this paper regards one year ago and one quarter ago as the important time nodes, and the data of these time nodes are inputted into the decoder through a fully connected layer for prediction. Based on these improvements, LTS-TP can effectively fit the periodicity and long correlation of network traffic data without significantly increasing the time consumption. In this paper, the proposed model is simulated by MAWI public data set and the results show that the LTS-TP can effectively improve the accuracy of network traffic prediction. LTS-TP model has certain improvement in performance for traffic prediction, but there are still some limitations, such as the burstiness characteristic of network traffic data cannot be well fitted, which could reduce the prediction accuracy. We will further improve our model in the future.

ACKNOWLEDGMENT

This work was supported by Open Subject Funds of Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory (SKX182010049), National Key Research and Development Programme of China (2016YFE0204500), the Industrial Internet Innovation and Development Project 2018 of China.

REFERENCES

[1] P. Wang, S. Zhang and X. Chen, "SFARIMA: A New Network Traffic Prediction Algorithm," 2009 First International Conference on Information Science and Engineering, Nanjing, 2009, pp. 1859-1863.

[2] X. Wang, C. Zhang and S. Zhang, "Modified Elman neural network and its application to network traffic prediction," 2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems, Hangzhou, 2012, pp. 629-633.

[3] P. Wang and Y. Liu, "Network Traffic Prediction Based on Improved BP Wavelet Neural Network," 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, Dalian, 2008, pp. 1-5.

[4] K. Zhang, L. Liang and Y. Huang, "A Network Traffic Prediction Model Based on Quantum Inspired PSO and Neural Network," 2013 Sixth International Symposium on Computational Intelligence and Design, Hangzhou, 2013, pp. 219-222.

[5] J. Feng, X. Chen, R. Gao, M. Zeng and Y. Li, "DeepTP: An End-to-End Neural Network for Mobile Cellular Traffic Prediction," in IEEE Network, vol. 32, no. 6, pp. 108-115, November/December 2018.

[6] J. Zhou, X. Yang, L. Sun, C. Han and F. Xiao, "Network Traffic Prediction Method Based on Improved Echo State Network," in IEEE Access, vol. 6, pp. 70625-70632, 2018.

[7] 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), Udupi, 2017, pp. 2353-2358